Harnessing the Power of Cryo Electron Microscopy to Visualise a Full-length Kinesin KIFC1 (HSET)

Varun Ramaswamy

Divisions of Cancer Therapeutics and Structural Biology

The Institute of Cancer Research

University of London

This thesis is submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy.

March 2023

DECLARATION

I, Varun Ramaswamy, declare that the work presented in this thesis is my own and, where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This thesis has not been previously submitted for a degree at this or any other university or institution, in part or in whole.

Signature

30/03/2023

Date

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my supervisor Dr. Rob Van Montfort, for giving me the opportunity to work on this exciting project. Rob has been a great mentor and advisor with regards to my project as well as my career. I would also like to thank my associate supervisors Dr. Catarina Rodrigues and Dr. Yann-Vaï Le Bihan. Catarina oversaw my experiments from the first day of my PhD, nurturing me to become an independent scientist, and was a constant pillar of professional and emotional support. Yann-Vaï took over the role of my associate supervisor when Catarina was on maternity leave and remained a part of my supervisory team till the very end. I am always grateful for his advice and the detailed discussions we had on the principles of different biophysical techniques. I extend my thanks to Dr. Craig McAndrew for his training on protein production and purification, as well as devoting time to have engaging discussions about my project. Craig is the epitome of patience and undying curiosity, and I have learned a lot about presentations and scientific communication while answering his questions during our casual discussions.

I am also immensely grateful to my secondary supervisor and EM facility manager Dr. Fabienne Beuron, who has patiently trained me throughout the four years of my PhD on all my EM experiments. In addition to sample preparation, imaging, and processing in negative stain and cryo EM, Fabienne has been an integral part of every moment of hope and despair that I experienced during my research. She is currently retiring after 17 years of service to the ICR, and I am fortunate to have worked with her until the last day of my PhD. The EM side of my project would also not have been successful without the numerous regular discussions with Dr. Ed Morris. I am incredibly grateful for his consistent mentorship despite his retirement from the ICR. The promptness, enthusiasm, and excitement that Ed showed towards every micrograph, class average and extra angstrom of resolution that I achieved has been nothing short of a blessing.

During my PhD, I had the opportunity to interact with people from different areas of research within the ICR, who made important contributions to my project. I sincerely thank Dr. Harry Saville of the Breast Cancer Research-Functional Genomics team for helping me out with the ADP-Glo assay, as well as Dr. Arjun Thapaliya and Dr. Harshnira Patani for advice on the molecular biology experiments, and Dr. Caroline Ewens of the Hit Discovery and Structural Design team for her valuable discussions on kinesin functionality. I am also thankful for the exciting brain-storming sessions with team leaders Dr. Basil Greber and Dr. Claudio Alfieri from the Division of Structural Biology. Additionally, I express my gratitude to Dr. Paula da Fonseca, for her valuable advice and encouraging words.

I am grateful to the biophysics facility manager Dr. Stephen Hearnshaw for his support with the Mass Photometer and DLS instrument, as well as Dr. Chris Richardson for managing and troubleshooting the computational resources. I would also like to acknowledge the divisional laboratory managers Jane Sandall and Lynn Winsor, who work tirelessly to ensure all laboratory equipment are always functional in the Division of Structural Biology. My project involved the significant usage of external facilities, and I would like to thank the facility managers- Dr. Nora Cronin, LONCEM and Dr. Katsuaki Inoue, Diamond Light Source, for the opportunities and the support with using the Titan Krios microscope and the SAXS beamline respectively.

Last but not the least, a PhD is as much of a learning experience outside the lab as it is inside the lab. I would therefore like to acknowledge all my friends in the Divisions of Cancer Therapeutics and Structural Biology, who made the ICR and London a home for me over the past four years. I would especially like to mention Dr. Roxanne Smith, Elena Mihalachi Luiza, Mandy Wan, and Maria Semiannikova, for helping me through difficult times in the year 2020.

ABSTRACT

Cancer cells are known to exhibit a phenomenon typically known as centrosome amplification (CA), which is responsible for multipolar spindle formation during mitosis. This severely impairs cytokinesis and could hence prove fatal to the cells. However, in these cancer cells, a specific microtubule-associated motor protein KIFC1 (also known as HSET) which has the ability to cluster centrosomes is overexpressed, thereby resolving the spindle multipolarity and facilitating bipolar mitosis. Recent studies have revealed that HSET knock-down renders the CA-positive cancer cells non-viable, while having no effect on the viability of non-cancerous cells. This specificity makes HSET an attractive target for anti-cancer drug discovery.

HSET is a minus-end directed motor and a member of the kinesin-14A family. It has a N-terminal cargo-binding domain followed by a long coiled coil stalk connected to a C-terminal catalytic motor domain. While a crystal structure of the motor domain of HSET bound to ADP has been reported, the inherent flexibility of the coiled coil and cargo-binding domains has rendered crystallization attempts of the full-length protein unsuccessful. Thus, little is known about the global structural changes that result from the communication between the three domains of HSET.

My aim is to determine a full-length structure of HSET in its Apo and ADP-bound forms using cryo electron microscopy combined with complementary biophysical characterization. Here I report 3D reconstructions of the Apo- and ADP-bound full-length HSET at a resolution of 20 Å using negative stain EM. An ADP-HSET structure was determined at a resolution of 8.5 Å by cryo EM using a Volta Phase Plate. The maps provide the first structural insight into the HSET homodimer and show the motor domains in an asymmetric arrangement with respect to the coiled coil stalk. The maps highlight two unique conformations of Apo-HSET and a single conformer of HSET in its ADP-bound form.

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1 The Puzzle of Intracellular Movement 1

1.1.1 ATPases- The Mode of Transport 1

1.1.2 Microtubules- The Road for Transport 2

1.1.3 The Translocators: Captured and Christened 3

1.1.4 The first look at the shape of kinesins 4

1.2 A Family of Kinesins- Where There's One, There's Many More.. 7

1.2.1 A seemingly limitless display of diversity 7

1.2.2 Initial attempts at kinesin classification 9

1.2.3 A promising solution to the consistency conundrum 10

1.2.4 A standardised nomenclature 11

1.3 A Divergence in the *Direction* of Kinesin Research 13

1.3.1 An opposing minority in the superfamily 13

1.3.2 Domain architecture of the kinesin-14 family 14

1.4 Decoding The Motion of a Kinesin-14 17

1.4.1 The processivity of kinesin-14 17

1.4.2 The power stroke of kinesin 14 18

1.4.3 A tale of two motor domains 20

1.4.4 The mechanism of kinesin-14A in visual snapshots 21

1.5 Biological Significance of kinesin-14A 25

1.5.1 Centrosomes and kinesin-14 25

1.5.2 HSET: a kinesin-14 that promotes the sustenance of cancer 26

1.5.3 HSET in Anti-Cancer Drug Discovery 28

1.6 The Aims of My PhD- a Future for HSET Drug Discovery 30

1.6.1 Cryo EM- the solution to the HSET conundrum 30
1.6.2 The aims of this project 31

2. MATERIALS AND METHODS 33

2.1 Protein Production and Purification 33
2.1.1 FL HSET Constructs 33
2.1.2 Protein Expression 34
2.1.3 Protein Purification 35
2.1.3.1 His-tagged FL HSET 35
2.1.3.2 2xStrep-tagged FL HSET 36

2.2 Biochemical and Biophysical Characterisation 37
2.2.1 ATP Hydrolysis Assay 37
2.2.2 Differential Scanning Fluorimetry 37
2.2.3 Size Exclusion Chromatography- Multi Angle Light Scattering (SEC-MALS) 38
2.2.4 Dynamic Light Scattering (DLS) 38
2.2.5 Mass Photometry 38
2.2.6 Small Angle X-ray Scattering (SAXS) 39

2.3 Negative Stain Electron Microscopy 41
2.3.1 Sample Preparation 41
2.3.2 Data Collection 41
2.3.3 Data Processing 42
2.3.3.1 Sample Characterisation 42
2.3.3.2 Structural Analysis 42

2.4 Cryo Electron Microscopy 44
2.4.1 Sample Preparation 44
2.4.2 Data Collection 44
2.4.2.1 Glacios 44
2.4.2.2 Titan Krios 45
2.4.3 Data Processing 45
2.4.3.1 Pre-Processing 45
2.4.3.2 Particle-Picking 45
2.4.3.3 2D and 3D Analysis 46

3. RESULTS 47

3.1 Production and Purification of FL HSET 47

3.1.1 His-tagged FL HSET 47

3.1.1.1 Step-1: IMAC using Ni^{2+}-NTA column 48

3.1.1.2 Step-2: SEC using Superdex 200 48

3.1.1.3 Step-3: IEC using cation-exchange 50

3.1.2 Twin-StrepII-tagged FL HSET 51

3.1.2.1 Step-1: Affinity Chromatography using StrepTrap 52

3.1.2.2 Step-2: SEC using Superdex 200 52

3.1.3 Optimising the twin-StrepII strategy 54

3.1.3.1 Improving the resolution in SEC 54

3.1.3.2 Preparation of ADP-HSET 55

3.2 Biochemical and Biophysical Characterisation of FL HSET 58

3.2.1 Confirming the functionality of HSET 58

3.2.1.1 Principle of ADP-Glo assay on HSET 59

3.2.1.2 Confirmation of HSET's enzymatic activity in the ADP-Glo assay 61

3.2.2 Investigating ADP-binding through DSF of HSET 62

3.2.2.1 Comparison of Apo and ADP-HSET by DSF 63

3.2.2.2 ADP-binding improves the thermal stability of HSET 63

3.2.3 Determining the oligomeric state of Apo and ADP-HSET 65

3.2.3.1 SEC-MALS of HSET 65

3.2.3.2 Apo and ADP-FL HSET are homodimers 66

3.2.4 Monitoring HSET's tendency for aggregation 67

3.2.4.1 DLS of HSET 67

3.2.4.2 HSET displays temperature-dependent polydispersity variation 68

3.2.5 Determining the salt and glycerol requirement of HSET 70

3.2.5.1 Mass Photometry of HSET 70

3.2.5.2 Apo-HSET tolerates a lower salt concentration than ADP-HSET 71

3.2.5.3 Glycerol proves to be essential for ADP-HSET as well 73

3.2.6 Observing the behaviour of HSET in solution 75

3.2.6.1 Basic Principle of SAXS 75

3.2.6.2 Size of HSET from Fourier space approach 76

3.2.6.3 Degree of structure in HSET 80

3.2.6.4 Shape of FL HSET from real-space approach 81

3.3 Image Formation in Single Particle Transmission Electron Microscopy 84
3.3.1 From object to image 84
3.3.2 Contrast- the essence of EM 87
3.3.2.1 Amplitude Contrast 87
3.3.2.2 Phase Contrast 88
3.3.3 Contrast Enhancement 90
3.3.3.1 A beneficial defect 90
3.3.3.2 Imaging out of focus 91
3.3.3.3 Phase Plates 95
3.4 Negative Stain Electron Microscopy- Data Processing 97
3.4.1 Visual Characterisation of FL HSET variants 97
3.4.2 Low-resolution structure elucidation of FL HSET 101
3.4.2.1 Particle picking 102
3.4.2.2 2D Classification 104
3.4.2.3 3D Analysis 109
3.5 Cryo Electron Microscopy- Sample Preparation 116
3.5.1 Blot-force 118
3.5.2 Grid-type 120
3.5.3 Protein Concentration 120
3.5.4 Glycerol Content 122
3.5.5 Support Films 123
3.6 Cryo Electron Microscopy- Data Collection and Processing 128
3.6.1 The Glacios Sessions 129
3.6.1.1 Live processing and pre-processing 129
3.6.1.2 Particle picking 135
3.6.1.3 2D Classification 137
3.6.1.4 3D Analysis 143
3.6.2 The Titan Krios Session 146
3.6.2.1 Particle picking 148
3.6.2.2 2D Classification 150
3.6.2.3 3D Analysis 154
4. DISCUSSION 162

4.1 Evaluating the Stability of FL HSET 162
4.1.1 Stabilising the structure of FL HSET 162
4.1.2 Minimising the polydispersity of HSET samples 163
4.1.3 Summary 164
4.2 Describing the Size and Shape of HSET 165
4.2.1 Oligomerisation state 165
4.2.2 Size and shape 166
4.2.3 Summary 167
4.3 EM-based Structure Determination of a Small and Elongated Target like FL HSET 168
4.3.1 Particle-picking 168
4.3.2 2D Classification 171
4.3.3 3D Analysis 173
4.3.4 Summary 175
4.4 Phase Plate- Advantageous or Disadvantageous? 176
4.4.1 Initial challenges to VPP data acquisition 177
4.4.2 The consequences of over-development 178
4.4.3 The Do's and Don'ts of using a VPP 178
4.4.4 Summary 179
4.5 Interpreting the FL HSET Structure 181
4.5.1 The overall structure of FL HSET 182
4.5.2 FL HSET: The neck and stalk 183
4.5.3 FL HSET: The motor domain 186
4.5.4 FL HSET: The CBD 189
4.5.5 Summary 191
5. THE NEXT STEPS IN THE STRUCTURAL INVESTIGATION OF HSET 193
5.1 Short-Term Aims 193
5.1.1 Preferred Orientation 193
5.1.2 Apo-HSET 194

5.1.3 Data Processing Strategies 194

5.2 Long-Term Aims 196

5.2.1 Shorter constructs 196

5.2.2 Microtubule binding 196

5.2.3 Compound binding 198

SUPPLEMENTARY INFORMATION 199

S1. Constructs for FL-HSET 199

S1A. His_6-TEV-FLHSET 199

S1B. $Strep_2$-HRV3C-FLHSET 199

S1C. Primers for PCR of $Strep_2$-FLHSET 200

S2. Purification of FLHSET 201

S2A. Molecular Weight Determination from SEC 201

S2B. Biochemical Properties of FL HSET 201

S2C. Concentration and Purity of Eluted HSET Fraction 203

S3. Secondary Structure Prediction of FLHSET 204

BIBLIOGRAPHY 206

LIST OF FIGURES

Figure 1.1 Domain Architecture of Kinesin Heavy Chain 5
Figure 1.2 Domain Architecture of a Kinesin-14 16
Figure 1.3 Reaction Mechanism of a Kinesin-14A 19
Figure 1.4 The mechanochemical cycle of kinesin-14A 23
Figure 3.1 Construct for His_6-FLHSET 47
Figure 3.2 Purification of His-tagged FL HSET 49
Figure 3.3 Construct for $Strep_2$-FLHSET 51
Figure 3.4 Purification of Twin-StrepII-tagged FL HSET 53
Figure 3.5 Comparison of SEC Columns - Superdex 200 10/30 vs. Superose 6 Increase 10/30 54
Figure 3.6 Purification of Strep2-FL HSET-II 56
Figure 3.7 Principle of ADP-Glo Assay 60
Figure 3.8 ADP-Glo Assay of FL HSET Constructs 61
Figure 3.9 Investigating the Thermal Stability of HSET by NanoDSF 64
Figure 3.10 Determining the Oligomeric State of HSET 66
Figure 3.11 Studying the Evolution of Molecular Heterogeneity in Apo and ADP-HSET with temperature 69
Figure 3.12 Salt titration of Apo and ADP-HSET 72
Figure 3.13 Glycerol Titration of ADP-HSET 74
Figure 3.14 SEC-SAXS of Apo and ADP-HSET 76
Figure 3.15 Guinier Analysis of Merged Apo and ADP-HSET Datasets, and Corresponding Full Scattering Plots 77
Figure 3.16 Kratky Analysis of Apo and ADP-HSET 81
Figure 3.17 Low Resolution 3D Modelling of FL HSET 83
Figure 3.18 Schematic of Transmission Electron Microscope (adapted from Orlova, E.V. and Saibil, H.R., 2011) 85
Figure 3.19 Image Formation by Objective Lens 86
Figure 3.20 Amplitude Contrast 88
Figure 3.21 Phase Contrast 89
Figure 3.22 Spherical Aberration in Lenses 91

Figure 3.23 Applied Defocus of the Objective Lens 92
Figure 3.24 Contrast Transfer Function (CTF) 93
Figure 3.25 Effects of Defocus on CTF 94
Figure 3.26 Effect of the Volta Phase Plate on In-Focus Data Acquisition 96
Figure 3.27 Visual Characterisation of FL HSET by Negative stain EM..... 98
Figure 3.28 ISAC2 Serves as an Effective Classification Tool for FL HSET 101
Figure 3.29 CRYOLO Pipeline for Picking Particles of FL HSET..... 103
Figure 3.30 Comparing Classifiers..... 105
Figure 3.31 Alignment Optimisation in RELION..... 107
Figure 3.32 Reference-free 2D Class Averages of Apo and ADP-FL HSET 108
Figure 3.33 Reference-based 2D Classification of Apo and ADP-FL HSET 110
Figure 3.34 3D Classification and Refinement of Apo and ADP-FL HSET 112
Figure 3.35 Final 3D maps of Negative stained Apo and ADP-FL HSET . 114
Figure 3.36 Holey Grids for Negative stain EM and Cryo EM 116
Figure 3.37 Optimisation of Grids for Cryo EM of HSET..... 118
Figure 3.38 Effects of Different Blot Forces on Unsupported Grid-types 119
Figure 3.39 Effects of Protein Dilution on C-flat Grids..... 121
Figure 3.40 Exploration of Ultra-Au-Foil 122
Figure 3.41 Working on Carbon Support..... 125
Figure 3.42 The Effect of Enhanced Phase Contrast..... 126
Figure 3.43 The Different Regimes of Phase Contrast..... 130
Figure 3.44 Distribution of Phase Shifts Through an Overnight Collection of ADP-HSET Data on the Glacios 132
Figure 3.45 Processing an Apo-HSET Dataset..... 134
Figure 3.46 Particle Picking Pipeline in CRYOLO..... 136
Figure 3.47 Reference-free 2D classification of Apo and ADP-HSET 138
Figure 3.48 Purification by Classification..... 139
Figure 3.49 Testing the Effects of a Circular Mask on the 2D Classes of ADP-HSET 140

Figure 3.50 Testing the Effects of High-pass Filters on 2D Classes of ADP-HSET 142
Figure 3.51 Reference-based 2D Classification of ADP-HSET 143
Figure 3.52 Refining the 3D Structure of ADP-HSET 145
Figure 3.53 Phase Development in an ADP-HSET Dataset Collected on the Titan Krios 147
Figure 3.54 Testing Different Denoising Models to Aid Particle Picking in CRYOLO 149
Figure 3.55 Iterative Optimisation of Picking Model in CRYOLO 150
Figure 3.56 Treating the Over-Averaging Issue in 2D Classes of ADP-HSET 151
Figure 3.57 Unravelling Secondary Structure Detail by Using a Circular Mask 153
Figure 3.58 3D Ab-initio Reconstruction of ADP-HSET 154
Figure 3.59 An Impasse in 3D Refinement 156
Figure 3.60 Optimising the Parameters for 3D Refinement 158
Figure 3.61 Final Refined Map of ADP-FL HSET 159
Figure 3.62 Homology Modelling Highlights Divergence from Crystal Structure 161
Figure 4.1 Negative stain 3D Reconstructions of Apo-HSET 183
Figure 4.2 Kinesin-1 v/s. Kinesin-14 184
Figure 4.3 Asymmetry in the Motor Domains of ADP-HSET 186
Figure 4.4 Nucleotide-binding Sites of the Motor Domains in ADP-HSET 188
Figure 4.5 Lengths of Domains in FL HSET 190

LIST OF TABLES

Table 1.1 Sequence Homology in the Kinesin-14 Family 15
Table 3.1 SAXS Analysis of Apo and ADP-HSET 79
Table 3.2 Narrowing Down the Causes of Blurred Images in Apo and ADP-HSET Datasets 133

LIST OF ABBREVIATIONS

°C - degrees Celsius

2D - Two Dimensional

3D - Three Dimensional

Å - Angstroms

AC - Affinity Chromatography

ADP - Adenosine 5'- diphosphate

Amp – ampicillin

AMP – Adenosine 5'- monophosphate

AMP-PNP – Adenylyl imidiophosphate

An - *Aspergillus nidulans*

ATP – Adenosine 5'- triphosphate

ATPase- ATP hydrolase

Au - Aurum (Gold)

AVEC – Advanced Video- Enhanced Contrast

BimC4 – Block In Mitosis C4

C terminal – Carboxy terminal

C. elegans – *Caenorhabditis elegans*

CBD - Cargo-Binding Domain

CHO2 – Chinese Hamster Ovary 2

CNN – Convolutional Neural Network

CPU - Central Processing Unit

CRYOLO - CRyo You Only Look Once

cryoSPARC - Cryo EM Single Particle Ab-initio Reconstruction and Classification

Cs- Coefficient of Spherical Aberration

CTF- Contrast Transfer Function

Cu - Cuprum (Copper)

CV - Column Volume

DAM – Dummy Atom Modelling

DAMAVER - Dummy Atom Model Averaging

DAMFILT - Dummy Atom Model Filtering

ΔT_m - Difference in Melting Temperature

DIC – Differential Interference Contrast

DLS - Dynamic Light Scattering

Dm – *Drosophila melanogaster*

D_{max} - Maximum Dimension

dRI - differential Refractive Index

DSF - Differential Scanning Fluorimetry

DTT – Dithiothreitol

EcoRI – *Escherichia coli* RY13 type 1

EDTA – Ethylene diamine tetra-acetic acid

Eg5 – kinesin family 5 spindle protein

EGTA – Ethylene Glycol tetra-acetic acid

EM - Electron Microscopy

F/T - Flow-through

FF - Fast Flow

Fig - figure

FL - Full Length

GO - Graphene Oxide

GPU - Graphic Processing Unit

HDSD - Hit Discovery and Structural Design

HEPES – 4-(2-Hydroxyethyl) PiperazineEthaneSulfonic acid

His – Histidine

HPLC – High Performance Liquid Chromatography

HRV3C – 3C Protease from Human Rhinovirus

Hs - *Homo sapiens*

HSET - Human Spleen Embryo Testes

I (0) - Intensity at 0° scattering angle

I (q) - Intensity as a function of scattering vector

I/P - Input

ICR - Institute of Cancer Research

IEC - Ion Exchange Chromatography

IMAC - Immobilised Metal Affinity Chromatography

ISAC - Iterative Stable Alignment and Clustering

JANNI - Just Another Noise2Noise Implementation

Kan - kanamycin

Kar3 - Karyogamy 3

kDa - kilo Dalton

keV - kilo electron Volt

KHC – Kinesin Heavy Chain

KIF - Kinesin family

KLC – Kinesin Light Chain

KLP - Kinesin-Like Protein

KOD – *Thermococcus kodakaraensis*

L - Ladder

LB – Luria Bertani

LONCEM- London Consortium for cryo EM

LS - Light Scattering

MALS - Multi Angle Light Scattering

Mant – Methyl anthraniloyl

mAU - milli Absorbance Units

Mg - Magnesium

mg - milligrams

$MgCl_2$ - Magnesium Chloride

Mins - minutes

mL - milli litre

mM - milli molar

Mm - Mus musculus

MP - Mass Photometry

MPI – Message Passing Interface

MRC – Medical Research Council file format

MT - Microtubule

MW - Molecular Weight

MWCO - Molecular Weight Cut-Off

N-terminal – Amino terminal

NaCl - Sodium Chloride

Ncd-MT - non-claret disjunction

NdeI – *Neisseria denitrificans*

Ni^{2+} - Nickel ion

nm – nanometer

Nod – No distributive disjunction

NTA – Nitrilo-Triacetic Acid

Osm3 – Osmotic Avoidance abnormal protein 3

P(r) - Pairwise distance distribution function

PAGE - Poly Acrylamide Gel Electrophoresis

PAUP – Phylogenetic Analysis Using Parsimony

PCR - Polymerase Chain Reaction

PDB - Protein Data Bank

Pen - penicillin

pH - potential of Hydrogen ions

pI - Iso-electric pH

Pi – Inorganic Phosphate

PSIPRED – Position-Specific Iterative PREDiction

q - Scattering vector

Rad - radians

RELION - REgularised LIkelihood OptimisatioN

R_g - Radius of Gyration

R_h - hydrodynamic radius

Rn – *Rattus norvegicus*

Rpm - revolutions per minute

SAS – Small Angle Scattering

SAXS - Small Angle X-ray Scattering

S - Svedberg

Sc - *Saccharomyces cerevisiae*

SDS - Sodium Dodecyl Sulphate

SEC - Size Exclusion Chromatography

Secs - seconds

Sf9 – *Spodoptera frugiperda*

SOC – Super Optimal Broth with Catabolite Repression

Sp – *Strongylocentrotus purpuratus*

SPARX - Single Particle Analysis for Resolution eXtension

SPHIRE - SPARX for High Resolution Electron Microscopy

Strep - streptomycin

Strep II - Streptactin II

TCEP – Tris 2-Carboxy Ethyl Phosphine

TEM - Transmission Electron Microscopy

TEV – Tobacco Etch Virus

TIFF – Tag Image File Format

T_m - Melting Temperature

µg - micrograms

µL - micro litre

µm – micro metre

µM - micro molar

UV - Ultraviolet

Vis - Visible

VPP- Volta Phase Plate

W – wash

XCTK2 – *Xenopus laevis* C Terminal Kinesin 2

1. INTRODUCTION

1.1 The Puzzle of Intracellular Movement

Perhaps one of the most intriguing puzzles in the field of cell biology was centred around the phenomenon of movement inside a cell, i.e. intracellular motility. In a review from 1976, MA Bisby stated that the very geometry of a nerve cell makes it susceptible to problems in communication, maintenance, and distribution of materials (Bisby 1976). In 1975, the process that achieves these tasks- was first shown to involve particle movement towards the synapse as well as towards the cell body (Sjostrand and Frizell 1975). Over the next 20 years, this process was further investigated by employing radio-labelled macromolecules and crushed nerve endings (Droz, Koenig et al. 1973).

1.1.1 ATPases- The Mode of Transport

By the early 1980s, it was successfully established that large membranous organelles and vesicles exhibited saltatory motion with average velocities of a few hundred millimetres/day, not affected by any electrical activity in the cell (Grafstein and Forman 1980). In 1982, with the advent of the video-enhanced Contrast - differential interference-contrast (AVEC-DIC) microscopy, Robert T. Allen's team managed to gain a visual understanding of the process in the giant axon of a squid (Allen, Metuzals et al. 1982). They reported a variety of particles, moving in one of the two possible directions within the axon, at a range of average velocities that increased with decreasing size of the particles. The axonal transport of smaller particles (30-50 nm) that comprised small vesicles and dense membranous bodies, seemed to be a continuous process that reached speeds of up to 5 μm/s, in contrast to the slow, intermittent transport of medium-sized (0.2-0.6 μm) and large particles (0.8-5 μm) such as mitochondria.

Further exploration of the mechanism behind these movements required either a change in ionic conditions or the introduction of chemical agents into the cytoplasm, both of which were prevented by the presence of a plasma

membrane. Future studies sought to eliminate this obstacle either by working on the extruded axoplasm (Brady, Lasek et al. 1982) or by permeabilising the plasma membrane with detergents (Forman, Brown et al. 1983). In one such permeabilization experiment with the detergent saponin, David Forman's team observed a loss in axonal transport, which could subsequently be revived by the addition of 1mM ATP. This rescue was also observed with other nucleotides. However, titration studies showed a significantly lower requirement for ATP when compared to the other nucleotides, thus providing the first evidence that fast axonal transport is ATP-dependent. Forman confirmed these results by inhibiting axonal transport using vanadate, a known ATPase inhibitor. The rapid and reversible nature of the inhibition by vanadate was characteristic of a dynein or a dynein-like ATPase (Forman, Brown et al. 1983).

1.1.2 Microtubules- The Road for Transport

Although most studies focussed on the mode of intracellular transport, certain studies decided to take an alternative approach and focus on the road of transport instead. By 1985 it was well known that the vesicles were transported on cytoskeletal filaments (Schliwa 1984). While endeavouring to determine the exact nature of the filaments, BJ Schnapp noted that two vesicles transported in opposite directions on the same path crossed each other without colliding. This could only be possible, he reasoned, if there were several tracks longitudinal to each other around the circumference of the filament, i.e., a complex of several units. Using AVEC-DIC microscopy on the extruded axoplasm of a squid, he was able to determine the diameter of the filaments as approximately 20-27 nm. Substructures of the filament showed a linear array of subunits with a repeat every 5.5 nm in a row and up to 3 nm between rows. Furthermore, replicated transport filaments appeared splayed or split at the ends, which on further investigation revealed regular oblique striations on the inner surface, 4 nm apart. All these observations were indicative of microtubules (MTs). He confirmed this by an immunofluorescence assay using an anti-α tubulin antibody, which was found to bind a protein that co-migrated with α-tubulin on an SDS polyacrylamide

gel. Thus, it was ascertained that the fast axonal transport of vesicles is achieved by a force-generating dynein/dynein-like translocator on MTs. Another important observation that Schnapps made was that the protofilaments of the MTs have the same directionality. Hence, the direction of transport must be determined by the vesicle or the translocator itself (Schnapp, Vale et al. 1985).

1.1.3 The Translocators: Captured and Christened

Using this knowledge, Ronald D. Vale made a breakthrough in the field by developing a strategy to purify the translocator in question from squid optic lobes and a bovine brain. He adopted an affinity-based approach, whereby the protein co-sedimented with MTs. The translocator migrated in two bands- a larger 110-120 kDa heavy chain and a smaller 60-62 kDa light chain. Subsequent size exclusion chromatography showed a peak corresponding to an object of approximately 600 kDa, suggesting that the translocator was a multimeric protein composed of the heavy and light chain subunits (Vale, Reese et al. 1985). A few years later, the translocator was confirmed to be dimeric complex by a sedimentation equilibrium experiment (Kuznetsov, Vaisberg et al. 1988). Furthermore, the purification had been carried out in the presence of a non-hydrolysable ATP analog-AMPPNP, as the binding to MTs was lost upon adding ATP. This behaviour was the opposite to what would be expected from a dynein. Combining this with the clear difference between the molecular weight of the protein complex and that of dynein, Vale proposed that this was indeed an entirely new class of ATPase translocators. He coined the term “kinesin” for the protein, from the Greek “*kine*” meaning movement. His work set the precedent for the isolation of kinesins from different organisms and opened the door to the biochemical exploration of the translocation process (Vale, Reese et al. 1985).

Vale had also performed two types of motility experiments on the purified kinesin in the presence of ATP. In one of these, the kinesin was immobilised onto glass coverslips and brought about the movement of MTs over the surface of the coverslip. The other involved adsorbing kinesins onto carboxylated latex beads

and demonstrated the unidirectional movement of these beads on immobilised MTs, opposite to the movement of MTs in the previous experiment. Based on these observations, Vale's team reported that kinesin molecules successfully hydrolysed ATP to in the presence of MTs, despite showing basal levels of ATPase activity on their own. This result sparked a series of studies exploring the kinetics of kinesin's ATP hydrolysis reaction, which served as the first attempts at modelling the movement of the protein. Kuznetsov and Gelfand, for instance, confirmed that microtubule-binding stimulated the ATPase activity of the kinesin by a nearly sixty-fold increase, compared to basal activity levels (Kuznetsov and Gelfand 1986). Further kinetic studies undertaken by David Hackney in 1988 to dissect the steps of kinesin's ATPase reaction reported that the purified kinesin fractions showed a near absence of ATPase activity at low concentrations of ATP. Additionally, the binding constant of the kinesin-ADP complex was found to be less than 5 nM, which was greater than the affinity constant for ATP by a factor of two. Taken together, these two observations led Hackney to propose that purified kinesin might exist in the ADP-bound state and thus, ADP release could be the rate-limiting step of kinesin's ATPase reaction (Hackney 1988). This hypothesis was proven experimentally by a luciferase reporter assay that colorimetrically tested the presence of ATP in the purified extracts as well as a centrifugation-gel permeation chromatography to separate the free nucleotides from those that were bound to kinesin. In this manner, another step of the kinesin's mechanochemistry had been uncovered- the existence of bound ADP which inhibited its ATPase activity at lower concentrations of ATP and the absence of MTs.

1.1.4 The first look at the shape of kinesins

Vale's strategy of extracting and purifying kinesin from the cell opened the door to the first studies to visualise the structure of kinesin translocators, reported by Linda A. Amos in 1987. She managed to purify a kinesin from a pig's brain using Vale's protocol and imaged it in detail by negative staining electron microscopy. The protein appeared to be shaped like a flexible rod of about 100 nm in length,

with a diameter of 2-4 nm. One end of the rod had a tiny fork that was shown to bind MTs in the presence of AMP-PNP, while the other end was large and branched, to bind organelles or vesicles (Amos 1987).

Further insight into the domain architecture of kinesin was obtained a year later, when Joy T. Yang successfully isolated the gene coding for the kinesin heavy chain (KHC) in *Drosophila*. Using an *in vitro* translation system, Yang's team managed to synthesize the heavy chain peptide which had the same nucleotide-dependent microtubule binding ability as the intact kinesin complex. This observation suggested that the nucleotide and microtubule binding sites were located in the heavy chain. They then proceeded to locate these sites through proteolytic experiments. Changes in conformation produced by the ATP and microtubule binding events exposed different parts of the kinesin heavy chain to proteolysis, resulting in fragments of different sizes. Yang's team studied these fragments and were able to narrow down the binding sites to a 50 kDa domain near the amino terminal of the heavy chain (Yang, Saxton et al. 1988).

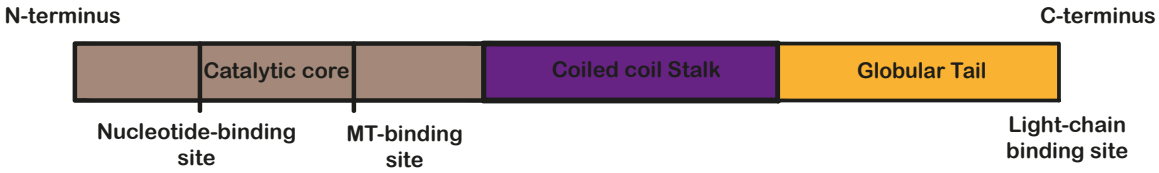


Figure 1.1 Domain Architecture of Kinesin Heavy Chain

Schematic showing the domains of the kinesin heavy chain and the locations of the binding sites based on experiments performed by Amos, L.A., 1987 and Yang, J.T. et al., 1988.

In conjunction with the sequence analysis and the electron microscopy performed by Amos, Yang's team delineated the entire primary structure of the kinesin heavy chain, as shown in fig 1.1. The chain consisted of a small globular N-terminal head domain of around 60 kDa that contained the ATP and microtubule binding

sites. Extending from the head was a 50-60 kDa alpha-helical stalk domain (fig 1.1) with a length of 56 nm, which fell well within the 50-70 nm range that had been measured by electron microscopy (Amos 1987). The alpha-helix was proposed to dimerize into a coiled coil configuration with the adjacent subunit to form the complex. The presence of proline and glycine residues towards the centre of the stalk made the region a flexible hinge, corroborating the findings of Amos in 1987. C-terminal to the stalk, the kinesin has a small 45 kDa tail domain at the C-terminal (fig 1.1), that was said to interact with the light chain to form branched structures that would hold vesicles that the protein was transporting (Yang, Laymon et al. 1989). Thus, in just a decade since its discovery, the translocator responsible for fast axonal transport had been isolated and its domain architecture had been characterized.

1.2 A Family of Kinesins- Where There's One, There's Many More

By the end of the 1980s, most of the studies on kinesins narrowed down to dissecting the mechanism of action of the KHC. However, these left many unanswered questions about the function of kinesins. When Linda Amos imaged porcine kinesin in 1987, she reported the existence of several cross-links between two MTs, suggesting that both ends of the protein bound to MTs (Amos 1987). This ability of kinesins to cross-link MTs had also been noticed by Ronald Vale a couple years earlier (Vale, Reese et al. 1985). Furthermore, Amos noticed that the cross-bridges differed in size, but the physiological relevance of these structures could not be ascertained. Interestingly, a plethora of similar cross-bridges had been observed five years earlier, before the first sighting of kinesins, in Nobutaka Hirokawa's electron micrographs (Hirokawa 1982). Evidently, the presence of a single type of kinesin couldn't justify the multiple cross-bridges that were observed. This demonstrated that intracellular transport is indeed highly sophisticated and indicated the potential existence of an entire family of kinesins within the cell. The search for the various sub-types of kinesins would create a branch of its own in the field of kinesin research.

1.2.1 A seemingly limitless display of diversity

The first two variants of kinesin were discovered in sea-urchin oocytes by two independent teams. Jonathan Scholey's team had managed to isolate a 134 kDa translocator whose behaviour during gel filtration, vanadate-mediated inhibition assays and *in vitro* motility assays were strikingly similar to that of the 110 kDa squid kinesin (Dinenberg, McIntosh et al. 1986). In addition, immuno-blotting experiments indicated that the anti-110 kDa antibodies recognised the 134 kDa protein, and the converse was true as well. All the evidence led Scholey's team to conclude that this translocator was indeed a kinesin. However, an immunoperoxidase-stained light microscopy experiment showed that the kinesins localized in the nucleus and bound to mitotic spindles. They concluded that this kinesin could be involved either in translocating chromosomes, moving proteins to and from the chromosomes, or sliding the mitotic spindles over each other. A

year later, Christine Collins and Richard Vallee discovered another type of kinesin in sea-urchin oocytes using sucrose gradient co-sedimentation studies. Apart from the presence of dynein and Scholey's kinesin, the sedimentation profile showed a peak corresponding to 10S, that increased sharply upon the addition of MTs. This protein showed an ATP dependence and yet was not inhibited by vanadate. Moreover, it remained bound to MTs despite the addition of ATP, indicating the existence of both ATP-dependent and independent modes of binding to MTs (Collins and Vallee 1986).

The advent of molecular genetics brought along a new method to identify potential kinesin family members by the complementation of mutant phenotypes that were observed in different organisms (Hartwell and Weinert 1989, O'Farrell, Edgar et al. 1989). In 1990, Annmarie Enos and Ronald Morris isolated a mutant strain of *A. nidulans* which was unable to carry out spindle-pole body segmentation and nuclear division. This was caused by a mutation in the *bimC4* gene, which encodes a protein with a sequence of 400 amino acids from the N-terminus that show an identity of 42% to KHC. Although the central domain of bimC4 was predicted to be alpha-helical, the absence of a heptad repeat implied that it wasn't a coiled coil. Considering the homology between the motor domain sequences and the absence of any similarity in the rest of the protein, Enos and Morris concluded that bimC4 could represent a sub-family of structurally different kinesins that slide mitotic spindles over each other (Enos and Morris 1990). In the same year, a mutation that disrupted nuclear division and delayed doubling of *S. cerevisiae* was mapped to the *kar3* gene by Pamela Meluh and Mark Rose. The Kar3 protein contained a C-terminal motor domain with 62% homology to KHC, contrary to the N-terminal motor of bimC4. In addition, Kar3 showed heptad repeats indicating a coiled coil structure for its central domain and had yet another unique cargo-binding domain that was shown to bind MTs, independent of ATP (Meluh and Rose 1990).

1.2.2 Initial attempts at kinesin classification

The first attempts at a kinesin classification system came from PCR screens of the *Drosophila* genome by two independent teams. Although incomplete, the two screens collectively managed to isolate 11 genes coding for unique kinesin-like proteins (KLPs), including homologs of previously reported kinesins. They classified them by assigning arbitrary numbers, usually based on the serial numbers of the primers used in the screens. However, to develop a more robust and consistent system, Isabelle Vernos's team decided to explore the functional basis for this diversity. Specifically, they were interested in whether the kinesins were expressed in a cell-type dependent basis, or if they were differentially regulated within a cell. They managed to identify four novel KLP transcripts in Xenopus oocytes, two of which were temporally regulated during embryogenesis. However, expression patterns couldn't achieve a successful segregation of these structurally and functionally distinct kinesin (Vernos, Heasman et al. 1993). An alternative attempt by Vale and Fletterick proposed a hierarchical approach to the classification. The first step was to classify them by the location of the motor domains in the sequence: N-terminal, C-terminal or inter domain. This was followed by a sub-categorisation based on the best-known features of the kinesin, i.e., oligomeric state, location within the cell, directionality of motion, and so on. While this did serve as a convenient initial basis for segregation, this convenience was its inherent flaw. The inconsistency of the sub-categories meant that that not all kinesins in one class would share the same properties and not all kinesins that share the same properties would be in the same class (Vale and Fletterick 1997).

The failures of each of these different approaches highlight the challenge that was at the core of the kinesin classification problem- the inconsistency in the information available for each newly discovered kinesin.

1.2.3 A promising solution to the consistency conundrum

A solution to this situation would involve a system that does not rely on structural and functional data, such as a phylogenetic analysis. Sharyn Endow, a pioneer in studying the diversity in the kinesin superfamily, arrived at the earliest consistent classification strategy. Endow's team analysed the amino acid sequences of the motor domains from the 34 different kinesins known at the time (Goodson, Kang et al. 1994). The sequences were used to construct phylogenetic trees with distance matrix algorithms such as CLUSTAL V (Higgins, Bleasby et al. 1992) and the maximum parsimony program PAUP (Wilgenbusch and Swofford 2003). This analysis resulted in five classes of kinesins:

a) The first class, known as the DmNcd (*Drosophila melanogaster* non-claret disjunction) class, comprised kinesins with motor domains at the C-terminal (or more simply, C-terminal kinesins) such as Ncd and Kar3. As both proteins were known to localise at the spindle poles, the other members in this class such as murine HSET were predicted to have the same properties.

b) The second class- MmKif2 (*Mus musculus* Kinesin family member 2) consisted of two central motor kinesins, namely MmKif2 and RnKrp2. Unlike the C-terminal kinesins, this class showed no similarities in their architecture or the amino acid sequence of the motor domain. The lack of information for this class of kinesins makes it difficult to establish their hallmarks, especially because they appear to share more similarities with kinesins in other classes than themselves. The fact that not all central motor kinesins are part of the same class confirms that the location of motor domains is not an evolutionarily consistent classification on its own.

c) The third class- AnBimC (*Aspergillus nidulans* Blocked in mitosis C) also comprised spindle-associated motors, akin to the DmNcd class. However, their cargo domains show no significant sequence similarities. Through mutant phenotype studies and antibody localisation experiments, it was established that this class of kinesins is required to maintain spindle bipolarity, hinting at a

molecular tug-of-war during spindle assembly, between the DmNcd and AnBimC classes of motors.

d) The fourth class contained kinesins that share a heterotrimeric state, quite unlike the conventional kinesin model of two heavy chains, two light chains and a coiled coil that mediates dimerization. For instance, the SpKRP85-KRP95 heterodimer (*Strongylocentrotus purpuratus* Kinesin Related Protein 85), can be isolated in complex with a third, non-kinesin protein (Cole, Chinn et al. 1993). This heterotrimer SpKinesinII, is thought to be the functional form and could potentially serve as a means of coordinating two different but related microtubule activities. Other members include *C.elegans* Osm3 and *M.musculus* Kif3.

e) The fifth class is known as the kinesin-1 class. Members of this class, such as DmNod are typically known to have less than 5% probability of possessing a coiled coil domain, as estimated by secondary structure prediction software. These kinesins are thought to function as monomers.

Although incomplete, this phylogenetic classification seemed more robust compared the previous attempts. Five years later, SA Endow and AJ Kim expanded the existing scheme to accommodate newly discovered kinesins. Rather than collapsing with the additional details, the model was able to assign the new kinesins into existing classes and in some cases, create new classes as well (Kim and Endow 2000).

1.2.4 A standardised nomenclature

The availability of a consistent and robust kinesin classification scheme for the kinesin superfamily certainly helped to define a standardised kinesin nomenclature. However, over the years, there were several independent attempts to override the customised names, each with their own minute discrepancies due to outlier kinesins (Hirokawa 1998, Kirchner, Woehlke et al. 1999). It was only in 2004, that a special interest subgroup of the American

Society of Cell Biology established a single, universal nomenclature (Lawrence, Dawe et al. 2004). The attendees successfully outlined a detailed list of rules for naming and classifying kinesins. A total of 14 kinesin families were defined, each being further divided into sub-families. Rather than using functionally derived names such as the human kinesin HSET, or prefixes like KLP (Kinesin-like Protein) or KIF (Kinesin Family), it was decided that the proteins would be referred to using the prefix 'Kinesin family'. For instance, the protein HSET mentioned above would be referred to as Kinesin family 14 member HSET. Any sub-classes would be depicted by letters following the family number, such as 'Kinesin family 14A'. This system has successfully avoided much of the confusion and redundancy surrounding the names of different kinesins and paved the way for a methodical incorporation of future members of existing families as well as new families (Lawrence, Dawe et al. 2004). As of today, sixteen years later, kinesins are still being discovered and named using the criteria listed in the standardised nomenclature of 2004.

1.3 A Divergence in the *Direction* of Kinesin Research

1.3.1 An opposing minority in the superfamily

The discovery and characterisation of the increasing number of kinesins highlighted the similarities between them that would reinforce the defining traits of this superfamily of motors. One of the most important properties that defined a kinesin was the direction of their motion, which was found to be towards the growing end i.e., plus-end of the MTs for most of the kinesins across different families (Dinenberg, McIntosh et al. 1986, Enos and Morris 1990, Vernos, Heasman et al. 1993). The fact that this opposed and thus balanced the role of the well-known minus-ended dynein family of MT-based translocators further satisfied the logic behind kinesins being defined as a family of plus-ended motor proteins. However, the 1990s saw the discovery of a minority in the superfamily that diverged from this norm. For instance, the claret locus in *Drosophila* was known to encode a product that played a key role in chromosomal segregation. Mutation studies indicated that the protein likely mediated chromosomal interaction with the spindle MTs, to prevent mitotic non-disjunction. Sequencing further revealed that this product of the claret non-disjunction locus showed homology to a kinesin motor domain, indicating that proper chromosomal segregation was dependent on the force generated by a kinesin (Endow, Henikoff et al. 1990). However, this theory implied that the kinesin possessed a trait that was not considered possible previously- a minus-ended polarity. Thus, the non-claret disjunction protein Ncd, was the founding member of a family of minus-ended kinesins- the kinesin-14 family (Goodson, Kang et al. 1994). The same year, Pamela Meluh and Michael Rose reported the identification of kar3p from *S. cerevisiae*, which was also involved in spindle-chromosome interactions (see section 1.2.1, *Page 7*). Four years later, kar3 was also found to be minus ended, similar to Ncd (Endow, Kang et al. 1994). It thus seemed that members of the kinesin family 14 were minus-end directed motors that localised within the nucleus. A detailed analysis of the structural and functional commonalities among the kinesin-14 members is presented below.

1.3.2 Domain architecture of the kinesin-14 family

The discovery of minus-ended kinesins like *Drosophila melanogaster* Ncd and *Saccharomyces cerevisiae* Kar3 paved the way for the discovery of homologs in different organisms, such as *Aspergillus nidulans* KlpA and *Homo sapiens* HSET, among others mentioned in table 1.1. The systematic classification of the kinesin superfamily in 2004 established this family of kinesins as kinesin family 14. The family is further divided into the subfamilies kinesin-14A and 14B (Lawrence, Dawe et al. 2004). While the mitotic kinesins such as Ncd, HSET and the others mentioned in table 1.1 belong to the kinesin-14A subfamily, the less-known kinesin-14B subfamily comprises members such as *Mus musculus* KIFC2, that are known to localise at late endosomes in the cytoplasm (Hirokawa and Noda 2008). Due to the unavailability of information about kinesin-14B members, I will describe the structural features and functional relevance of kinesin-14A. Table 1.1 shows the similarity between members of the kinesin-14A sub family by comparing the sequence identity shared by some kinesin-14A members and the founding member of the family, DmNcd.

Aside from being the only family of kinesins exhibiting minus-ended motion, kinesin-14s are also the only family with C-terminal motor domains (Hirokawa and Noda 2008) (fig 1.2). The motor domains are preceded by a sequence of heptad repeats, that are characteristic of a coiled coil and thus points to dimerization (fig 1.2). The heptad repeats are interrupted towards the N-terminus by a proline residue that typically kinks or breaks helices (Endow, Henikoff et al. 1990), thereby indicating the start of a third domain- the N-terminal tail that is rich in basic residues and presumably involved in binding cargo (fig 1.2). Despite the common domain architecture, kinesin-14 members show a significant degree of intra-class dissimilarity. As table 1.1 shows, the amino acid sequences in the motor domains of Ncd and Kar3 show less than 50% identity, while the other domains are only 17% identical. AnKlpA, the protein in the list that is most similar to Ncd is still only 20% identical to Ncd in its non-motor regions. Such a marked divergence is bound to lead to some functional distinctions as well. For instance, while proteins like Ncd were found to be homodimers, Kar3

was found to be heterodimeric, with the ability to bind to two candidate proteins-Cik1 or Vik1 to carry out its mechanical cycle (Manning, Barrett et al. 1999).

Kinesin-14A members		% Motor Identity	% Non-Motor Identity
DmNcd	AnKlpA	49.8	19.9
	SpoPkl1	48.8	15.8
	ScKar3	46	17.6
	HsHSET	43.9	20.8

Table 1.1 Sequence Homology in the Kinesin-14 Family

The % sequence identity between the sequences of the motor domains and non-motor domains of some kinesin-14A members with the respective domains of Ncd, obtained from Clustal Omega.

While the amino acid sequence of the domains varies significantly, their functions remain largely similar in all members of the family. For instance, the C-terminal motor domains of kinesin-14A members were shown to move MTs over glass coverslips in the presence of ATP (McDonald, Stewart et al. 1990, Endow, Kang et al. 1994), similar to the N-terminal motor domains of kinesin-1. Additionally, the basic N-terminal CBD of kinesin-14A was shown to interact with the acidic E-hook of tubulin on the MTs via electrostatic interactions (Furuta and Toyoshima 2008). In other words, kinesin-14A dimers are capable of binding MTs on both ends. This results in a phenomenon known as MT-bundling, first noted in 1990 (McDonald, Stewart et al. 1990). The existence of a minus-ended kinesin family that can interact with two layers of MTs is antagonistic to the kinesin-5 family whose members move towards the plus-end while making similar contacts with

MTs in mitotic spindles. This scenario would thus lead to a tug of war between kinesin-5 and kinesin-14 members alluded to earlier (see section 1.2.1, *Page* 7), that could play a key regulatory role in spindle maintenance.

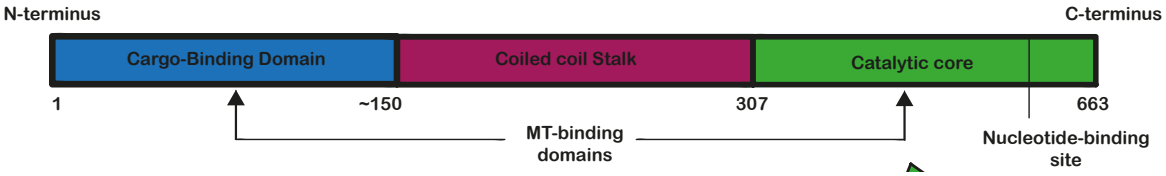


Figure 1.2 Domain Architecture of a Kinesin-14

Schematic representation of the domains that make up a typical kinesin-14A chain; C-terminal motor domain (green), central coiled coil dimerization domain i.e., coiled coil stalk (magenta) and N-terminal cargo-binding domain (blue).

Thus, in addition to the reversed directionality of motion, kinesin-14A members also showed a divergence in the position of the domains within the amino acid sequence and the type of cargo that they interacted with. Having established the domain architecture of the family, the focus of a large part of the kinesin community now turned towards determining the source of the reversal in directionality in the kinesin-14 motor domains, which showed a significant degree of homology to the plus-end directed kinesin-1 motor domains.

1.4 Decoding The Motion of a Kinesin-14

Perhaps the most intriguing feature of a kinesin motor protein is how the chemical reaction of ATP hydrolysis is coupled to the generation of mechanical output, that manifests as the transport of cargo within the cell. This cycle has been the subject of several biochemical and biophysical studies on kinesin-1 and kinesin-14A proteins. A decade after the discovery of the kinesin-1, Andrew Lockhart and Robert Cross proposed a model based on kinetic and mobility studies, to delineate the steps in its mechano-chemical cycle (Lockhart and Cross 1994). They suggested that one of the kinesin motor domains undergoes ADP release as it encounters a MT and establishes contact in this empty state. ATP-binding to the empty motor domain triggers a conformational change that locks the MT-bound motor in an inflexible state and generates a force to propel the unbound motor domain forward. This generation of forward thrust is referred to as the power stroke of a kinesin (Lockhart and Cross 1994). The binding of the motor domain in the front, post-power stroke, is accompanied by the ATP hydrolysis in the motor behind, setting it up for the next propulsion. The concerted binding and unbinding of the two motor domains ensure that at least one motor domain is MT-bound at any stage, a property that is referred to as the processivity of kinesin-1. Simply put, the motion of kinesin-1 can be regarded as walking on MTs. Lockhart and Cross also observed that the MT-bound modes of the empty motor domains in a plus-end directed kinesin-1 and a minus-end directed kinesin-14 were stereo-chemically indistinguishable i.e., the directionality could not be determined by visualising the motor-MT interaction. What could this imply about the power stroke and the processivity of a kinesin-14?

1.4.1 The processivity of kinesin-14

The first hint about the processivity of kinesin-14 was derived from the initial motility experiments that immobilised the N-terminal of Ncd to a glass slide to study the ATP-dependent movement of MTs that is caused by the C-terminal motor domains (Stewart, Semerjian et al. 1998). One of the key findings of this study was the existence of a minimum density of kinesin molecules for the diffusion of the MTs. Specifically, a density of 20 Ncd motor molecules/μm^2 or

less was insufficient to allow even the transient association of MTs to the Ncd in the presence of ATP. Binding was rescued by the replacement of ATP with its non-hydrolysable analog AMP-PNP, demonstrating that Ncd did indeed possess the ability to bind MTs, albeit reversibly. The apparent necessity of a threshold density of Ncd was accompanied by the observation that shorter MTs dissociated a more frequently than longer MTs on the addition of ATP. The threshold motor density, length-dependency and the reversible binding were crucial hints in arriving at the conclusion that the minus-ended kinesin Ncd is a non-processive motor (Stewart, Semerjian et al. 1998). The same non-processivity was also observed in Kar-3, and other Ncd homologs such as XCTK2 from *Xenopus laevis* (Walczak, Verma et al. 1997) and CHO2 in Chinese Hamster Ovary cells (Matuliene, Essner et al. 1999). In other words, the kinesin-14A subfamily appeared to possess yet another unique characteristic compared to the rest of the superfamily- the ability to "hop".

1.4.2 The power stroke of kinesin 14

The non-processive motion was also confirmed through stopped-flow measurements of the MT-dependent ATP hydrolysis of Ncd, which showed a stoichiometric pre-steady state burst, as opposed to the super-stoichiometric burst that is expected for a processive motor protein (Foster and Gilbert 2000). Such stopped-flow kinetic experiments that used fluorescent analogs of nucleotides such as methyl anthraniloyl ATP (mantATP) as well as equilibrium sedimentation were the most common approaches to characterise the multistep enzymatic cycle of a kinesin (Foster, Correia et al. 1998). In addition, single molecule studies proved extremely useful in characterising motor proteins in terms of the step sizes, periodicity and processivity of their motion. In one such experiment, a three-bead suspended MT was engineered by optical trapping to study the stepping mechanism of an immobilised Ncd motor (deCastro, Fondecave et al. 2000). The results showed that the MT was displaced from its mean position twice on the addition of ATP- the first was a combination of axial and lateral displacement of around 2-3 nm, while the following displacement was

axial with a magnitude of 6 nm after which the MT immediately returned to mean position. The initial tug was proposed to be the establishment of the motor-MT contact following the loss of ADP, which generates a mild torque as the empty motor domain locks on to the MT forming a rigor state. The second displacement, larger in magnitude and axial in direction, was thought to signal a power stroke which led to motor release from the MT based on the return of the MT to mean position. Reduction of ATP concentrations stalled the rigor state for longer durations, implying that unlike a kinesin-1 that generates power from ATP-binding, the power stroke of a kinesin-14 was generated during ATP hydrolysis (deCastro, Fondecave et al. 2000).

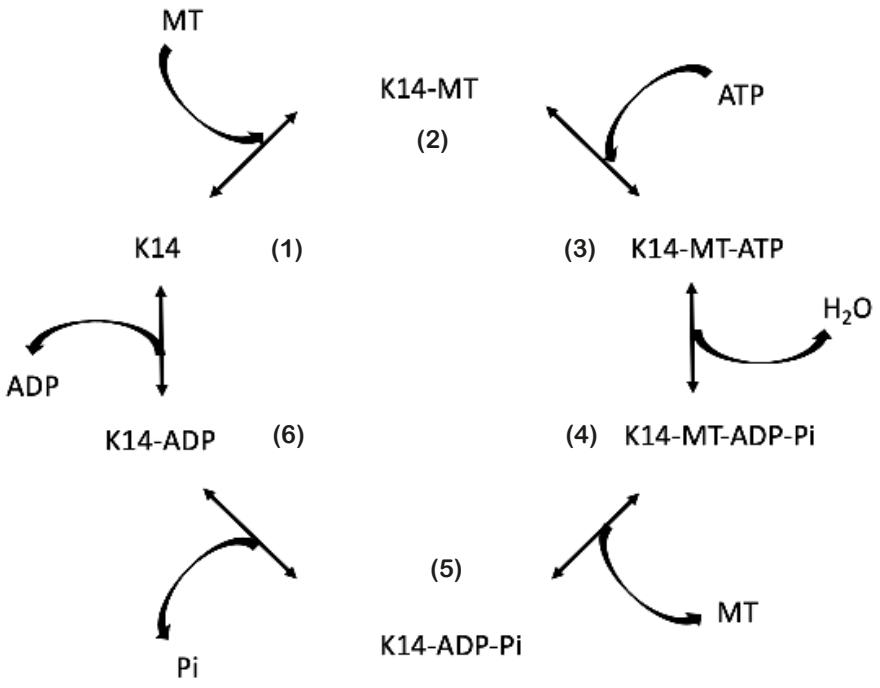


Figure 1.3 Reaction Mechanism of a Kinesin-14A

Pathway depicting the steps of the enzymatic reaction of a kinesin-14 on MTs, as deduced by kinetic studies (Foster and Gilbert 2000). The reaction is initiated by (1) an Apo kinesin-14A approaching an MT and forming (2) an Apo kinesin-MT complex; this is followed by (3) ATP-binding to the MT-bound kinesin-14A motor domains; ATP hydrolysis results in the formation of (4) ADP-Pi which triggers (5) detachment of the kinesin-14A from the MT followed by (6) the loss of Pi to result in ADP-bound kinesin-14A; the loss of ADP is the rate-limiting process that would regenerate Apo kinesin-14A and restart the cycle.

This was supported by the independent observation by Kelly Foster's group that Ncd sedimented with MTs in the presence of all nucleotides except Mg-ADP-Pi, which was the product of ATP hydrolysis. Moreover, stop flow measurements of MT-Ncd interactions with ATP or AMP-PNP revealed that ATP hydrolysis was necessary to promote the detachment of Ncd from the MTs, thereby establishing the biochemical basis for the kinesin-14A power stroke (Foster, Correia et al. 1998). The enzymatic cycle of a kinesin-14A could thus be summarized as shown in fig 1.3.

1.4.3 A tale of two motor domains

A missing piece of the puzzle that could not be resolved from the above studies was the presence of the second motor domain of the Ncd dimer. This suggested that the two motor domains could either interact simultaneously with the MTs or take turns to generate the power stroke (deCastro, Fondecave et al. 2000). A key result that provided the first hint towards understanding this issue was obtained in the sedimentation experiment of Ncd-MT in the presence of Mg-ADP-Pi. While dissociation of Ncd from MT was observed, the dissociation seemed to increase as a non-Michaelis Menten sigmoidal function of the concentration of inorganic phosphate (Pi) (Foster, Correia et al. 1998). This curve was characteristic of cooperative binding and was thus the first evidence of the second motor domain's involvement in the mechanical cycle (Foster, Correia et al. 1998). Subsequently, ADP-release was shown to be the rate-limiting step of the entire reaction (Foster and Gilbert 2000).

However, another observation about the ADP-release step would be made just a year later, that would provide a solution to the question about the second motor domain. In 2001, it was observed that the ADP release showed signs of being a two-step process. The first step was an exponential decay followed by a second significantly slower, linear decay. The second one could be unambiguously identified as the familiar rate-limiting ADP-release (Foster, Mackey et al. 2001). Considering the cooperativity between the motor domains,

the first step would also point to an ADP-release from a motor domain. The near 700-fold difference in the rates of the two steps had the astounding implication that one of the motor domains had a weaker affinity for ADP than the other. In short, Kelly Foster's team reported that kinesin-14A members are asymmetric homodimers (Foster, Mackey et al. 2001). This is where kinetic studies reached their limit. In spite of the chemically generated outline of the reaction, the mechanics of force generation was still an unsolved mystery. While biophysical studies could quantitatively estimate the force being generated, they could not provide an explanation about how force generation occurred. Further completing the story regarding the mechanism of kinesin-14s required the ability to visualize them.

1.4.4 The mechanism of kinesin-14A in visual snapshots

The first attempts at visualisation came in the form of crystal structures of the motor domain of ADP-bound monomeric Ncd at a resolution of 2.5 Å (Sablin, Kull et al. 1996). The organization of the secondary structure elements of Ncd closely resembled that of the kinesin-1 motor domain (Kull, Sablin et al. 1996), whose amino acid sequence shares a 40% identity to that of Ncd. The striking similarity extended to the MT-binding site, which suggested identical mechanisms for MT-motor interactions in both families, thereby supporting the indistinguishable MT-bound forms of both motors (Lockhart and Cross 1994). The structure of dimeric Ncd comprising the motor domain and a portion of the stalk was subsequently solved to a resolution of 2.5 Å a couple of years later, and showed a twofold symmetry about the axis of the stalk (Sablin, Case et al. 1998). Further information could be obtained from structures that provided the context of MT-binding, which could not be achieved by crystallisation. This created a paradigm shift in the focus of structural studies on kinesin-14A towards cryo EM.

The first cryo EM structure of MT-bound Ncd was solved by Keiko Hirose alongside Andrew Lockhart and Rob Cross in 1995, in the presence of AMP-PNP (Hirose, Lockhart et al. 1995). The map showed density at a resolution of 10 Å

and could be interpreted by fitting the crystal structure of Ncd. The fit showed that the motor domain that was not bound to the MT, pointed towards Ncd's direction of motion- the minus-end. A similar MT-bound structure of kinesin-1 revealed the detached motor domain pointing towards the plus-end. On the other hand, the neck-linkers were shown to be tilted backward, pointing towards the plus-end in the case of Ncd and minus-end in the case of kinesin-1 (Hirose, Lockhart et al. 1995). This observation suggested a level-arm mechanism for the neck linker, further supported by single molecule studies that demonstrated a change in the angle between the stalk and the motors accompanying the power stroke. Soon, additional nucleotide-bound and apo intermediates of the Ncd cycle were solved, sampling more of the lever-arm's rotation (Wendt, Volkmann et al. 2002, Endres, Yoshioka et al. 2006) to enable a better understanding of the power stroke.

Moreover, the visualisation of different binding modes of the motor domain confirmed several observations that were reported in kinetic studies and answered questions that couldn't be resolved with biophysical measurements. For instance, a study that used unidirectional metal shadowing showed that Kar3 and Vik1, the two homologous subunits of heterodimer bind to adjacent protofilaments of the MT in the presence of ADP (Rank, Chen et al. 2012). Considering that ADP-release from Kar3 is a rate limiting step as well as the fact that Vik1 exhibits nucleotide-independent binding to MTs, one could deduce that the first collision with the MT was through Vik1. The Kar3 motor domain soon follows by binding to an adjacent protofilament, accompanied by ADP-release. The subsequent ATP hydrolysis in the Kar3 motor domain triggers the release of Vik1, which now pointed towards the minus end with the neck domain pointing towards the plus end. The release of inorganic phosphate (Pi) leads to the power stroke, where the neck domain rotates 90° towards the minus end and the Kar3 motor is finally detached (Chen, Porche et al. 2012). This model could be extrapolated to Ncd, where the motor domain with the weaker affinity to ADP plays the role of Vik1 and initiates first collision with the MT (Step-2, fig 1.4) (Zhang, Dai et al. 2015). This important feature of asymmetry between the motor domains also explains the time lag the binding of the second motor following rate-limiting ADP release (Step-3, fig 1.4) and the detachment of the first motor following ATP hydrolysis (Step-5, fig 1.4).

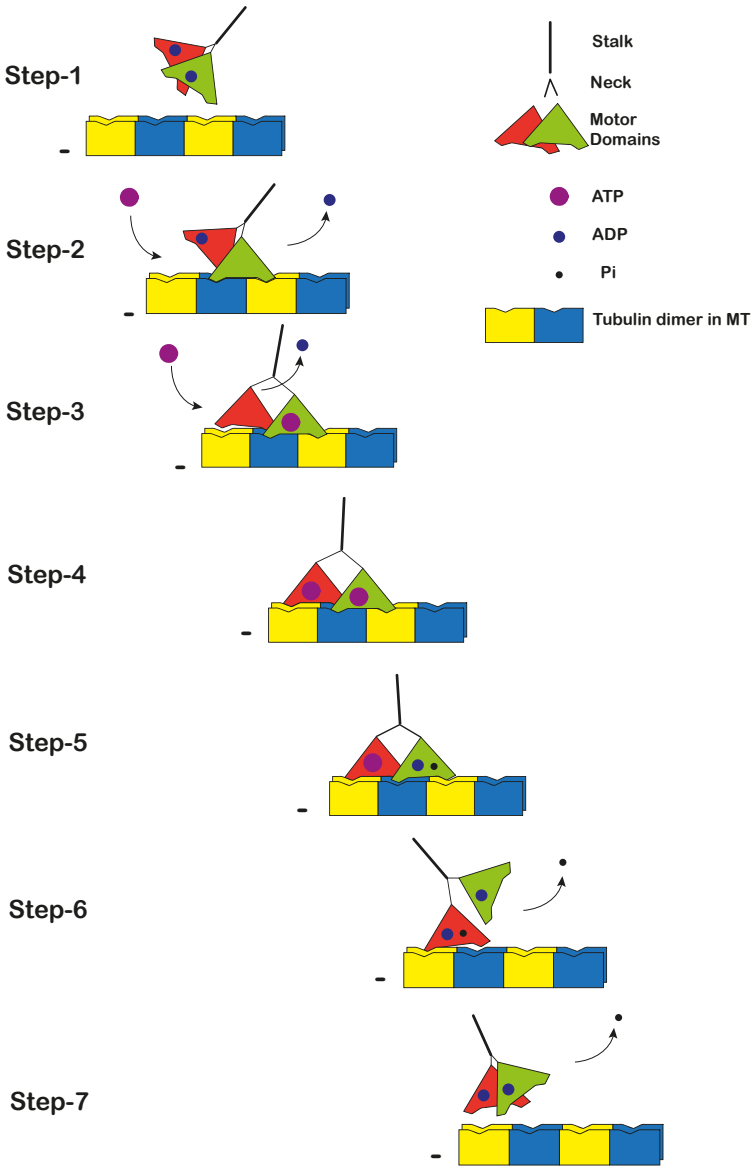


Figure 1.4 The mechanochemical cycle of kinesin-14A

Schematic depiction of the non-processive mechanism of a homodimeric kinesin-14A (Zhang, Dai et al. 2015), showing transition states that represent every binding mode of kinesin-14A on MTs, obtained through cryo EM and kinetic studies; both motor domains are depicted in different colours to highlight the asymmetry in the homodimer. The cycle begins with (Step-1) ADP-bound kinesin-14A approaching MTs; (Step-2) the subsequent loss of ADP from one of the motor domains initiates contact of the empty motor domain with the MT; (Step-3) ATP binding to the MT-bound empty motor triggers a conformational change that promotes ADP-loss and contact initiation between the second motor domain and the MT; (Step-4) both motor domains are ATP-bound and attached to the MT on adjacent protofilaments; (Step-5) one of the motor domains undergoes ATP hydrolysis to form ADP and Pi and (Step-6) detaches from the MTs with the liberation of the inorganic phosphate (Pi) while the second motor undergoes hydrolysis; (Step-7) both motor domains exit the MTs bound to ADP.

The great degree of similarity between the mechanisms of heterodimeric Kar3-Vik1 and homodimeric Ncd reveals yet another property that ties the kinesin-14 family together- an evolutionarily conserved power stroke mechanism (Zhang, Dai et al. 2015). Thus, from a concerted effort of several biochemical, biophysical, and structural studies spanning two decades, the mechanochemical cycle of kinesin-14 on MT was proposed to occur as depicted in the schematic diagram in fig 1.4.

1.5 Biological Significance of kinesin-14A

The previous section discussed the mechanism underlying the biochemical function of the kinesin-14A subfamily. As mentioned in section 1.3.2 (*Page 14*), kinesin-14A members use the force generated by their motion to bundle microtubules in the mitotic spindle. What is the impact of the bundling on the processes of nuclear and cellular division? What are the circumstances that dictate the role of the kinesin-14A motors in the cell? These questions about the biological significance of the kinesin-14A subfamily will be addressed in this section.

The absence of kinesin-14A members in cells contributed to a similar phenotype in the different organisms they were isolated from- a failure to segregate chromosomes effectively during mitosis or meiosis (Endow, Henikoff et al. 1990, Meluh and Rose 1990, O'Connell, Meluh et al. 1993). This is especially interesting when one considers the plethora of motor proteins such as dynein (Rieder and Alexander 1990) and other kinesin families (Theurkauf and Hawley 1992, Maney, Hunter et al. 1998), that are known to coordinate the movement of chromosomes in metaphase and anaphase (Brunet and Vernos 2001). Among the several mitotic motors, kinesin-14A members are known to play a role in specific cellular contexts that are dictated by an organelle in the cell- the centrosome.

1.5.1 Centrosomes and kinesin-14

Centrosomes are non-membranous organelles (1-2 µm in diameter) within the nucleus, that primarily function as sites of nucleation for MTs. In other words, they are MT organising centers (MTOCs) (Fukasawa 2005) that direct the formation of spindles during mitosis. Typically, cells contain single centrosomes which duplicate during the S phase of the cell cycle and migrate to opposite ends of the nucleus during mitosis to form the spindle poles (Stearns 2001). The mitotic spindles tether to these poles on one end (minus-end) while gradually polymerising on the other end (plus-end) until they contact the chromosomes at the center of the nucleus. This facilitates an arrangement where the

chromosomes are pulled apart towards the two spindle poles, which is followed by cytokinesis. Thus, a bipolar segregation of chromosomes is critical to the accurate distribution of genetic material between the two daughter cells and is tightly regulated.

The dysregulation of centrosome numbers within the nucleus leads to the existence of more than two spindle poles during mitosis- a phenomenon known as Centrosome Amplification (CA) (Brinkley 2001). This phenomenon can cause multipolar mitosis, which would result in genetic instability in the daughter cells, rendering them non-viable (Ghadimi, Sackett et al. 2000, D'Assoro, Lingle et al. 2002). Interestingly, Ncd was shown to be essential for the viability of *Drosophila* cells that exhibited centrosome amplification (CA) (Skold, Komma et al. 2005). Despite the presence of another minus-end directed motor like dynein, Ncd mutants that were CA- positive showed a high degree of centrosomal mobility, indicating a possible role of Ncd in clustering centrosomes (Kwon, Godinho et al. 2008). CA positive cells had been known to resort to centrosomal clustering that restructures the multipolar geometry of spindles and effectively results in a bipolar mitosis (Quintyne, Reing et al. 2005). A genome-wide RNAi screen revealed that the kinesin-14 gene was crucial to centrosome clustering, while dynein was not. Expression patterns in *R. norvegicus*, *S. pombe*, *S. cerevisiae* and humans also revealed an elevation in kinesin-14 expression during mitosis with too many or too few centrosomes (Kwon, Godinho et al. 2008). Thus, while seemingly redundant in a normal cellular environment undergoing bipolar mitosis, kinesin-14s are important for the survival of cells exhibiting an anomaly in centrosome count.

1.5.2 HSET: a kinesin-14 that promotes the sustenance of cancer

Genetic instability is a characteristic of most cancers (Yao and Dai 2014). One of the leading factors that is proposed to destabilise the cancer genome is CA (Brinkley 2001, D'Assoro, Lingle et al. 2002). Several types of cancers including adenocarcinomas, triple negative breast cancer and ovarian cancers are

characterized by the presence of more than two centrosomes (Landen, Lin et al. 2007, Pannu, Mittal et al. 2015). Similar to the other cell types discussed earlier, CA-positive cancer cells also exhibit the ability to bypass multipolar mitosis by clustering their centrosomes (Basto, Brunk et al. 2008). In fact, the phenomenon of centrosomal clustering was first observed in a cancer cell line forty years ago (Ring, Hubble et al. 1982). Cancer cells use the Spindle Assembly Checkpoint (SAC) to delay the progression of the cell cycle into mitosis, triggering a cascade that ultimately leads to the recruitment of the kinesin-14A member KIFC1 (She and Yang 2017) for the clustering process. Originally discovered in the human spleen, embryo, and testes, KIFC1 is commonly referred to as HSET (Ando, Kikuti et al. 1994).

Due to its ability to cluster the spindle poles by crosslinking spindle MTs, HSET is known to have a significant effect on spindle morphology (She and Yang 2017). For instance, the loss of HSET reportedly results broad spindles and unfocussed poles (She and Yang 2017). This is because the absence of the force generated by HSET leads to the prevalence of the forces exerted by other motors that move opposite to the direction of HSET on the MTs. The primary example of a motor that acts antagonistically to HSET on the mitotic spindles is the plus-end directed kinesin family 5 (Goodson, Kang et al. 1994). Like kinesin-14s, kinesin-5 members have also been shown to crosslink spindle MTs in vitro (Kashina, Baskin et al. 1996). However, they differ from the homo or heterodimeric arrangements adopted by kinesin-14A proteins. Kinesin-5 family members adopt a homo-tetrameric dumbbell arrangement that allows them to interact with MTs through the pairs of motor domains on their opposite ends (Sawin, Mitchison et al. 1992). This generates an outward force to move the layers of parallel MTs apart and thereby counteracts the inward focussing action of the kinesin-14 (Hentrich and Surrey 2010). The absence of kinesin-14 HSET results in an inability to focus spindle poles and form asters (a radiating array of microtubules arising from a centrosome in a dividing cell), while the cells that show an absence of kinesin-5 Eg5 exhibit poor pole separation (She and Yang 2017). Simply put, aster formation by parallel MTs at spindle poles is the result of the tug of war between the two kinesin families in question. The balance in the interplay between HSET and Eg5 has also been shown to play a role in regulating the

assembly, organisation, and cross-linking of antiparallel mitotic spindles in the prometaphase (Tao, Mogilner et al. 2006, Fink, Hajdo et al. 2009).

In addition to its interaction with MTs, HSET has also been found to interact with unpolymerized tubulin *in vivo* (Hentrich and Surrey 2010) to form multimeric clusters that show a high degree of processivity (Reinemann, Norris et al. 2018). These heterogeneous clusters of HSET around soluble tubulin are known as asters and are evidence for the role of tubulin levels in the regulation of HSET's activity within the cell.

1.5.3 HSET in Anti-Cancer Drug Discovery

As discussed above, the biological significance of kinesin-14A members occupies opposite ends of a spectrum from being functionally redundant in cells that contain two centrosomes to being crucial for the survival of cells that exhibit CA (Park, Ma et al. 2017). In the case of HSET, this presents a valuable opportunity for therapeutic intervention. In 2012, Julia Kleylein-Sohn's team conducted an siRNA-based knock down of HSET in CA-positive cancer cells and non-cancerous cells. The cancer cells were unable to resolve the multipolarity of mitotic spindles and died, while the non-cancerous cells were remained unaffected (Kleylein-Sohn, Pollinger et al. 2012). This further highlighted the potential of HSET as a target for drug discovery.

Over the course of the last decade, several HSET inhibitors have been identified within and outside the ICR based on high throughput biochemical and biophysical screens (Wu, Mikule et al. 2013, Myers and Collins 2016, Zhang, Zhai et al. 2016). The inhibitors are designed to target the motor domain, aiming to arrest HSET’s mechanochemical cycle and thereby preventing it from clustering centrosomes (Saint-Dizier, Matthews et al. 2023). The pharmaceutical company AstraZeneca reported the HSET inhibitor AZ82, that is proposed to target the MT-motor domain interaction allosterically (Wu, Mikule et al. 2013). However, the detailed mode of action of any of these inhibitors is largely unknown due to the

absence of HSET-inhibitor structures. Multiple groups including Rob van Montfort's Hit Discovery and Structural Design (HDSD) Team at the ICR have attempted to crystallise inhibitor bound HSET motor domains but have been largely unsuccessful (Park, Ma et al. 2017). The only available structure of HSET is that of the ADP-bound motor domain, solved to a resolution of 1.9 Å by Sharyn Endow's team in 2017 (Park, Ma et al. 2017). This largely resembles the crystal structure of Ncd (Sablin, Kull et al. 1996) but lacks the neck linker. Structures of HSET in the unbound (Apo) form or bound to ATP analogs and MTs have also not been reported to date. Moreover, the poor conservation of amino acid sequence in the coiled coil stalk and N-terminal CBD of HSET as mentioned in table 1.1 implies that the generation of full-length structures through homology modelling are also not very reliable, despite the advent of deep-learning based tools like Alphafold (Jumper, Evans et al. 2021). A structural understanding of HSET inhibition and comparisons with the structures of various nucleotide-bound transition states of HSET could help to accelerate the drug discovery programmes.

1.6 The Aims of My PhD- a Future for HSET Drug Discovery

As mentioned in the previous section, the dearth of HSET structures in its different states hinders the structural understanding of its molecular mechanism and of small molecule inhibition of HSET. In section 1.4.4 (*Page 21*) structural biology has been shown to play a pivotal role in deciphering the mechanochemical cycle of kinesin-14s. A full-length structure of the HSET dimer would provide clearer insight into the binding modes attained by HSET in different nucleotide or compound-bound states. Moreover, structures of HSET-MT complexes would build a complete picture of the mechanism of HSET inhibition.

With a high throughput and the ability to yield structures of nearly atomic resolutions (< 2Å), X-ray crystallography would be the ideal candidate to obtain structures of the different states of HSET. However, attempts at crystallising full-length HSET and obtaining compound-bound structures of its motor domain have been consistently unsuccessful. While kinesin-14 dimers have been successfully crystallized previously (Kozielski, De Bonis et al. 1999), the non-native state of the protein during crystallization may create structural features that are not entirely representative of physiological conditions. A key example of this is the symmetric appearance of the motor domains which were predicted to be asymmetric. Moreover, crystallography may not be suitable for large macromolecular complexes, such as an HSET-MT complex.

1.6.1 Cryo EM- the solution to the HSET conundrum

In contrast, cryo EM samples are aqueous solutions of proteins that are rapidly frozen to preserve their native state in suspended animation (Taylor and Glaeser 1974, Dubochet, Adrian et al. 1988). Due to this reason, one can use cryo EM to obtain structures of full-length targets, flexible targets, nucleotide-bound targets, and large complexes. Therefore, cryo EM constitutes a complementary approach to obtain the information beyond the limits of X-ray crystallography. As is the case with EM maps prior to 2014, high resolution crystal structures were required for interpretation of the features in the density. However, since 2014, single particle cryo EM has undergone a rapid improvement in resolving power due to the

invention of sophisticated direct electron detectors that provide an improved signal to noise ratio particularly at high spatial frequencies. In addition, they can capture data as multi-frame movies rather than still images with a long exposure (Kuhlbrandt 2014). The long exposures were responsible for beam-induced motion in the sample, which resulted in loss of information in cryo EM data in previous decades. The high sensitivity of the cameras today enables the collection of short exposures, and the movie format provides a means to weight the frames according to the dose to retain the frames showing minimal radiation damage to the sample. This "resolution-revolution" (Kuhlbrandt 2014) accompanied by further advances in hardware over less than a decade have resulted in EM maps reaching resolutions as high as 1.7 Å, rivalling the resolution of crystal structures (Nakane, Kotecha et al. 2020). Over the last decade, the number of EM-based high-resolution structures being solved has exponentially increased, which includes numerous structures of MTs decorated with kinesin motor domains (Benoit, Asenjo et al. 2018, Cook, Roberts et al. 2021). However, despite the ability to obtain high resolution structures, there has been no report of a full-length structure from the kinesin superfamily, until very recently. Structural studies on kinesins have always focussed on the power stroke at the site of catalysis i.e., the motor domains. A structural insight into the mechanism of cargo-binding is lacking for most kinesin families. This is especially important for kinesin-14s as the interactions of the cargo-binding domain render the non-processive proteins processive and cross-link MTs.

1.6.2 The aims of this project

Therefore, the goal of my PhD project is to obtain a structural understanding of the inter-domain crosstalk that drives the mechanical cycle of parts of the kinesin KIFC1 (HSET). In order to do so, I plan to determine the structures of full-length HSET in its Apo and ADP-bound forms using single particle cryo EM, aided by biophysical characterisation. Specifically, I plan to answer the following questions:

1. What is the oligomeric state of FL HSET and how does it change on ADP-binding?
2. How does ADP-binding affect the structural stability of FL HSET?
3. Does either form of HSET exist in more than one conformational state?
4. What is the structure of the N-terminal cargo-binding domain?

In the long run, the direction of this project would proceed towards establishing a cryo EM-driven drug discovery pipeline for HSET at the ICR by the structural analysis of the following:

1. Inhibitor-bound FL HSET
2. MT-bound FL HSET (Apo and ATP analogs)
3. Inhibitor-FL HSET-MT complex

2. MATERIALS AND METHODS

2.1 Protein Production and Purification

2.1.1 FL HSET Constructs

The construct for His_6-FLHSET was a gift to the ICR from Cancer Research Technologies, now known as Cancer Research Horizons (London, UK). The construct was generated by inserting a PCR amplicon encoding residues 2 to 673 of human HSET into the pFastBacHTb vector (Thermo Scientific, Waltham, USA) downstream of the sequence encoding the N-terminal His_6 tag and TEV protease site.

For the $Strep_2$-FLHSET construct, a PCR amplicon encoding residues 2 to 673 of human HSET was inserted into pFastBac vector that had been modified to encode an N-terminal double StrepII tag followed by a HRV3C protease site (courtesy of Dr Catarina Rodrigues). PCR was performed using the His_6-FLHSET construct as a template and KOD Hot start polymerase (Merck, Darmstadt, Germany). Primers were designed to include fifteen nucleotide overhangs in accordance with the guidelines for In-Fusion cloning (Takara Bio, Boston, USA), and were based on the sequence of pFastBac following restriction digestion by the enzymes NdeI and EcoRI. Primers were purchased from Eurofins MWG Operon, Ebersberg, Germany. The primer sequences can be found in Supplementary Information S1C (*Page 200*). The thermal cycling protocol for the PCR was as follows: 15 cycles at 95 °C for 20 seconds, 65 °C for 10 seconds, and 70 °C for 41 seconds, using the ProFlex PCR System (Life Technologies, Thermo Fisher Scientific). The PCR amplicon was purified using the QIAquick® PCR Purification Kit (QIAGEN, Hilden, Germany). The pFastBac was prepared by digestion with NdeI & EcoRI enzymes (New England Biolabs, Ipswich, USA) and gel purification on a 1% w/v Agarose gel in 1x Tris-Acetate-EDTA buffer (pH 8.0). The corresponding band was excised and purified using the QIAquick® Gel Extraction Kit (QIAGEN). The purified amplicon was inserted into the digested vector following the In-fusion HD cloning protocol (Takara Bio, Boston, USA). 10

µl of the In-fusion products was incubated with 90 µl of RapidTrans™ TAM1 competent *E. coli* (Active Motif, La Hulpe, Belgium) in ice for 15 minutes following which the cells were heat-shocked at 42 °C for 45 seconds. The cells were subsequently cooled on ice, resuspended in 300 µl of SOC media, and incubated in a shaker culture at 37 °C and 240 rpm for an hour to recover. 50 µl of the culture was plated onto LB-Agar plates supplemented with 50 µg/ml ampicillin and incubated overnight at 37 °C. Single colonies were used to inoculate 5 ml cultures of LB broth containing 50 µg/ml ampicillin, which were grown overnight at 37 °C and 240 rpm. The plasmid DNA was isolated from the cultures using QIAprep spin MiniPrep kit (QIAGEN) and sequenced using the forward and reverse PCR primers at Source Biosciences to identify the bacterial colony that contained the successful construct.

2.1.2 Protein Expression

All insect cell cultures were performed at 27 °C using Sf9 cells (Invitrogen, Waltham, USA) and Sf-900™ III SFM supplemented with 30 µg/ml Penicillin and 50 µg/ml Streptomycin (Gibco, Waltham, USA). Shaking cultures were agitated at 120 rpm.

The in-fusion constructs were used to generate recombinant baculoviruses according to the Bac-to-Bac Baculovirus Expression System (Invitrogen, Waltham, USA). Transfection supernatants containing recombinant virus were titrated by infecting small (50 ml) shaker cultures with 3, 10 or 30 µl virus per 10^7 cells, and measuring cell growth & viability at 24-, 48- and 72-hours post infection. Subsequently, large scale cultures (400 ml) were infected with 20 µl virus which was found to be sufficient to cause growth cessation within 24 hours. Cultures were harvested 72 hours post infection by centrifuging at 4,000 rpm for 20 minutes using a JLA8100 rotor (Beckmann, Brea, USA) and cell pellets were stored at -80 °C. The His_6-FLHSET construct was expressed by Dr Craig McAndrew.

2.1.3 Protein Purification

Purifications of all expressed constructs were carried out with the following procedures on an ÄKTA Pure (GE Healthcare Life Sciences, Buckinghamshire, UK) at 4°C, using UV absorbance at 280 nm to monitor protein elution. All chromatography columns were from GE Healthcare Life Sciences, and the columns for Size Exclusion Chromatography (SEC) were calibrated (Supplementary fig S2, *Page 201*) using Bio-Rad's Gel Filtration Standard (Bio-Rad Laboratories, California, USA). Eluted fractions were analysed by SDS-PAGE on NuPAGE® Novex 12% Bis-Tris gels (Thermo Fisher) using a SeeBlue™ 2 Pre-stained Protein Standard (Invitrogen) and subsequent staining by InstantBlue™ Protein stain (Sigma Aldrich). The concentration of the protein samples was estimated using the NanoDrop UV-Vis Spectrophotometer.

To prepare lysates, cell pellets were re-suspended in 60 ml of lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP and 5% v/v Glycerol), containing 1x complete Ultra EDTA-free protease inhibitor tablet (Roche, Welwyn Garden City, UK), 1 mM $MgCl_2$ and 12.5 U/ml of Benzonase nuclease (Millipore Ltd., Hertfordshire, UK) and lysed by sonication (18 cycles of 5 seconds on, 55 seconds off, at 50% Amplitude). Lysed cells were then centrifuged at 21,500 rpm for 55 minutes at 4°C, using a Beckmann-JA25.50 rotor, to separate the insoluble material. The supernatant was then filtered through a 1.2 µm syringe filter prior to loading onto the Affinity Chromatography (AC) column.

2.1.3.1 His-tagged FL HSET

Filtered lysate containing His_6-FLHSET was loaded onto a 1 ml HisTrap™ Fast Flow Immobilised Metal-ion Affinity Column (IMAC) pre-equilibrated in IMAC buffer A (20 mM HEPES, 500 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol at pH 7.5) at 1 ml/min. Unbound components were washed off with 5 column volumes (CV) each of 8% and 16% IMAC buffer B (IMAC Buffer A containing 250 mM imidazole), and HSET was subsequently eluted with 5 CV of 100% IMAC buffer B. Peak fractions were pooled and concentrated to a final volume of 500 µl using 30,000 MWCO Pierce™ PES protein concentrator (Pierce Biotechnology, Waltham, USA). Concentrated fractions were loaded onto a Superdex 200

10/300 Size Exclusion Chromatography (SEC) column pre-equilibrated in running buffer (20 mM HEPES, 200 mM NaCl, 1 mM TCEP and 5% v/v Glycerol), at 0.5 ml/min. Peak fractions were pooled and subsequently loaded onto a 1 ml Mono-S Ion Exchange Chromatography (IEC) column at 1.25 ml/min after equilibration in IEC buffer A (20 mM HEPES, 1 mM TCEP, 5% v/v Glycerol). HSET was eluted using a continuous gradient of 20% to 60% IEC buffer B (IEC Buffer A + 1M NaCl) over 20 CV. Purified fractions were selected by SDS PAGE and pooled, concentrated to 1 mg/ml and flash frozen for storage at -80 °C.

2.1.3.2 2xStrep-tagged FL HSET

In the case of $Strep_2$-FLHSET, filtered lysates were loaded onto a 5 ml StrepTrap HP pre-equilibrated in Strep buffer A (20 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% v/v Glycerol) at 1 ml/min, followed by a 5 CV wash at 3 ml/min in the same buffer and elution using 5 CV Strep-buffer B (buffer A with 3 mM Desthiobiotin) at 2 ml/min. Eluted fractions were pooled and concentrated to a volume of 500 µl as previously explained prior to SEC. SEC was initially performed using an Superdex 200 10/300 column before changing to a Superose 6 Increase 10/300 GL which gave better resolution.

To obtain Apo HSET, concentrated HSET was used with no added nucleotide, and the Apo-SEC running buffer constituted 20 mM HEPES pH 7.5, 400 mM NaCl, 1 mM TCEP and 5% v/v Glycerol. To obtain HSET bound to ADP, ADP and $MgCl_2$ were added to a final concentration of 0.5 mM and the ADP-SEC running buffer comprised of 20 mM HEPES pH 7.5, 200 mM NaCl and 0.05 mM Mg-ADP. Samples were loaded and run at 0.5 ml/min. The eluted fractions were analysed by SDS-PAGE and stored at 4 °C.

2.2 Biochemical and Biophysical Characterisation

2.2.1 ATP Hydrolysis Assay

The ATP hydrolysis activity of HSET was studied using the ADP-Glo™ Max Assay system (Promega, Southampton, UK) which was performed by Dr. Harry Saville in a PROXIPLATE 384 PLUS WHITE plate (PerkinElmer, Buckinghamshire, UK). Each well contained a total assay volume of 20 µL made up of 5 µL reaction components (HSET, Microtubules and ATP), 5 µL ADP Glo Reagent and 10 µL Kinase Detection Reagent. Concentrations of His_6-FLHSET and $Strep_2$-FLHSET-I in the 5 µL reaction mixture were varied from 0.8 nM to 100 nM in two-fold increments in a buffer composed of 40 mM HEPES pH 6.8, 20 mM $MgCl_2$, 0.5 mM EGTA, 0.2 mM Triton X-100 and 2 mM DTT. ATP hydrolysis was carried out in the presence of 100 µM ATP and 0.07 mg/ml microtubules (Cytoskeleton Inc, Universal Biologicals). The reaction was stopped by the addition of ADP-Glo reagent in separate wells every 10 minutes for an hour. Fluorescence was observed by the addition of Kinase Detection reagent and measured using an Envision 2013 MultiLabel Reader (PerkinElmer) with an excitation wavelength of 340 nm and emission wavelength of 450 nm.

2.2.2 Differential Scanning Fluorimetry (DSF)

The ability of FL HSET to bind ADP was studied by label-free Differential Scanning Fluorimetry, also referred to as nanoDSF, using the Prometheus NT.48 (NanoTemper, München, Germany). 25 µM each of the pre-purified Apo and ADP-bound forms of $Strep_2$-FLHSET-II were loaded into separate nanoDSF Grade Standard Capillaries (NanoTemper) that were subjected to a temperature increase in the Prometheus from 20°C to 95°C over 70 minutes, while monitoring and recording the intensities of fluorescence at 330 nm and 350 nm.

2.2.3 Size Exclusion Chromatography- Multi Angle Light Scattering (SEC-MALS)

The SEC-MALS of Apo- and ADP-bound $Strep_2$-FLHSET-II was performed using an Agilent HPLC system coupled with a mini-DAWN Helios Light Scattering detector, measuring the diffusion of the light generated by a laser of wavelength 658 nm, and an Optilab Refractive Index detector (Wyatt Technologies, Santa Barbara, USA). Each sample of 100 µl at a concentration of 22 µM was separated by a Superose 6 Increase 5/15 GL column (GE Healthcare) at a flow rate of 0.25 ml/min. The running buffers contained 20 mM HEPES buffer at pH 7.5 and 200 mM NaCl, with an additional 1 mM TCEP and 5% (v/v) Glycerol for Apo-HSET and 0.1 mM each of $MgCl_2$ and ADP for ADP-HSET. The data was analysed by the adjoining ASTRA v7.3 software to calculate the molar mass at every position of the SEC chromatograms for both forms of HSET.

2.2.4 Dynamic Light Scattering (DLS)

Aggregation tendencies of $Strep_2$-FLHSET-II were monitored by DLS using the SpectroLight 600 DLS system (Xtal Concepts GmbH, Hamburg, Germany). 2 µl of Apo-HSET or ADP-HSET at concentrations of 1 mg/ml each were loaded under paraffin oil in a 96-well Vapour batch plate (Douglas Instruments, Hungerford, UK). Scattering profiles were collected at 10 and 20 °C, with the plate equilibrated for ten minutes at each set temperature prior to acquiring measurements. The scattering data was processed by the Xtal Concepts instrument software.

2.2.5 Mass Photometry

Mass photometry measurements were carried out using the Refyn One^{MP} system (Refyn Ltd., Oxford, UK). Samples were added to wells that were formed by placing a 3mm x 1mm silicone casket (Grace Bio-Labs, Bend, USA) on top of a

Corning 24 x 50 mm glass coverslip (Corning, Inc. New York, USA). Buffer (13 µl) was loaded into each well to focus the sample and optimise illumination settings. This was followed by the addition of 2 µl of sample that had been diluted to 200 nM. Binding of the sample to the surface of the coverslip was observed and recorded for 60 seconds using the Acquire^{MP} software and around 900-1800 binding events were collected per movie. Data was processed to obtain a distribution of binding event counts in Discover^{MP} software. Counts were converted to molar mass (kDa) using the calibration that was performed on standards provided by Refyn, Inc. The different dilution buffers tested contained 20 mM HEPES at pH 7.5, differing only in their salt content as follows: buffer-1 (200 mM NaCl), buffer-2 (100 mM NaCl), buffer-3 (50 mM NaCl).

2.2.6 Small Angle X-ray Scattering (SAXS)

All measurements were conducted on the B21 beamline at Diamond Light Source (Didcot, UK), which generates 13 keV X-rays of wavelength 0.95 Å. The scattered photons were collected within a scattering vector (q) range of 0.0026-0.34 Å^{-1}, on a detector that was positioned 3.7 m away from the sample. Apo and ADP-HSET samples of 50 µl each, that were concentrated to 3.0 mg/ml in Apo-SEC and ADP-SEC buffers respectively, were flown through a Shodex™ KW-404 SEC column (Shodex™, Munich, Germany) mounted on an Agilent 1200 HPLC, prior to entering a quartz capillary for X-ray exposure. The samples in the capillary were imaged at 4 °C, and a total of 619 exposures of one second each were collected for each form of HSET.

The scattering data from each exposure were imported into the B21 beamline's custom software ScÅtter-3 to view the overall scattering profiles along the SEC chromatogram as a plot of intensity "log I(q)" vs scattering vector "q", from which the signal corresponding to the buffer was subtracted. For each separate exposure within the SEC peak of each HSET form, the Guinier region of the scattering curve (range of scattering data corresponding to low-q values) was carefully inspected to select the part of the peak giving consistent values for

the radius of gyration (R_g). The scattering curves for the selected exposures were then scaled and averaged. The averaged data were then imported into Primus from the ATSAS v3.2.1 suite (EMBL, Hamburg, Germany) (Manalastas-Cantos, Konarev et al. 2021) for further analyses. First, the radius of gyration (R_g) and scattering intensity at angle zero (I(0)) of the samples was determined using the Guinier analysis. The I(0) value was then used to calculate the molecular weight of the sample in a similar way to SEC-MALS. Subsequently, the scattering profile was used to calculate a dimensionless Kratky plot (normalised to I(0) and R_g) to assess the degree of structure of the two forms of HSET (Durand, Vives et al. 2010). This was followed by the computation of an Indirect Fourier transform (IFT) to obtain a pair-wise distance distribution "P(r)" function that was adjusted to ensure a smooth drop-off of the function at the maximum dimension "D_{max}" of the molecule (Konig, Svergun et al. 1992). Dummy atom 3D models were computed from the P(r) function using ten rounds of DAMIF software, followed by averaging using DAMAVER and refinement by DAMIN software to generate a single reconstruction for each form of HSET (Franke and Svergun 2009).

2.3 Negative Stain Electron Microscopy

2.3.1 Sample Preparation

Negative staining was performed on 300-mesh Copper Quantifoil grids- R1.2/1.3 (SPT Labtech Ltd, Hertfordshire, UK) coated with a thin layer of carbon support film. Grids were glow-discharged at 0.26 mBar for 60 seconds using PELCO easiGlow™ (Ted Pella, Inc., Redding, USA) prior to sample application. The HSET sample (3 µl) was applied onto the glow-discharged grids and incubated for 60 seconds at room temperature. The grids were subsequently washed twice in distilled water and then stained twice in 2% uranyl acetate such that the second staining step lasted 30 seconds. Each step was accompanied by manual blotting with a Whatman™ No. 1 filter paper and the stained grid was air-dried before loading into the microscope. Negative stain EM was used for the purpose of characterisation of the different purification strategies of HSET as well as structural analysis of Apo and ADP-HSET. For characterisation experiments, Apo-HSET peak fractions from the different purification strategies were diluted to a concentration of 0.2 µM (0.016 mg/ml) in the respective SEC buffers, prior to application on the grids. For structural analysis, samples of Apo and ADP-bound $Strep_2$-FLHSET-II were each diluted in a two-step process as follows: buffer-A up to a concentration of 0.7 µM (0.05 mg/ml) followed by buffer-EM (20 mM HEPES pH 7.5, 100 mM NaCl) to a final concentration of 0.2 µM.

2.3.2 Data Collection

The negative stain grids were imaged on the FEI (Thermofisher Scientific, Waltham, USA) Tecnai F20 microscope at the Division of Structural Biology, the Institute of Cancer Research, Fulham Road, equipped with a field emission gun operating at 200kV, and a TVIPS F416 CMOS detector (TVIPS GmbH, Gauting, Germany). Images were collected 1 to 2 µm under focus at a magnification of 50,000x (corresponding to a pixel size of 1.73 Å). Images were acquired with a dose rate of 35-45 electrons per $Å^2$ with a 1-second exposure. Data was collected

as images in TIFF format manually using EM-Menu software (TVIPS) for the sample characterisation, while the collection was automated for structural analysis, using the EM Tools module.

2.3.3 Data Processing

2.3.3.1 Sample Characterisation

TIFF images from the TF20 were converted to MRC format using the EM2EM tool in the IMAGICv110817 processing suite (van Heel, Harauz et al. 1996). The MRC micrographs were then imported into RELION v3.0 (Zivanov, Nakane et al. 2018) for manual particle picking. Micrographs were low-pass filtered to 20 Å for manual-picking. Particles were picked with a diameter of 300 Å from each dataset and extracted with a box-size of 240 pixels which corresponds to 420 Å. The extracted particles were subjected to un-masked reference-free 2D classification in RELION-3.0 without correction of the Contrast Transfer Function (CTF), using default parameters for image alignment and a 20 Å limit in E-step resolution. The particles were also classified using ISAC2 from the SPHIRE v1.3 processing suite (Yang, Fang et al. 2012) with a diameter of 300 Å and a pixel-error-threshold of 2.1 into classes containing 60-100 particles each.

2.3.3.2 Structural Analysis

Datasets for Apo and ADP-HSET were converted from TIFF to MRC format using the EM2EM tool in the IMAGIC processing suite. The converted micrographs were imported into RELION v3.1 to estimate CTF parameters using CTFFind v4.1.8 (Rohou and Grigorieff 2015). Particles were auto-picked with a diameter of 300 Å and a picking threshold of 0.3 in CRYOLO v 1.5.6 (Wagner, Merino et al. 2019) by a general picking model that was trained using a part of the dataset constituting 200 manually picked particles from a selection of twelve micrographs low-pass filtered in CRYOLO to a cut-off of 0.1. The picked coordinates were imported into RELION to extract the particles from the micrographs with a box size of 440 Å after down sampling by a factor of two. The extracted particles stack was subsequently used to test classifications in cryoSPARC v 3.0 (Punjani,

Rubinstein et al. 2017), ISAC2 and RELION v3.1 as outlined below. In cryoSPARC, the particles were classified unmasked into 50 classes over the course of 20 online-EM iterations that marginalized over poses and shifts, while the ISAC2 classification utilised the parameters described in section 2.3.3.1 (*Page 42*). In the case of RELION, a range of values were tested for the alignment parameters before applying the optimal values to the entire stack of particles. Ab-initio models for each dataset were generated with C1 symmetry using initial and final subsets of 200 and 1000 particles respectively by applying 15º angular sampling and a 6-pixel translational search range in 2-pixel search steps in RELION. For 3D alignment, particles from successful 2D classes were classified in 3D into a single class using the ab-initio reference with 7.5º angular sampling and 5-pixel translational search range in single pixel steps. The 3D-aligned particles were subsequently reclassified in 2D without applying any in-plane image alignment to preserve the 3D alignments. Particles from successful 2D classes were subjected to further classification in 3D into two classes using the ab-initio reference and custom mask with the alignment parameters described above. The 3D class averages were refined with their respective stacks of particles and custom soft-edge masks in cryoSPARC v3.0's Homogeneous Refinement Legacy tool without windowing or masking the dataset (Punjani, Rubinstein et al. 2017). All 3D maps were viewed and analysed in UCSF Chimera (Pettersen, Goddard et al. 2004). To estimate the resolution, the refined models were low pass filtered to a series of thresholds in IMAGIC.

2.4 Cryo Electron Microscopy

2.4.1 Sample Preparation

A sample of 4 µl of Apo or ADP-HSET was applied onto glow-discharged grids in a FEI Vitrobot Mark IV (Thermo Fisher) which was maintained at 100% humidity and temperatures of 4 °C or 18 °C. The sample was allowed to incubate on the grid for a defined waiting period which was followed by blotting with specified settings of blot force and blot time, and then plunged into liquid ethane cooled by liquid nitrogen. A range of physical parameters such as temperature, blot force, blot time and wait time as well as chemical parameters such as concentration of protein and buffer composition were tested on the grids. The following types of unsupported grids were used: 300-mesh Ultra-Au-Foil and Cu/Au Quantifoil (SPT Labtech, Ltd.), and 400-mesh C-flat 4-Cu-T-50 (Protofilaments, Inc. Morrisville, USA); all grids were of dimensions R1.2/1.3. Support films on grids such as amorphous carbon, Graphene Oxide and Molecular Graphene (Electron Microscopy Sciences, Hatfield, USA) were also trialled.

2.4.2 Data Collection

2.4.2.1 Glacios

Initial screening and data collection efforts on cryo-grids were carried out on the FEI Glacios TEM at the Division of Structural Biology, the Institute of Cancer Research, Fulham Road, equipped with a Falcon 4i direct electron detector. Movies were recorded in EER format with 183 frames at 0.5 to 0.7 µm under focus using a dose of 60 electrons per $Å^2$ and an exposure time of 5.73 seconds per movie. Imaging was performed at a magnification of 150,000x which corresponds to a pixel size of 0.94 Å using a Volta Phase Plate (VPP) whose activation time and periodicity of exposures was optimised. The data collection was automated, aided by the software EPU provided by Thermo Fisher Scientific.

2.4.2.2 Titan Krios

Another batch of cryo-grids was imaged on the FEI Titan Krios at the London cryo EM consortium LONCEM at the Francis Crick Institute, equipped with a BioQuantum K3 direct detector and an in-column Energy Filter (Gatan, Inc., Pleasanton, USA). Three exposures were acquired for every foil hole at 0.5 µm under focus with a dose rate of 70 electrons per $Å^2$ and a total exposure time of 2.67 seconds. Movies contained 53 frames and were acquired in TIFF format in super-resolution mode at a magnification of 130,000x which corresponded to a pixel size of 0.335 Å. Imaging was performed using a VPP collecting 97-102 exposures per position of the VPP. The collection was carried out using EPU software.

2.4.3 Data Processing

2.4.3.1 Pre-Processing

Movies from the Glacios and Krios datasets were independently imported into RELION v3.1.2 where the beam induced motion in every frame was corrected using 4 by 4 patches and a B-factor of 150 in RELION's implementation of MotionCorr2. The Glacios data was grouped into fractions containing 30-50 frames each and used unbinned for the motion correction while the Krios data was binned by a factor of 2 for the same. The corrected frames were dose-weighted and combined into MRC micrographs, and CTF parameters were subsequently estimated by CTFFind v4.1.14. Micrographs were selected based on the following parameters - defocus range of -0.2 to -1.2 µm, phase shifts between 0.3 and 0.7π rad and an estimated resolution greater than 10 Å. Additionally, in the case of the Glacios dataset, the selection of micrographs was manually curated based on the power spectra to eliminate any whose CTF was incorrectly estimated by the program.

2.4.3.2 Particle-Picking

Particle picking was performed in CRYOLO v1.8.4. A small subset of 20 micrographs from the Glacios and 36 micrographs from the Krios were used to

manually pick 200 and 400 particles constituting the respective training sets. In each case, the images were denoised by JANNI (Wagner, Merino et al. 2019) to train a picking model iteratively. The trained model was subsequently used to pick particles from all micrographs with a threshold of 0.3 and a picking diameter of 250 Å. Default and customised denoising models were tested on the Krios dataset. The picked coordinates were imported into RELION where the particles were down-sampled and extracted from the CTF-based selection of micrographs with a box size of 300 Å in both datasets for further processing. The datasets were down sampled by a factor of 2 for the Glacios data and 3 for the Krios data, bringing the respective pixel sizes to 1.88 Å and 2.01 Å.

2.4.3.3 2D and 3D Analysis

The extracted particles were subjected to multiple rounds of reference-free 2D classifications by iteratively selecting the best classes from each round to reclassify in the next round. Following this, particles from both datasets were subjected to classification with the application of a circular mask of 200 Å diameter, while the Glacios dataset also tested the application of high pass filters to varying cut-off values. The ab-initio 3D reference was generated in C1 symmetry with a small subset of 20,000 particles from the stack using the default 15 º angular sampling and 6-pixel translational search range in 2-pixel steps with initial and final particle subsets of 200 and 1000 for both datasets. The reference was used to classify particles in 3D with a suitable custom soft-edge mask and alignment parameters of 3.7 º angular sampling, 10-pixel search range, and 1-pixel steps for the Glacios datasets. Different values of the parameters for alignment and classification were tested during the 3D classification of the Krios dataset and the optimised parameters were used to refine the 3D class that showed the highest level of detail using RELION's 3D Refine. The 3D classes from the Glacios data were imported into cryoSPARC along with their respective masks and subsets of particles to be refined using the Homogeneous Refinement Legacy algorithm, without windowing or masking the dataset.

3. RESULTS

3.1 Production and Purification of FL HSET

The biophysical and structural characterization of FL HSET required a significant quantity of protein at a high level of purity. Thus, the first goal of my project was to develop a reproducible strategy to produce and purify FL HSET at the desired quality. I designed constructs that I cloned and expressed using the Baculovirus-Sf9 expression system described in section 2.1.2 (*Page 34*). The expressed protein was purified using a sequence of chromatography steps. Each chromatography step was based on a unique property to separate the proteins in the lysate and were used in the following order: Affinity Chromatography (affinity) -> Size Exclusion Chromatography (size) -> Ion Exchange Chromatography (charge).

3.1.1 His-tagged FL HSET

The first construct I tested is shown in fig 3.1 and will be denoted as His_6-FLHSET. His_6-FLHSET was first purified using a HisTrap FF IMAC column, followed by SEC with a Superdex 200 10/300 GL column and lastly, IEC using a Mono-S column (fig 3.2).

Figure 3.1 Construct for His_6-FLHSET

The first construct used to express and purify recombinant FL HSET was designed with an N-terminal Hexa-Histidine tag followed by a TEV protease cleavage site.

3.1.1.1 Step-1: IMAC using HisTrap column

The HisTrap column is composed of a Ni^{2+}-NTA matrix which has a high affinity for the imidazole ring in the amino acid Histidine (Bornhorst and Falke 2000). For this reason, a buffer supplemented with 250 mM imidazole was used to bind competitively to the column and elute bound protein, as shown in the chromatogram (fig 3.2 a). The affinity chromatography proceeded in three phases: binding, washing, and elution. After loading His_6-FLHSET, the column was washed with 8% and 16% buffer that corresponds to 20 mM and 40 mM imidazole (W1 & W2 respectively) before elution with buffer containing 250 mM imidazole. The SDS-PAGE analysis in fig 3.2 a showed that both washes contained a considerable number of contaminants whereas His_6-FLHSET was only detected in the elution fractions. Although His_6-FLHSET was the predominant band, there were still substantial contaminants at this stage. Eluates A1 to A4 were pooled & concentrated for the next step- SEC.

3.1.1.2 Step-2: SEC using Superdex 200

Unlike IMAC which is based on preferential binding affinity (Schmitt, Hess et al. 1993), the SEC columns use a porous matrix of spherical particles to separate a mixture of proteins based on their ability to pass through the inter-particle or intra-particle spaces. This thus relates to the dimensions of the proteins, more specifically expressed by the hydrodynamic radius i.e., a combination of size, shape, and hydration state (Lindqvist and Storgårds, 1955). In a run that lasts for a single CV, the SEC resolves components of the injected sample with the smaller species eluting nearer to the end of the run. Using a standard mixture of proteins of known molecular masses, the SEC can be calibrated such that the elution volume of a protein can be used to roughly estimate its molecular weight (Kostanski, Keller et al. 2004).

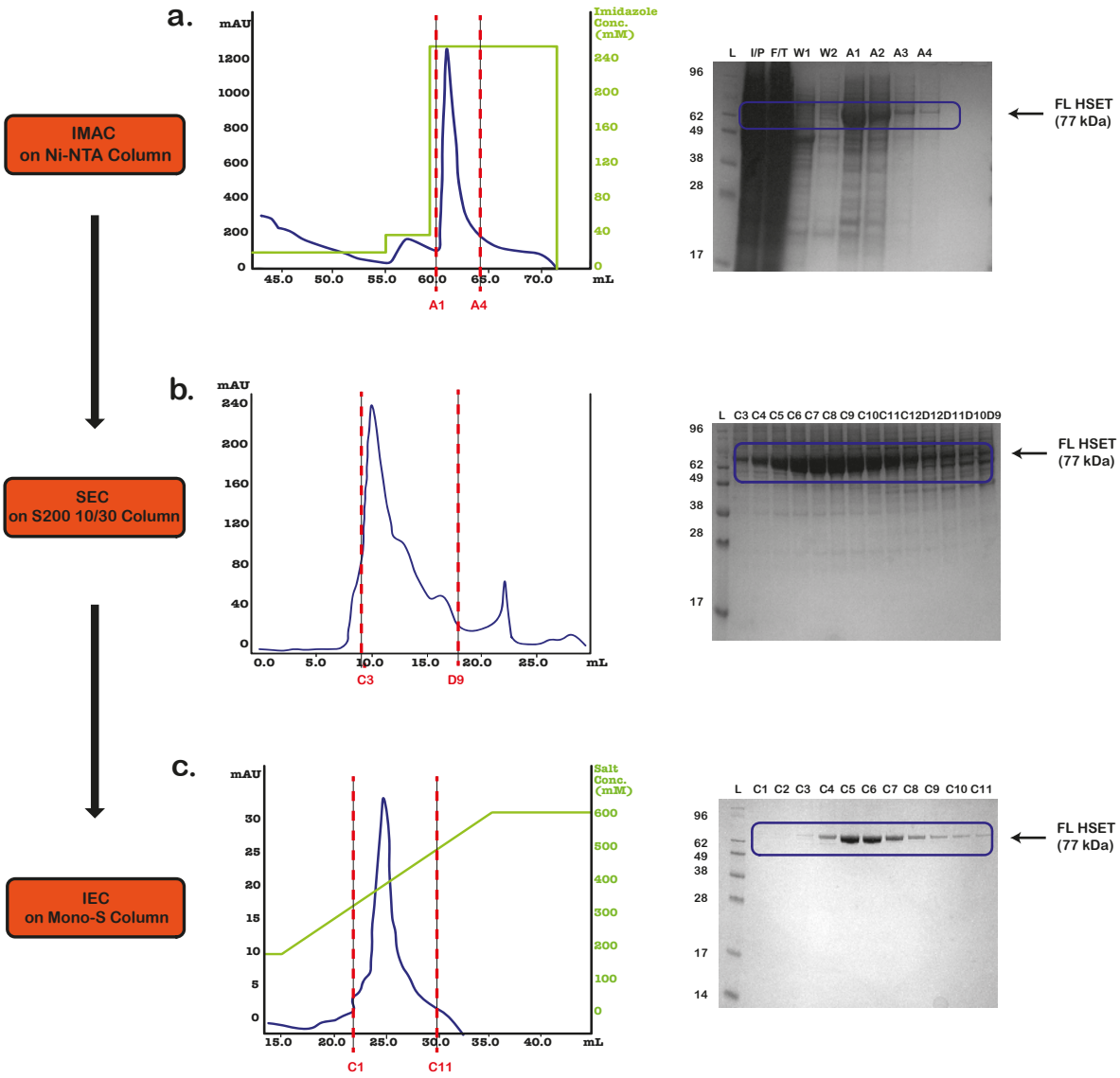


Figure 3.2 Purification of His-tagged FL HSET

His_6-FLHSET was purified as follows- (a) IMAC on a Ni2+-NTA column (1 ml HisTrap™ Fast Flow), (b) SEC through a Superdex 200 10/300 column and (c) an IEC using a 1 ml Mono-S column. The left panel shows chromatograms for each step, absorbance in mAU (blue line) is shown on the left- Y-axis and where applicable, the respective eluant concentrations in mM (green line) is shown on the right-Y-axis, both are plotted against the volume in ml along the X-axis. Please note that the chromatogram for the Ni2+-NTA & Mono-S columns have been cropped and do not include the loading phase. The eluted fractions were analysed by SDS-PAGE and are shown on the right panel; the red dashed lines on the chromatograms depict where eluted fractions were loaded onto the gel. L- denotes the molecular weight ladder, I/P- Input (injected sample), F/T- flow-through, W1 and W2- washes.

The SEC profile for His_6-FLHSET showed a large peak at a volume of 9.5 ml that begins eluting around the void volume of the S200 column (8 ml) followed by a shoulder peaking around 12 ml and two smaller peaks at around 18 and 23 ml. SDS-PAGE analysis indicated that the peak elution of His_6-FLHSET occurs around 10.5 ml which corresponds to the region in between the void and the shoulder peaks. Based upon calibrations established in the laboratory (Supplementary fig S2, *Page 201*), the peak of His_6-FLHSET has an estimated molecular weight (MW) between 400 and 500 kDa, which would be consistent with a hexameric species. Similarly, the large initial peak is consistent with very high MW species (>600 kDa) and can be referred to as the aggregate peak.

Although SEC improved the level of purity of His_6-FLHSET compared to the HisTrap eluates, it was still not pure enough for EM studies and required further purification. Furthermore, the elution profile of His_6-FLHSET overlaps significantly with the initial aggregate peak which appears to contain a non-protein contaminant judging by the discrepancy between large mAU signal and lack of protein seen on SDS-PAGE. To improve the purity of His_6-FLHSET, fractions C3 to C12 were selected for IEC.

3.1.1.3 Step-3: IEC using cation-exchange

IEC separates proteins based on their net charge in the pH of the buffer being used (Adhikari, Manthena et al. 2010). The predicted isoelectric point (pI) of His_6-FLHSET is 9.1 (Supplementary Information S2B, *Page 201*) (Gasteiger, Gattiker et al. 2003) which means that it will have an overall positive charge in buffers with a pH below the pI. Based upon a predicted positive charge, a negatively charged cation IEC column was chosen. The binding & wash phases (not shown) were performed in the presence of 200 mM NaCl (20% buffer B). The starting concentration was chosen to match the salt content in the injected sample. A continuous gradient of NaCl from 200 - 600 mM was subsequently used for elution and a single peak of protein eluted between 300 mM and 450 mM NaCl. The gel reflects the broadness of the HSET peak from fraction C4 till it becomes undetectable in C10 (gel, fig 3.2 c). The purity appears to be the highest in the peak fractions C5 and C6, whose concentrations were found to be around 3 μM from the NanoDrop UV-Vis Spectrophotometer (Supplementary table S2, *Page*

203). A small sample from C5 was used for negative stain EM. Fractions C5 to C8 were subsequently pooled and concentrated to 10 µM for use in quality control assays. Overall, the His-trap strategy resulted in a total yield of 3.3 mg of HSET per Litre of Sf9 culture.

Although it was possible to obtain His_6-FLHSET with reasonable purity for EM purposes, it required a three-step purification which resulted in a significantly low yield of protein recovered. In order to simplify the purification, it was decided to try and replace the His_6 tag with a more specific affinity tag that would improve the purity after the initial step of purification.

3.1.2 Twin-StrepII-tagged FL HSET

To improve the yield of HSET while ensuring a high level of purity, I produced a new construct with a 2xStrepII i.e., twin-Strep affinity tag, which encodes two repeats of a sequence of 8 amino acids that mimic the interaction site of biotin (Schmidt, Koepke et al. 1996). The matrix of the StrepTrap HP column contains the biotin-binding domain of Streptavidin and is known as Streptactin. In other words, the Strep tag- Streptactin column interaction is designed to mimic that of biotin to streptavidin. The specificity of this interaction along with the lack of biotinylated proteins in the lysate implies a higher probability of successfully purifying HSET from contaminants.

Figure 3.3 Construct for $Strep_2$-FLHSET

The second construct used to express and purify recombinant FL HSET was designed with an N-terminal 2xStrep tag followed by a downstream HRV 3C protease cleavage site.

As shown in fig 3.3, the position of the tag was the same as before but followed by an HRV 3C protease cleavage site because HRV 3C protease is more robust than TEV. This construct, referred to as $Strep_2$-FLHSET was initially purified using a StrepTrap affinity column followed by a Superdex 200 10/300 GL SEC column.

3.1.2.1 Step-1: Affinity Chromatography using StrepTrap

In the case of the $Strep_2$-FLHSET, the elution from the StrepTrap column was facilitated by the incorporation of 3 mM Desthiobiotin in the binding buffer, which was not used during the binding and washing phases, in contrast to the imidazole used to elute His-tagged FLHSET from the HisTrap column. Eluted fractions showed a marked reduction in the number of contaminant bands (gel, fig 3.4 a) compared to those from the HisTrap column (gel, fig 3.2 a). Furthermore, HSET was present in more fractions than in the HisTrap eluates (gel, fig 3.2 a). Thus, the use of the twin-Strep tag noticeably improved the recovery and purity of HSET compared to the His-tag. The eluates (A1 to A6) were pooled and concentrated for SEC.

3.1.2.2 Step-2: SEC using Superdex 200

Unlike the case of His_6 -FLHSET that showed multiple overlapping peaks over a span of 8 ml of the column, the SEC for $Strep_2$-FLHSET showed two peaks (profile, fig 3.4 b), both of which contained HSET (gel, fig 3.4 b). The first peak begins close to the void volume of the column and represents the higher MW species and aggregated proteins, including HSET. The second peak elutes around 10 ml, similar to the peak of HSET elution that was observed in the case of His_6-FLHSET (profile, fig 3.2 b). From the SDS page gel (fig 3.2 b), most of the contaminants co-elute with the aggregate peak. Given the high level of purity of HSET in the second peak there was no need for a third chromatography procedure. The total yield was 12 mg HSET per litre of culture, which is four times higher than that of His_6-FLHSET.

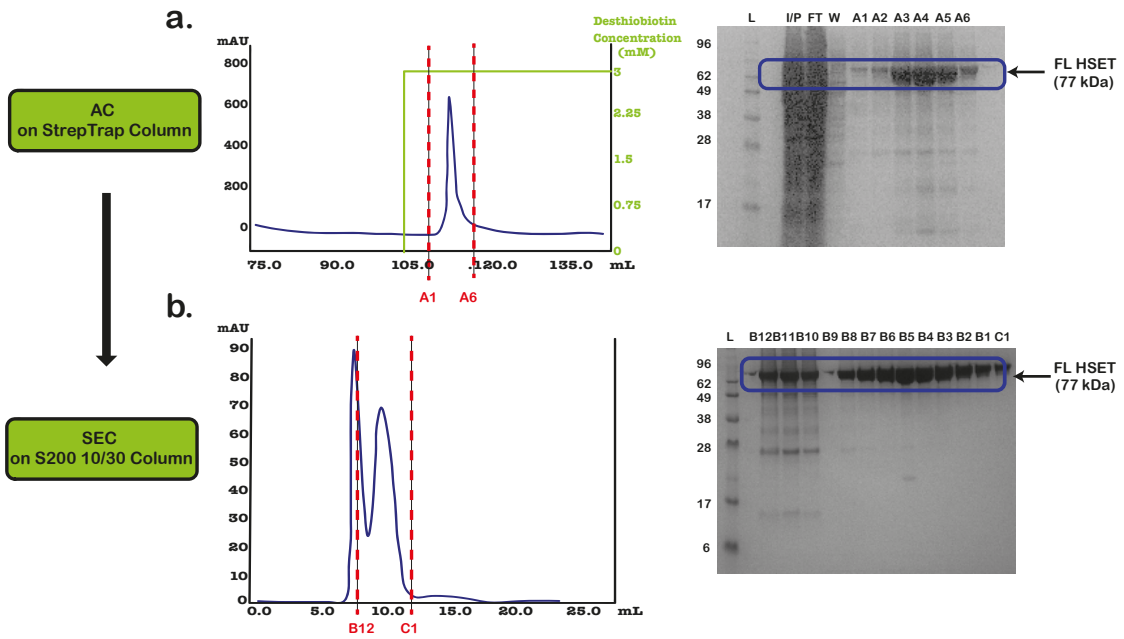


Figure 3.4 Purification of Twin-StrepII-tagged FL HSET

The initial approach of purifying FL HSET from the filtered cell lysate involved the following two-step strategy- (a) Affinity chromatography through a StrepTrap HP column, followed by (b) SEC through a Superdex 200 10/300 column. The left panel shows chromatograms for each step, absorbance in mAU (blue line) is shown on the left- Y-axis and the eluant concentrations in mM (green line) is shown on the right-Y-axis, both are plotted against the volume in ml along the X-axis. Please note that the chromatogram for the StrepTrap HP column has been cropped and does not include the loading phase. The eluted fractions were analysed by SDS-PAGE as shown on the right panel; the red dashed lines on the chromatograms depict where eluted fractions were loaded onto the gel. L- denotes the molecular weight ladder, I/P- Input (injected sample), F/T- flow-through, W- wash.

It is thus evident that changing to a twin-Strep tag showed a significant improvement in yield of HSET compared to the His_6-tag. Due to the high concentration of HSET in each fraction, the samples were not concentrated prior to storage.

3.1.3 Optimising the twin-StrepII strategy

Although switching to a twin-Strep tag significantly improved the yield and purity of HSET, further modifications to the purification were made. These are discussed below and the HSET fractions obtained from it are referred to as $Strep_2$-FLHSET-II, while its predecessor as $Strep_2$-FLHSET-I.

3.1.3.1 Improving the resolution in SEC

Although it was possible to resolve aggregated and non-aggregated HSET, the peaks were not fully resolved which could lead to heterogeneity in the sample.

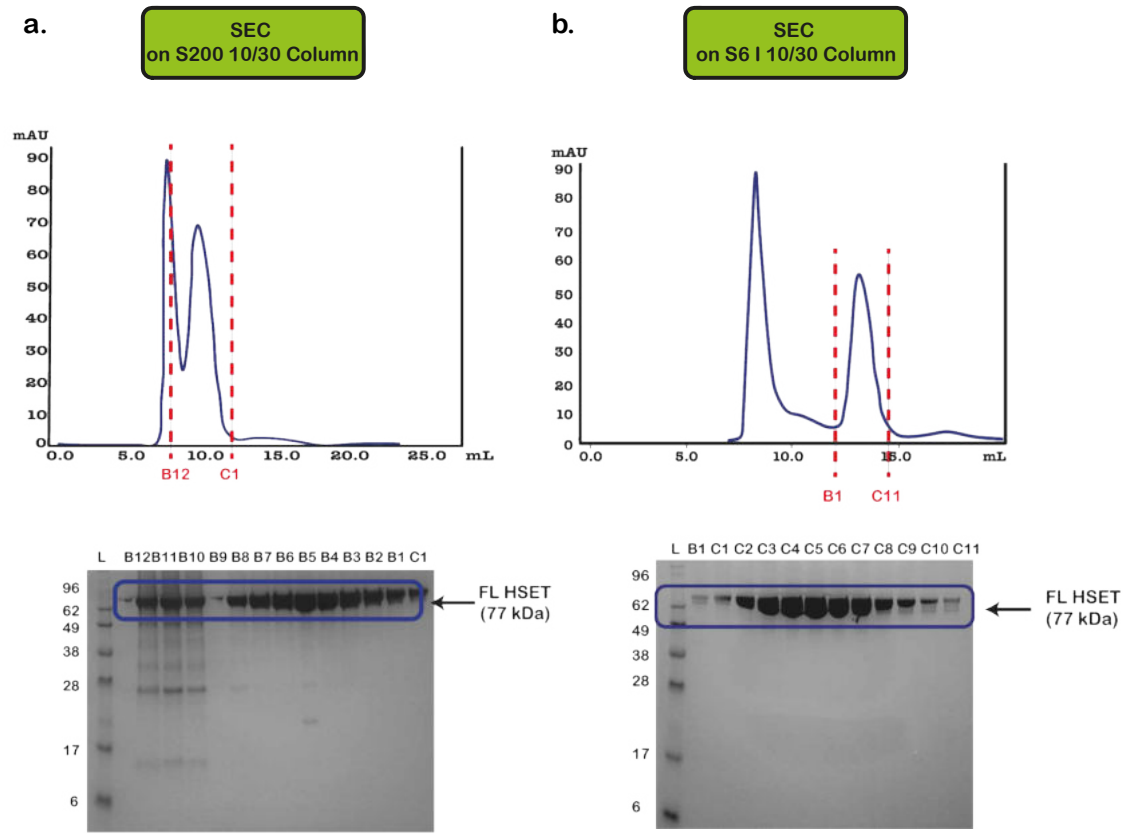


Figure 3.5 Comparison of SEC Columns - Superdex 200 10/30 vs. Superose 6 Increase 10/30

SEC of $Strep_2$-FLHSET using (a) Superdex 200 10/30 GL column vs. (b) Superose 6 Increase 10/30 GL column; the chromatograms of each step show absorbance in mAU (Y-axis) plotted against the volume in ml (X-axis). The eluted fractions were analysed by SDS-PAGE; the red dashed lines on the chromatograms depict the range of eluted fractions that were loaded on the gel. The remaining lanes represent the following: L- molecular weight ladder.

To avoid this, I decided to change the SEC column from Superdex 200 to the Superose 6 Increase, which provides a higher resolution range for the same column volume (Hong, Koza et al. 2012). As shown in fig 3.5, the Superose 6 column provides a much better separation between the two peaks of HSET compared to the S200 column. The SDS-PAGE analysis of the eluted fractions from the second peak showed a high level of purity (gel, fig 3.5 b).

3.1.3.2 Preparation of ADP-HSET

Another modification to the purification of $Strep_2$-FLHSET was the plan to divide the pool of StrepTrap fractions into two halves to independently purify Apo-HSET and ADP-HSET. The aim was to generate a version of HSET that would be stable in low-salt, glycerol-free buffer to make the sample more suitable for cryo EM. ADP-binding was used to stabilise HSET in the absence of the above components of the buffer.

To generate ADP-HSET, 0.5 mM each of Mg and ADP were added to half of the pool of StrepTrap fractions, which was subsequently concentrated for SEC using the Superose 6 Increase in the ADP-SEC running buffer (see section 2.1.3.2, *Page 36*). The SEC profile for ADP-HSET shows a significant reduction in the contents eluting at the void volume compared to Apo-HSET (fig 3.6). The elution of HSET shows a peak of similar height and width and appears in the same position of the column as Apo-HSET. While the overall yield of 12 mg HSET per litre of culture remained the same as the earlier preparation of $Strep_2$-FLHSET, the concentration of HSET per fraction measured by the NanoDrop UV-Vis spectrophotometer halved as expected (Supplementary table S2, *Page 203*).

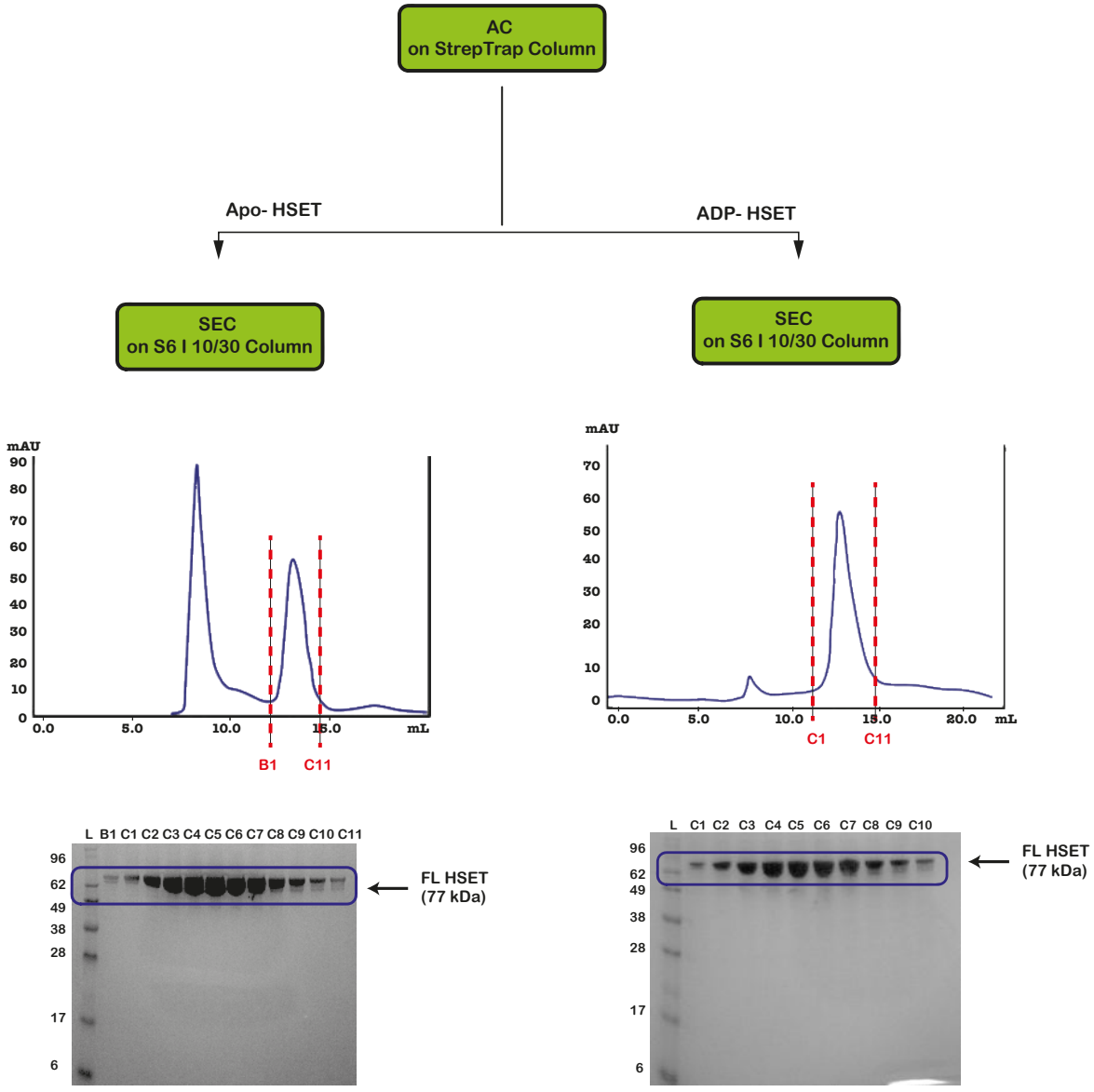


Figure 3.6 Purification of Strep2-FL HSET-II

Top panel shows the schematic of the improved alternative approach to purify twin-Strep-tagged FL HSET as follows- After affinity chromatography through a StrepTrap HP column, the eluted fractions were pooled together and split in two; ADP and $MgCl_2$ were added to one half to a final concentration of 0.5 mM. Both samples were subsequently concentrated and independently resolved through SEC using a Superose 6 Increase 10/300 column; the chromatograms of each step show absorbance in mAU (Y-axis) plotted against the volume in ml (X-axis). The eluted fractions were analysed by SDS-PAGE; the red dashed lines on the chromatograms depict the range of eluted fractions that were loaded on the gel. The remaining lanes represent the following: L- molecular weight ladder, I/P- Input (injected sample).

In this manner, I made modifications to the initial pipeline to create a new version of $Strep_2$-FLHSET showing a higher level of homogeneity in the fractions and the retrieval of two forms of HSET for cryo EM. Apo and ADP-HSET were compared biophysically to evaluate the differences in quality and thus determine their suitability for cryo EM.

3.2 Biochemical and Biophysical Characterisation of FL HSET

In this section, I describe the results of experiments that were undertaken to study the stability, stoichiometry, size, and shape of HSET. Like many structural biology techniques, several of the methods I employed to estimate these parameters were based on the scattering of electromagnetic radiation by the sample. Light scattering is a versatile physical phenomenon that can provide a plethora of information about macromolecules in solution based on the design of the experiment and the ratio of the wavelength of light to the size of the molecule (Sahin and Roberts 2012, Stetefeld, McKenna et al. 2016).

For instance, the stability or the lack thereof is a term that encompasses visual and biochemical information about the system. For instance, a less flexible structure is easier to visualise in structural techniques such as X-ray crystallography, cryo EM, or NMR spectroscopy. In addition, the heterogeneity of a sample due to its tendency to aggregate, denature or degrade, can refer to its biochemical stability as a function of time, temperature, and the presence of certain contaminants, buffer components, and cofactors. These experiments are crucial as they inform strategies for sample preparation for structural characterisation of a protein under study. Prior to any characterisation, it was imperative to ensure that the purified HSET sample I had obtained was biochemically active. Hence, I first performed a biochemical assay for quality control as explained in section 2.2.1 (*Page 37*).

3.2.1 Confirming the functionality of HSET

As HSET is known to hydrolyse ATP during its mechanical cycle, I decided to test its activity through an ATPase enzymatic assay, the ADP-Glo™ Assay (Promega, UK). As this assay was used to assess the activity of HSET rather than establish a kinetic study on the ATPase reaction, the concentration of ATP was maintained in a 1,000-fold molar excess of HSET to prevent the concentration of the substrate from limiting the rate of the reaction. The steps involved in the assay are delineated below.

3.2.1.1 Principle of ADP-Glo assay on HSET

In a nutshell, the assay consists of three steps (Zegzouti, Zdanovskaia et al. 2009)- reaction, depletion, and detection (fig 3.7). The first step is the ATPase reaction that occurs as HSET is exposed to a fixed concentration of pre-polymerised microtubules in the presence of the substrate- ATP. The next step involves the addition of the ADP-Glo reagent, which stops the ATPase reaction and simultaneously depletes the unused ATP in the reaction mixture, leaving the product- ADP. Following this, the kinase detection reagent is added to commence the third step of the assay. This final step reconverts the ADP that remains in the reaction mixture to ATP, which is then utilised to drive a Luciferase/Luciferin reaction that generates a bioluminescent product. Hence, the amount of light generated by the product is proportional to the amount of ATP that has been hydrolysed by HSET.

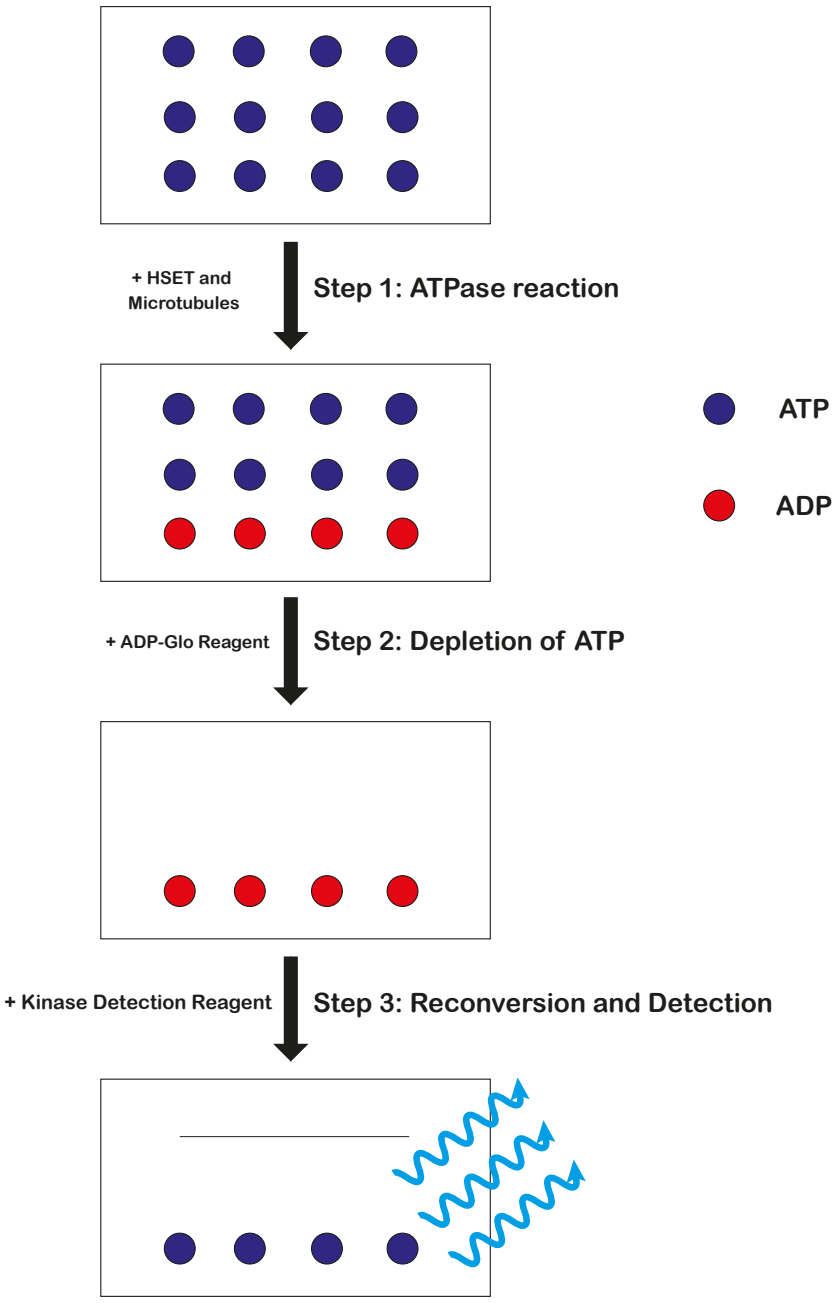


Figure 3.7 Principle of ADP-Glo Assay

The ADP-Glo Assay proceeds in three steps: Step 1- enzymatic reaction occurs, hydrolysing ATP and releasing ADP; Step 2- The reaction is stopped, and the unused excess substrate ATP is depleted; Step 3- The ADP is reconverted to ATP to be utilised by a Luciferase-catalysed reaction that emits light at a wavelength of 450 nm.

3.2.1.2 Confirmation of HSET's enzymatic activity in the ADP-Glo assay

Fig 3.8 shows the activity of the two HSET variants that were purified, as measured by ADP Glo assay.

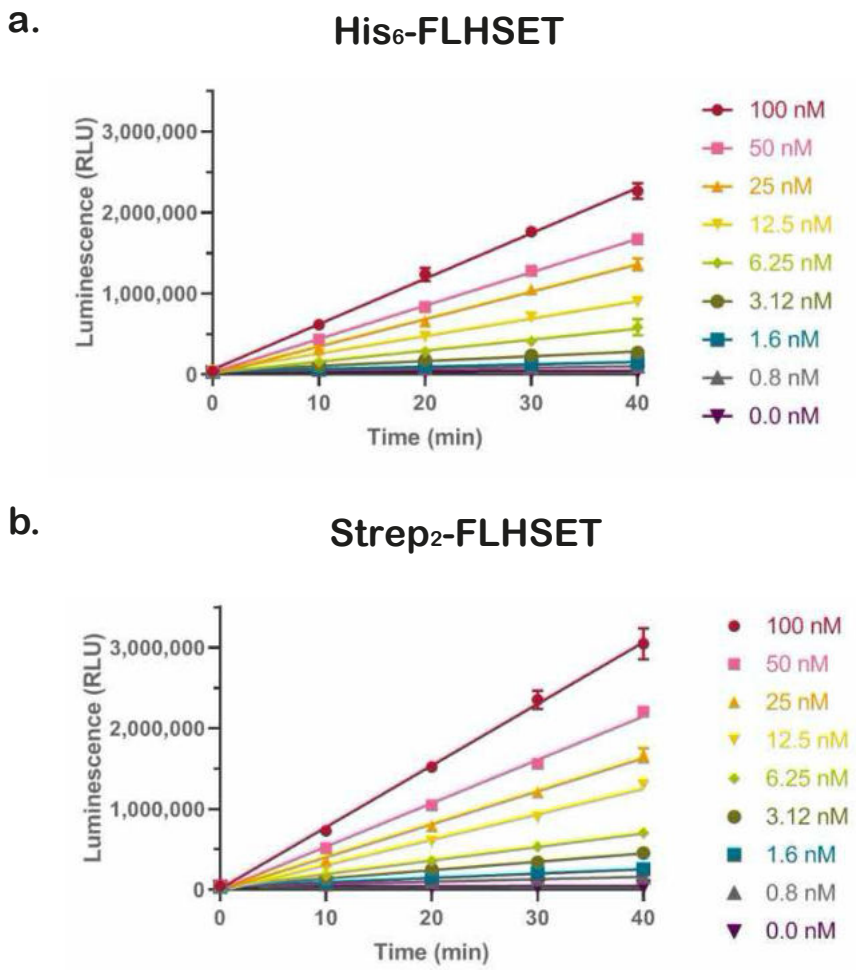


Figure 3.8 ADP-Glo Assay of FL HSET Constructs

The ADP-Glo assay was used to confirm and compare the ATPase activity of the batches of (a) His_6-FLHSET and (b) $Strep_2$-FLHSET-I; the formation of bioluminescent product was measured in terms of the detected luminescence in Relative Luminescence Units (RLU) and plotted as a function of time in minutes; the protein was serially diluted to measure the enzymatic activity over the different concentrations displayed on the right.

The plots show the recorded luminescence increasing linearly with time in both cases, indicating an accumulation of the product and therefore, turnover of

substrate over time. Moreover, the titration of HSET in both cases results in a trend wherein the slope of the luminescence plots increases consistently with increasing concentration of the enzyme HSET. However, there is a difference between the slopes of the two samples at each concentration. For instance, at a concentration of 0.8 nM, the slope of the plot for $Strep_2$-FLHSET-I (fig 3.8 b) is twice that of His_6-FLHSET (fig 3.8 a), indicating a greater ATP turnover in the case of the former. As the HSET concentration increases, the activity of $Strep_2$-FLHSET-I continues to overtake that of His_6-FLHSET by a factor of 1.2-1.5. Overall, the positive slope of the plots along with the concentration-dependent response confirmed the presence of enzymatically active HSET from both purification strategies. Moreover, a higher protein activity was observed in the case of $Strep_2$-FLHSET, which is further confirmation of the purity of the sample.

3.2.2 Investigating ADP-binding through DSF of HSET

To study the effects of ADP binding on the structure of HSET, I undertook different approaches. The first was to analyse the difference between Apo and ADP-HSET in terms of structural stability. As HSET transitions through several nucleotide-bound states in its mechanism as a kinesin (Foster and Gilbert 2000), a hierarchy in the relative stability of each state is likely to exist. Crystal structures of HSET and its homologs have only been obtained for the ADP-bound form, rather than the Apo form (Sablin, Kull et al. 1996, Kozielski, De Bonis et al. 1999, Park, Ma et al. 2017) . Based on this, I hypothesised that ADP binding could stabilise the structure of Apo-HSET, making it easier to visualise through structural studies. Thus, I chose to investigate the structural stability of HSET by subjecting it to thermal denaturation and observing any modification in its denaturation profiles upon nucleotide addition. This type of experiment is otherwise referred to as differential scanning fluorimetry (DSF) and is described below.

3.2.2.1 Comparison of Apo and ADP-HSET by DSF

DSF is a biophysical technique that typically uses a fluorescent dye, such as SYPRO Orange, to monitor the gradual unfolding of a protein structure by subjecting it to steadily rising temperatures (Semisotnov, Rodionova et al. 1991). Another type of DSF, usually referred to as nanoDSF (Alexander, Wanner et al. 2014), is label-free, and this is what I performed using the Prometheus NT48 (NanoTemper). Instead of using dyes, the Prometheus measures the intrinsic fluorescence of tryptophan (Trp) and to a lesser extent tyrosine (Tyr) residues (Alexander, Wanner et al. 2014) in the protein to obtain a denaturation profile. It can make precise measurements from capillary tubes that contain just 5 µL of sample at concentrations as low as 10 µM. These residues are typically buried in the hydrophobic core of the protein and become exposed to the aqueous environment as the structure unfolds. The denaturation profiles or melting curves follow the gradual exposure of these residues, which results in a reduction of the fluorescence intensity generated by Trp and Tyr residues at 330 nm, as well as a shift in maxima of emitted fluorescence from 330 nm to 350 nm for Trp residues (Ghisaidoobe and Chung 2014). To follow both changes, the Prometheus records the fluorescence intensities at each wavelength as a function of temperature and then plots the ratio of these intensities, which accounts for any variation in wavelength of the maximal fluorescence emitted by Trp residues. I thus used the Prometheus to study the effects of ADP binding on the thermal stability of HSET, as shown by the melting curves presented below.

3.2.2.2 ADP-binding improves the thermal stability of HSET

The curves in fig 3.9 have significantly different shapes depending on the presence or absence of ADP. The melting temperature- T_m shown in the figure is defined as the temperature at which half of the molecular population is unfolded (top, fig 3.9). The T_m in presence of ADP (T_m-ADP = 50 ºC, fig 3.9) is shown to be 18 degrees higher than the T_m of Apo HSET (T_m-Apo = 32 ºC, fig 3.9). In addition, there was a marked difference in the onset of unfolding between Apo- and ADP-HSET as well as the temperature range over which the denaturation was taking place, which translated into a steeper slope for ADP-HSET in the fluorescence ratio plot. This also manifests as a difference between the width and

height of the peaks in the first derivative plots of each sample- a broad, small peak for Apo-HSET and a narrow, tall one for ADP-HSET (bottom, fig 3.9).

Based on the difference in T_m (denoted ΔTm, fig 3.9) obtained for the two forms of HSET, it is evident that ADP binding increases the thermal stability of HSET. This confirms that HSET is effectively able to bind ADP. This represents another form of confirmation that my purified HSET is functional, in addition to the biochemical characterisation provided by the ADP-Glo assay. Moreover, the steeper transition for ADP-HSET vs Apo-HSET is a hallmark of a more homogeneous population of conformers for ADP-HSET (Gao, Oerlemans et al. 2020). An important implication of these results was that ADP-HSET was given priority in subsequent structural studies compared to the apparently more unstable and conformationally heterogeneous Apo-HSET.

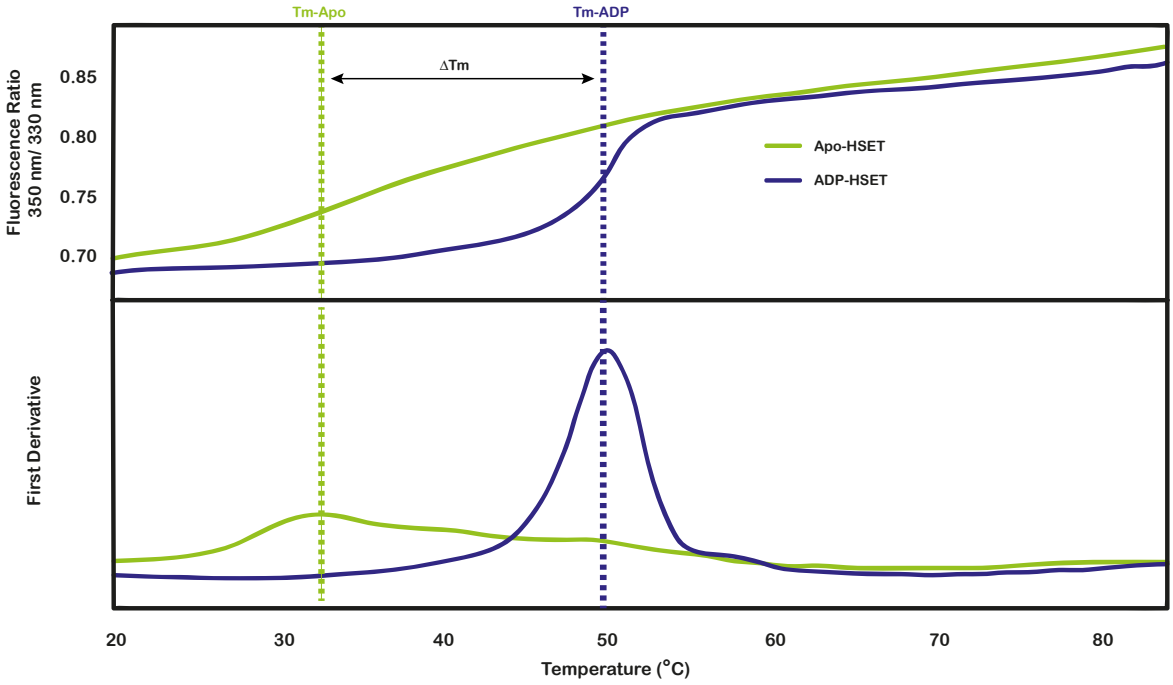


Figure 3.9 Investigating the Thermal Stability of HSET by NanoDSF

Top-Fluorescence ratio 350 nm/ 330 nm of Apo (green) and ADP-bound HSET (blue) is plotted to generate melting curves that follow the denaturation of the proteins as a function of temperature; Bottom- First derivative of the fluorescence ratio curves for Apo-HSET (green) and ADP-HSET (blue) as a function of temperature; melting temperatures (T_m) of both samples are marked by dashed lines and the difference in T_m is denoted by ΔT_m.

3.2.3 Determining the oligomeric state of Apo and ADP-HSET

My second approach in characterising the two forms of HSET was to determine their molar masses and hence their oligomeric states. These defining characteristics provide information that would serve as a valuable reference for visual molecular identification during cryo EM. Furthermore, any difference in oligomeric state of HSET subunits occurring due to ADP binding would give a better understanding of the dynamics involved in HSET's mechanical cycle. One of the techniques I used was Size Exclusion Chromatography coupled to Multi-Angle Light Scattering (SEC-MALS).

3.2.3.1 SEC-MALS of HSET

In this technique, the protein molecules in the sample are first separated through a size exclusion chromatography (SEC) column, followed by two types of detectors. The MALS detector enables measurement of the average intensity of the light scattered (also referred to as Static Light Scattering) by the sample at multiple angles, upon irradiation by a laser of wavelength (λ) 658 nm. As shown below in Eq.3.1, the intensity of the light (I(θ)) is proportional to the molar mass (M) of the sample, its concentration (c), the square of the specific refractive index increment (dn/dc, which has a value of around 0.1855 ml/g for proteins), and the angle between the directions of the incident beam and the scattered light (P(θ)) (Sahin and Roberts 2012).

$$I(\theta) = \left(\frac{4\pi^2 n_o^2}{N_A \lambda^4}\right)\left(\frac{dn}{dc}\right)^2 c\ M\ P(\theta) \qquad (3.1)$$

N_A – Avogadro's number,

n_o – refractive index of the solvent.

If the size of the analysed molecules is sufficiently less than the wavelength of the incident laser light, which is the case for HSET, then the intensity of the scattered light is the same at every measured angle, giving rise to the most

accurate results in the downstream molar mass analysis. The sample then goes through a differential refractive index (dRI) detector, which measures the concentration of the sample at every position of the SEC chromatogram (Striegel 2017). Knowing the protein concentration makes it possible to calculate the molar mass at every point of the chromatogram from the measured intensity of scattered light.

3.2.3.2 Apo and ADP-FL HSET are homodimers

The light scattering (LS) profile in fig 3.10 a and b depicts the elution profile of HSET from the SEC column, which occurs at nearly identical positions in the chromatogram for Apo- and ADP-bound HSET. Apart from measuring the protein concentration, the dRI detector enabled the measurement of the specific refractive index increment (dn/dc) of both samples (Striegel 2017), which was showed a reading of 0.1895 ml/g. This was within the range of values that are representative of an aqueous protein solution devoid of any other type of macromolecule (Some, Amartely et al. 2019). The nearly horizontal appearance of the molar mass distribution across the HSET peaks indicates size homogeneity of both samples.

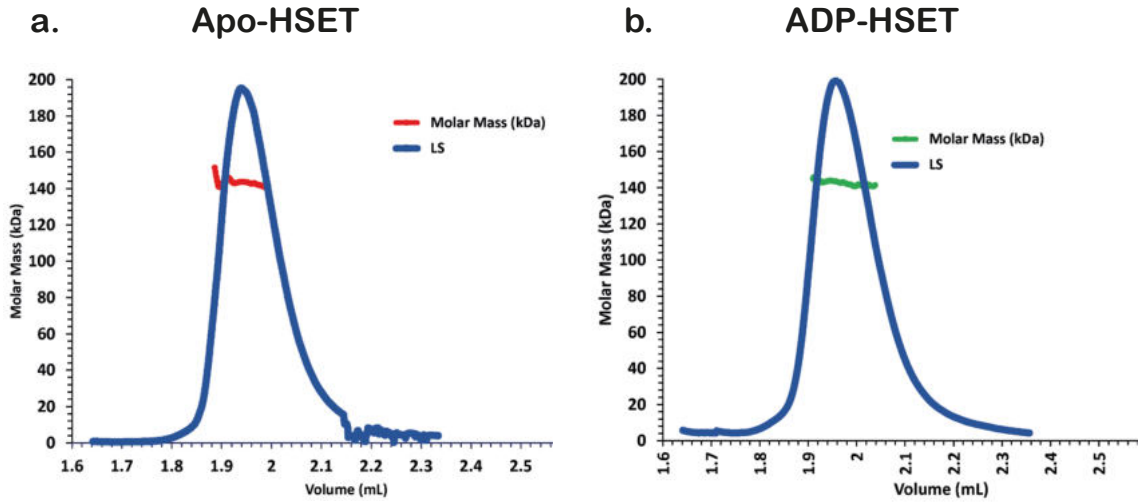


Figure 3.10 Determining the Oligomeric State of HSET

SEC-MALS analysis of (a) Apo-HSET and (b) ADP-HSET, showing the Light Scattering intensity (LS) of the sample as it elutes off a Superose 6 Increase 5/15 GL size exclusion column; The red and green lines show the molar mass distribution across the volume range of elution for Apo-HSET and ADP-HSET respectively.

The molar mass calculated for both HSET samples show a value close to 144 kDa. On comparison with the molar mass predicted from the sequence (Supplementary Information S2B, *Page 201*) which is 77.5 kDa for the monomer, I concluded that HSET is present as a homodimer in solution, in both its Apo and ADP-bound forms.

3.2.4 Monitoring HSET's tendency for aggregation

Having obtained information about the oligomeric state of HSET, I wanted to investigate the tendency of HSET to aggregate and was also curious about the distribution of molecular weights in the sample. This was especially important for cryo EM, which involves identifying individual particles of HSET in electron micrographs. Any tendency exhibited by HSET for non-specific association between molecules would lead to the occurrence of numerous molecular species in a spectrum of sizes. This heterogeneity within the molecular population of HSET would hinder the ability to isolate dimeric HSET while increasing the chances of including off-target molecules in the EM dataset. Moreover, the greater the aggregation, the fewer the individual HSET molecules available to solve the structure. Thus, I used another light scattering technique called Dynamic Light Scattering- DLS, to evaluate the level of size heterogeneity in the sample, referred to as polydispersity in DLS experiments, and its variation with temperature as described below.

3.2.4.1 DLS of HSET

Unlike its Static light scattering counterpart, SEC-MALS, which measures the average scattered intensity, DLS uses the evolution of the scattered signal with time to study diffusing macromolecules in solution. DLS works on the principle that the speed of diffusion of a solute molecule by Brownian motion (Tyndall, 1868) is inversely proportional to its size, shape, and hydration, represented by the hydrodynamic radius (Stetefeld, McKenna et al. 2016). The instrument operates by irradiating a molecular sample with a laser and observing the evolution of the light scattered by the molecules over time. This is plotted as an

auto-correlation function, which is used to extract information about the distribution of hydrodynamic radii in the sample. Additionally, as big objects scatter light more efficiently than small objects, DLS is particularly sensitive to the presence of small amounts of aggregates in solution. I used DLS to study the polydispersity of freshly purified Apo- and ADP-HSET. Additionally, as temperature is a controllable parameter in cryo EM sample preparation, DLS measurements were made at 10 °C and 20 °C to assess if the temperature would influence the polydispersity of the samples.

3.2.4.2 HSET displays temperature-dependent polydispersity

The distribution plots in fig 3.11 a and b were obtained from DLS measurements on Apo- and ADP-HSET at two different temperatures- 10 °C and 20 °C. They provide information on the molecular species present in the sample. At 10 °C, the distribution plots of both samples resemble a straight line centred around a hydrodynamic radius of 10 nm (fig 3.11 a). This is accompanied by some scattering events on either side of the 10 nm line, mostly towards higher values, representing a tendency to multimerise or aggregate. However, based on the mass-weighted histogram shown in fig 3.11 c, majority of the species in both Apo and ADP-HSET have a radius of 10 nm, with a relatively narrow distribution around this value. As the temperature is increased to 20 °C, the distribution of both samples undergoes significant change, with the central peak shifting slightly higher than 10 nm (fig 3.11 b).

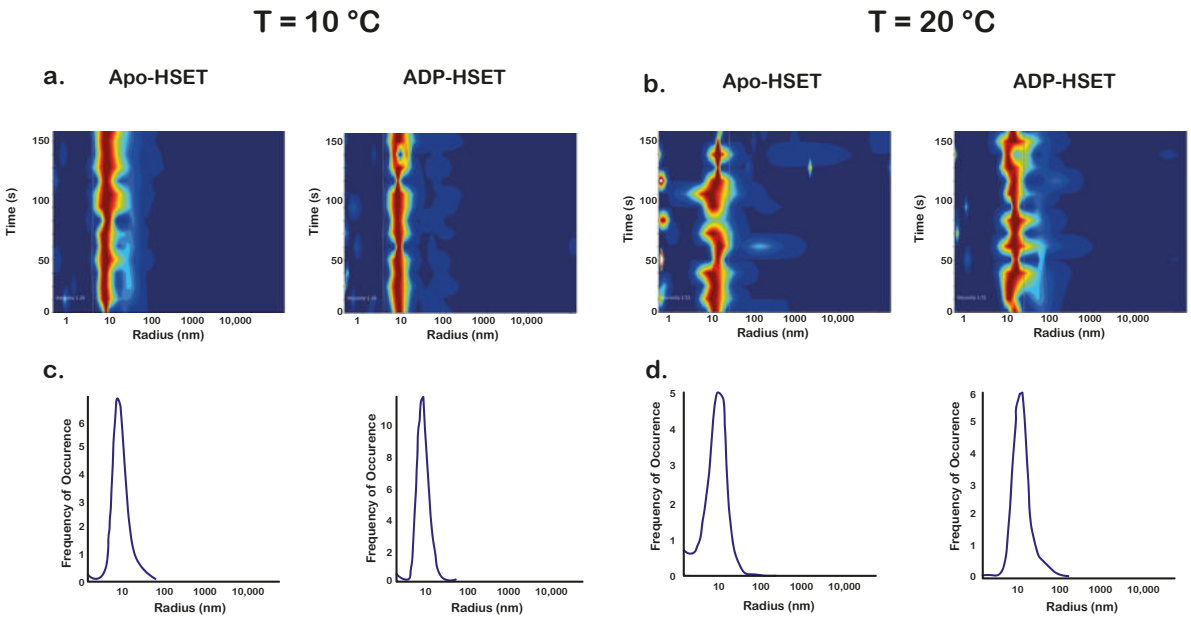


Figure 3.11 Studying the Evolution of Molecular Heterogeneity in Apo and ADP-HSET with temperature

Analysis of HSET by DLS reveals 3D distribution plots showing the evolution of heterogeneity in the Apo and ADP-HSET represented by estimated particle radius (X-axis) with time (Y-axis) at a specified temperature of (a) 10 °C and (b) 20 °C; the colours represent the intensity of light scattered by the molecular species in the sample; mass-weighted histograms show the prevalence of molecular species of Apo and ADP-HSET at (c) 10 °C and (d) 20 °C.

For both samples, one can observe a less narrow distribution of the species at 20 °C, with a greater amount of signal at high radius corresponding to aggregates than was observed at 10 °C. Accordingly, one can observe a broadening of the peaks in the mass-weighted histograms for both samples, highlighting a higher polydispersity of the samples at 20 °C in comparison to 10 °C. In conclusion, DLS showed that the samples of Apo and ADP-HSET must always be kept at temperatures of 10 °C or below to avoid aggregation.

3.2.5 Determining the salt and glycerol requirement of HSET

In addition to sample temperature, heterogeneity can occur due to inadequate concentrations of stabilising agents like salt and glycerol often used in recombinant protein preparations. Cryo EM is an inherently low-contrast technique and HSET is near the lower limit of the size range that can be clearly visualised in cryo-micrographs. The presence of salt and glycerol in sample buffers further reduces the contrast by contrast-matching, which hinders the visibility of HSET particles in the images (Drulyte, Johnson et al. 2018). To prevent this, it is common practice to dilute these components to the greatest possible extent during sample preparation. However, both components also play a pivotal role in keeping the protein stable in solution and thus the extent of their dilution must be decided based on HSET's requirements.

3.2.5.1 Mass Photometry of HSET

Mass photometry is a technique that is designed to determine the molar mass of macromolecules with light, using the process of scattering interferometry (Young, Hundt et al. 2018). In this technique, a light source is used to irradiate the molecular sample, which is placed on the other side of a reflective surface. The interference between the light that is scattered from the sample and the light that is reflected from the surface provides a detectable signal, which is proportional to the molar mass of the sample. Thus, a real-time count of the number of binding events generated by the molecules adhering to the surface is obtained as a histogram of ratio-metric contrast. On calibration, these values can be converted to obtain molecular mass distributions of the sample, as shown in fig 3.12.

To determine the minimal concentrations of the buffer components that HSET can afford in its Apo and ADP-bound forms, I performed a series of mass photometry experiments. To study the response of HSET to salt-depletion, I performed dilutions of Apo and ADP-HSET to 200 nM in three buffers (see section 2.2.5, *Page 38*) to obtain final NaCl concentrations of 200, 140 and 90 mM respectively (fig 3.12). However, it is important to note that Apo-HSET was purified in a buffer containing 5% v/v glycerol while ADP-HSET was purified glycerol-free (see section 2.1.3.2, *Page 36*). Thus, the plots obtained for Apo-

HSET in fig 3.12 b, c and d reflect the response of the sample to salt-depletion in an aqueous environment that contains a small amount of glycerol (0.8% v/v) post dilution. On the other hand, the response of ADP-HSET to salt-depletion studies the stabilising effect of cofactor ADP as a substituent to glycerol.

3.2.5.2 Apo-HSET tolerates a lower salt concentration than ADP-HSET

The molar mass distributions obtained by mass photometry for the salt titration experiments on Apo and ADP-HSET are shown in fig 3.12. The first dilution for Apo-HSET was performed in its original Apo-SEC buffer (400 mM NaCl, 5% v/v glycerol) and the corresponding molar mass distribution shows the initial conditions of the sample (fig 3.12 a). In the case of ADP-HSET, the salt concentration of buffer-1 matched its original buffer (200 mM NaCl, no glycerol), thereby functioning as a representation of the initial conditions of the sample. Apo-HSET appeared completely stable in its original buffer, while ADP-HSET already showed noticeable signs of monomerization and higher order multimerization (right, fig 3.12 b). Moreover, the proportion of dimers in the population was also significantly less in the case of ADP-HSET when compared to an apparently homogeneous Apo-HSET.

As both samples were exposed to progressively decreasing salt concentrations, the percentage of monomers and higher order oligomers increased for both forms of HSET and that of the dimer consequently decreased. The dimeric proportion in Apo-HSET dropped below 90% in buffer-3 (left, fig 3.12 d) while ADP-HSET crossed that margin earlier in buffer-2 (right, fig 3.12 c). This indicated that the lowest concentration of salt that could be afforded by Apo-HSET was around 140 mM, which is more tolerant than ADP-HSET that seemed to barely handle 200 mM of salt. In other words, while ADP binding might improve the structural stability of HSET, it cannot replace glycerol as a biochemical stabilizer.

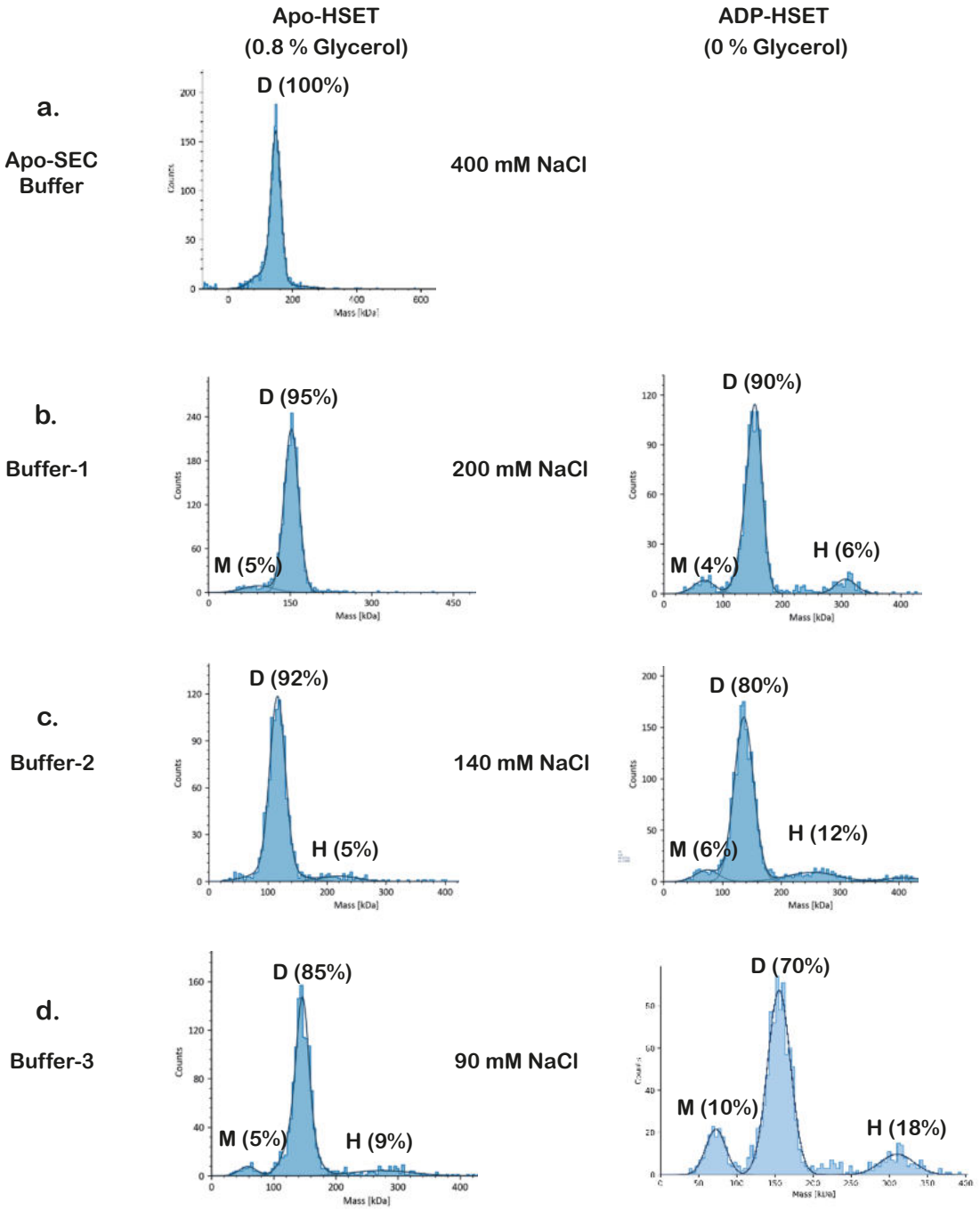


Figure 3.12 Salt titration of Apo and ADP-HSET

Mass photometry was used to determine the effects of NaCl depletion in the aqueous environment of HSET by observing its molecular size distribution when diluted to the following final salt concentrations (corresponding buffers within brackets) - (a) Apo-SEC buffer (400 mM), (b) buffer-1 (200 mM), (c) buffer-2 (140 mM) and (d) buffer-3 (90 mM); Dilutions of Apo-HSET brought the glycerol content down to 0.8% v/v while ADP-HSET was glycerol-free from the start. Histograms of the molecular weights show the abundance of each of the following species in the diluted samples: M- monomer, D- dimer, H- higher order multimers.

3.2.5.3 Glycerol proves to be essential for ADP-HSET as well

Despite having a population that contained 90% dimeric HSET in its original buffer conditions of 200 mM NaCl (right, fig 3.12 b), it was evident that ADP-HSET was more heterogeneous than its Apo counterpart in a buffer of equivalent salinity. I could already detect species whose molecular weights are greater than and less than the dimer when the sample was diluted in conditions that matched its parent aqueous environment (right, fig 3.12 b). I hypothesized that the difference in the stabilities of the two forms of HSET was due to the absence of glycerol in ADP-HSET. To test this, I decided to equalise the conditions used for both samples by incorporating an equivalent amount of glycerol to ADP-HSET.

I managed to obtain a sample of diluted ADP-HSET with identical final conditions to Apo-HSET in buffer-2, namely 140 mM NaCl and more importantly, 0.8% v/v glycerol. The result was a negligible level of multimerisation and denaturation accompanied by an increase in HSET dimers from 80% of the population in glycerol-free conditions (right, fig 3.12 c) to 96% (fig 3.13 a). I subsequently titrated the glycerol content in the ADP-HSET sample from 0.8%, while maintaining the salt concentration at 140 mM. Halving the glycerol content to 0.4% still resulted in a 90% HSET dimeric population (fig 3.13 c). Further attempts to titrate the glycerol down were associated with increased instability in the form of monomerisation rather than multimerisation (fig 3.13 c and d). This contrasted with the effect of the salt depletion experiments which showed a visible increase in both lower and higher molecular weight species.

In summary, both forms of HSET require a final concentration of at least 140 mM NaCl and 0.4% v/v glycerol to remain homogeneous in solution. If higher dilutions of HSET are necessary for cryo EM grid preparation, they will have to be achieved in more than one step to ensure that the final concentration of the two components does not drop below the limit mentioned above.

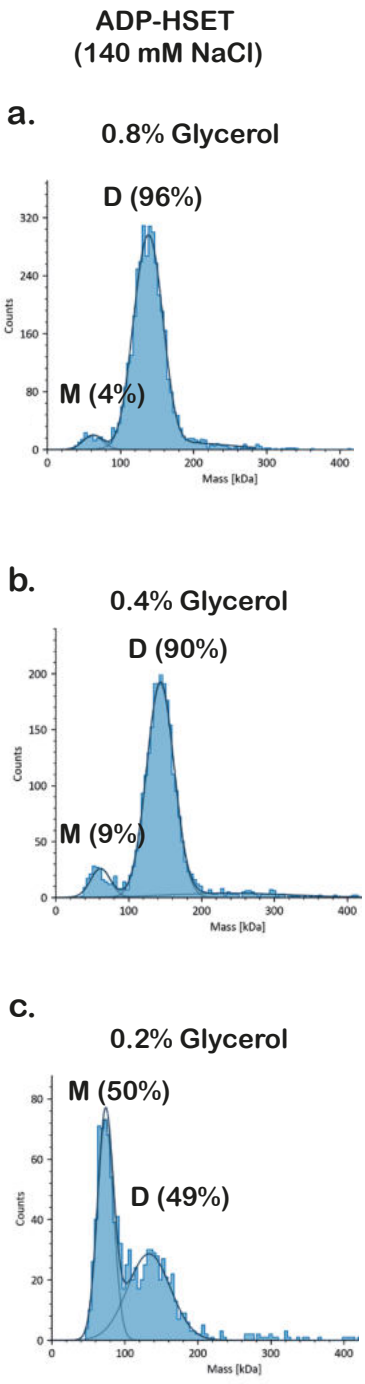


Figure 3.13 Glycerol Titration of ADP-HSET

Study of molecular weight distributions in ADP-HSET at identical final salt concentrations of 140 mM in the presence of varying glycerol content (a) 0.8 %, (b) 0.4 % and (c) 0.2 %; all percentages are represented in v/v. Histograms of the molecular weights show the abundance of each species in each diluted sample: M- monomer, D- dimer.

3.2.6 Observing the behaviour of HSET in solution

Lastly, I compared the two forms of HSET in their native states in aqueous solution using Small Angle X-ray Scattering (SAXS), a scattering technique which unlike the earlier techniques, also provides information about the overall shape and dimensions of HSET molecules, as well as the compactness and rigidity i.e., the level of structure in HSET (Rambo and Tainer 2010). Simply put, SAXS could provide a low-resolution structural analysis of HSET without the restrictions imposed by structural techniques such as crystallization in X-ray crystallography and vitrification in cryo EM. While SAXS is inherently a low-resolution technique, the results obtained can in principle provide information about molecules in solution without any restriction of their mobility, bridging the gap between biophysical and higher resolution structural studies on HSET (Hammel 2012).

3.2.6.1 Basic Principle of SAXS

The operating principle of Small Angle X-ray Scattering (SAXS) has some similarities to MALS described in section 3.2.3 (*Page 65*), with some significant differences being the wavelength of the radiation used and the small angles at which SAXS measurements are made. X-rays used in SAXS typically have a wavelength of around 1 Å, more than 6000 times shorter than the red laser used in MALS, which tips the scales such that the molecular size is now significantly greater than the wavelength of incident light i.e., the light scattered by the sample is now anisotropic (Glatter and Kratky, 1982). As a result, SAXS can provide additional information about the shape, size, and level of structure in the sample compared to MALS.

Similar to MALS, injecting the sample through a SEC column prior to performing the scattering measurement is a preferred option in SAXS, to avoid any heterogeneity in the sample (see methods section 2.2.6, *Page 39*), and to separate the molecules of interest from any aggregates that could “pollute” the scattering signal. It is important to note that a blank curve for the solvent must be measured and subtracted from the data obtained with the sample, to suppress the scattering contribution of the solvent. This data can then be transformed to obtain parameters related to size and shape such as maximum dimension (D_{max})

and radius of gyration (R_g) of the molecule. The transformations can be performed in real or reciprocal space, each revealing unique information about the sample apart from the R_g. These transformations and the results obtained from them are described below.

3.2.6.2 Size of HSET from Fourier space approach

The elution profile obtained by SEC of either sample shows two peaks- peak 1 which corresponds to Apo-HSET (fig 3.14 a) and ADP-HSET (fig 3.14 b) and peak-2 which signals the end of the column volume.

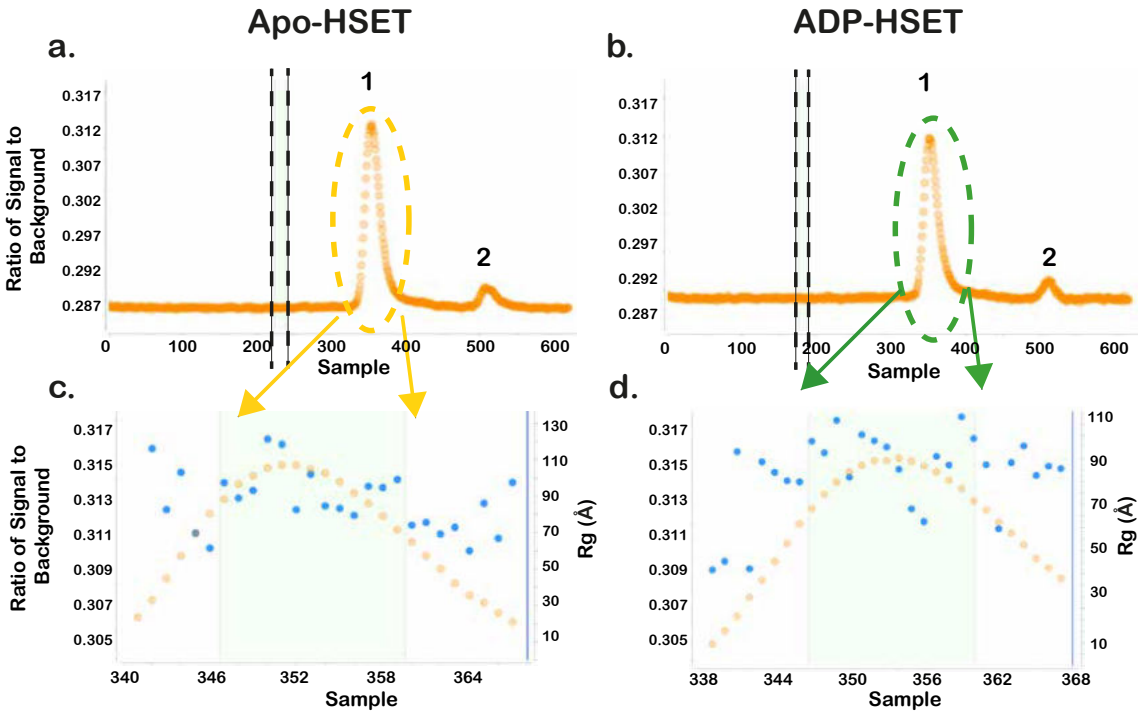


Figure 3.14 SEC-SAXS of Apo and ADP-HSET

(a, b) Chromatograms show the elution of the two forms of HSET (peak 1) through a Shodex KW-404 column during SEC-SAXS, the ratio of the signal scattered by the sample to the background is denoted in arbitrary units in the left Y-axis and the exposure number is shown in X- axis, the region used for buffer subtraction is shaded and delimited by dashed lines in both profiles; (c, d) A close-up view of the R_g distributions across the peak fractions in each case, R_g is depicted as blue dots and measured in Å, as shown in the right Y-axis.

A notable feature of the SEC profiles is the absence of any peak corresponding to aggregates near the void volume which would occur upstream of peak 1, indicating a negligible presence of aggregated protein in the injected samples of HSET. The scattering signals represented by the shaded portion in fig 3.14 a and b were used for solvent subtraction. Based on the distribution of R_g across the eluted fractions estimated by SIGNAL PLOT in ScÅtter-3 [http://www.bioisis.net/tutorial/9], the shaded regions in fig 3.14 c and d were used for the analysis.

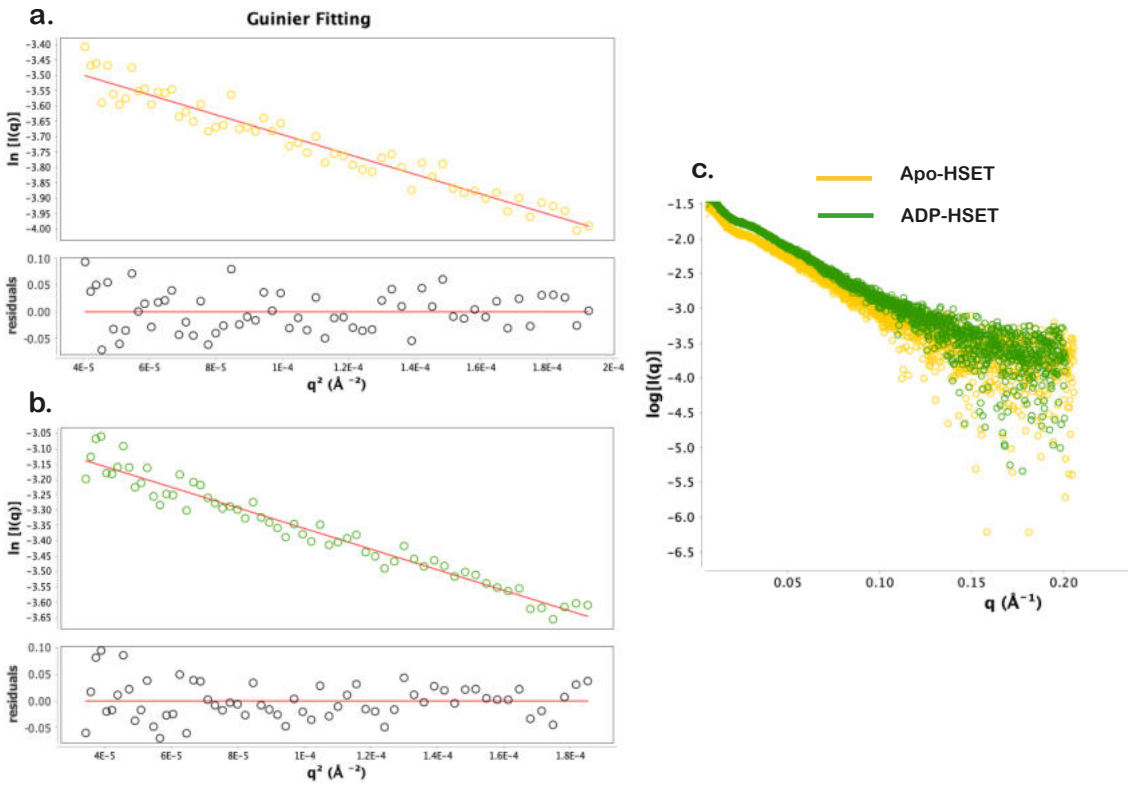


Figure 3.15 Guinier Analysis of Merged Apo and ADP-HSET Datasets, and Corresponding Full Scattering Plots

Close-up view of the linear fit in the Guinier region of (a) Apo-HSET and (b) ADP-HSET; (c) Overlay of the intensity plots of the scaled and averaged data from the subtracted SEC fractions of Apo (gold) and ADP-HSET (green), the scattering intensity is depicted in logarithmic scale (Y-axis) as a function of scattering vector q (X-axis).

Prior to merging the subtracted data from each fraction, the scattering data at the low-q region i.e., the Guinier region, were manually filtered to remove any outliers and select the portion of the peak giving the most consistent values of R_g, limiting the residuals to a range of -0.2 to +0.2. The absence of a sudden increase in slope closer to q = 0 $Å^{-1}$ (angle of 0º) further confirmed the absence of aggregation in the sample (Serdyuk, Tsalkova et al. 1987). The filtered data was then merged by scaling followed by averaging to result in the logarithmic scattering intensity "log I(q) vs q" plots shown in fig 3.15 c.

A closer look at the Guinier region of the individual plots of Apo and ADP-HSET in fig 3.15 a and b shows that most of the values fit a linear distribution and the low value range of residuals are proof of the good quality of fit. The parameters of the linear regression that the data was fitted to were subsequently used to estimate the R_g and the I(0)- the scattering intensity at a scattering angle of 0°, in an approach that is popularly known as the Guinier approximation (Guinier and Fournet, 1955). In this approach, the variation of I(q) with q can be approximated to an exponential function at low-q values as shown in Eq 3.2, which can be linearized by converting to logarithmic scale in Eq 3.3.

$$I(q) \approx I(0)\, e^{\frac{-q^2 Rg^2}{3}} \qquad (3.2)$$

$$\log I(q) = \log I(0) - \frac{q^2 Rg^2}{3} \qquad (3.3)$$

Thus, the slope of the log I(q) vs q^2 graph can be used to calculate the R_g, while the extrapolation of the data to find the Y- intercept yields the value of I(0)- the scattering intensity at a scattering angle of 0°. I(0) is proportional to the number of electrons in the sample (Orthaber and Glatter 2000) and can be used to calculate the molecular weight of the protein using the molar volume of the protein (Eq 3.4), albeit with a lower accuracy than SEC-MALS (Serdyuk, Tsalkova et al. 1987).

$$MW = \left(\frac{V^2}{cRg}\right)^k \quad (3.4)$$

where $c = 0.1231$ and $k = 1$ for proteins,

$$V\ (Volume) = \frac{2\pi^2 I(0)}{\int_0^\infty q\ I(q) dq}$$

In the case of both Apo and ADP-HSET, the R_g values obtained from the Guinier analysis are found to be between 95 and 100 Å (table 3.1). The molecular weights of both forms of HSET shown in table 3.1 were found to be close to the calculated value for a dimer, which confirms that the species being analysed here are indeed dimeric forms of Apo and ADP-HSET.

Property (Method of Determination)	**Sample**	
	Apo-HSET	**ADP-HSET**
R_g (Guinier fit)	95.2 Å	99.8 Å
MW calculated from I(0) (Guinier fit)	176.55 kDa	169.69 kDa
R_g (Indirect Fourier Transform)	111.2 Å	111.8 Å
D_{max} (Indirect Fourier Transform)	368 Å	374 Å

Table 3.1 SAXS Analysis of Apo and ADP-HSET

The numerical parameters describing the shape and size of either form of HSET are obtained from the analysis of the X-ray scattering curves.

The other end of the plots for both samples is seen to get progressively noisier at high values of q i.e., at higher scattering angles (fig 3.15 c). This noise is an indication of low protein concentration and limits the analysis to low q-values terminating a little over 0.05, corresponding to a maximum resolution of around 20 Å.

3.2.6.3 Degree of structure in HSET

The scattering data can be transformed to unravel unique information about the sample. One such transformation, called the Kratky analysis, involves plotting "I(q). q^2" against "q" to study the level of structure in a polypeptide chain (Glatter and Kratky, 1982). The shape of the Kratky plot varies depending on factors such as the shape of the protein i.e., compact or elongated, the extent of folding and lastly, the size. To allow the comparison of proteins of different sizes and MW, the Kratky analysis can be made size-independent by normalising it using I(0) and R_g, now plotting "I(q)/I(0). $(q.Rg)^2$" against "$q.R_g$". The units of q ($Å^{-1}$) and R_g (Å) are reciprocals of one another, which makes their product dimensionless (Durand, Vives et al. 2010) as shown in fig 3.16 a and b. The shape of this normalised plot can vary from a bell shape that is obtained for well-folded globular proteins to an exponentially increasing curve that is typical of an unfolded random coil (Receveur-Brechot and Durand 2012). However, the deviation from a bell shape is not sufficient evidence to suggest unfolding, as shown by the plateau in the orange curve in fig 3.16 a. This curve represents a protein made of more than one folded domain connected by flexible linkers, whose presence dominates the shape of the plot. Similarly, a bell-shaped curve does not imply a compact globular protein. An important distinction between the Kratky plots for globular and elongated proteins is the location of the peak. As shown in fig 3.16 a, globular proteins show a bell-shaped curve with a peak of 1.1 at $qR_g = \sqrt{3}$, while elongated proteins peak further away from the crossover point (Receveur-Brechot and Durand 2012).

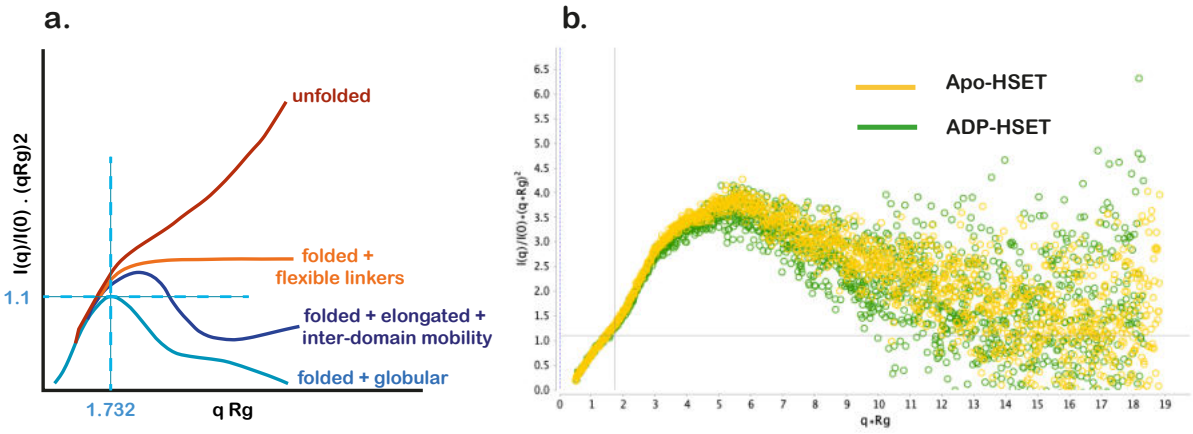


Figure 3.16 Kratky Analysis of Apo and ADP-HSET

(a) typical dimensionless I(0) and Rg-normalised Kratky plots for different types of proteins, adapted from Receveur-Brechot, V. and Durand, D., 2012; (b) an overlay of the dimensionless Kratky plots that estimate the level of structure in Apo (gold) and ADP-HSET (green), in both images, transformed and normalised intensities (Y-axis) are plotted as a function of dimensionless quantity "q.R_g" (X-axis).

Based on the occurrence of a peak and its location, both Apo and ADP-HSET appear to be well-folded and elongated (fig 3.16 b). The nearly complete overlap between the curves in fig 3.16 b shows an indistinguishable degree of structure to both forms of HSET at the experimental temperature of 4°C. The tendency of the distribution to tail upwards at higher values of qR_g is an indication of inter-domain mobility as illustrated in fig 3.16 a.

3.2.6.4 Shape of FL HSET from real-space approach

As described in the methods (section 2.2.6, *Page 39*), the next step was to perform a Fourier transform on the scattering data to determine the shape of the object. However, as SAXS measurements are obtained from molecules that are constantly tumbling in solution, the intensities observed are rotationally averaged values (Glatter and Kratky, 1982). Thus, the Fourier transform of the intensities provides a distribution of interatomic distances known as the P(r) function (Eq 3.5), rather than an electron density map and is thus considered an indirect fourier transform (Guinier and Fournet, 1955).

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} \, dq \qquad (3.5)$$

Albeit not giving the exact shape of the molecule, the P(r) does contain information that can be used to determine the R_g and maximum dimension D_{max} and infer its molecular envelope. The D_{max} is defined as the value of distance "r>0" for which the P(r) function reaches zero. Prior to setting P(r = D_{max}) at 0, the value of D_{max} was manually adjusted to observe a smooth drop off to 0. The resultant P(r) curves for Apo and ADP-HSET in fig 3.17 a are largely similar in overall shape, sharing a tall peak at smaller interatomic distances, followed by a shoulder and a short third peak. Both curves drop off at a value in between 360 and 380 Å, mentioned in table 3.1. The ADP-HSET curve shows an additional minor fourth peak around 350 Å and a slightly higher first peak (fig 3.17 a), indicating the higher frequency of occurrence of this range of distances in the ADP-HSET molecule. The P(r) functions for both forms of HSET are representative of a near cylindrical object, implying an elongated shape for FL HSET (Orun, Koch et al. 2003). The R_g values calculated from the P(r) curve of both forms of HSET agree with each other, while being higher than the value quoted by the Guinier-fit (table 3.1). This discrepancy is likely due to the increased precision of the P(r) method arising from the use of more data. The slightly higher D_{max} values and increased frequencies of the interatomic distances in the ADP-HSET curve are consistent with a slightly more elongated HSET molecule when bound to ADP as compared to HSET.

The P(r) data was used to generate 3D reconstructions by dummy atom modelling algorithms mentioned in the methods. The envelope obtained for both forms of HSET are elongated (fig 3.17 b and c) but possess different features. The biggest difference between the two envelopes is a protrusion at the bottom-left of the Apo-HSET molecule which might be one of the motor domains. While further information about the feature will hopefully be visible at higher resolution through EM, it seems to hint at an asymmetry within the homodimer of Apo-HSET.

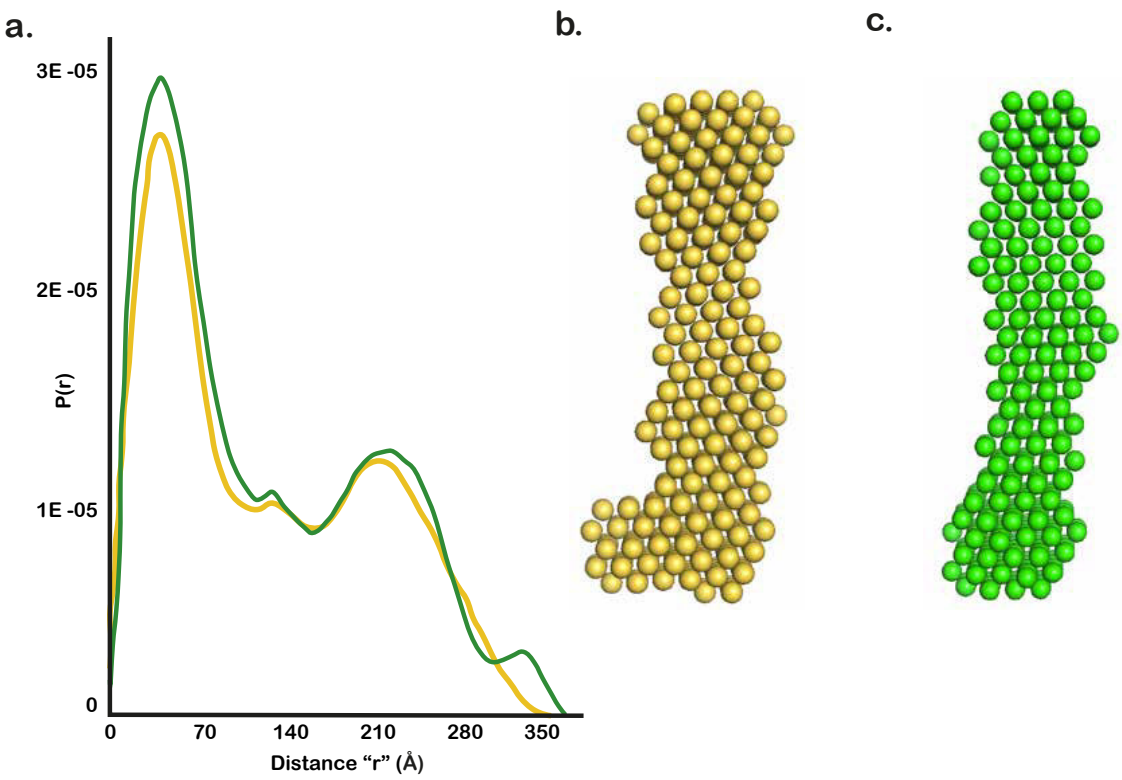


Figure 3.17 Low Resolution 3D Modelling of FL HSET

(a) Overlay of the pairwise intensity distribution P(r) functions of Apo (gold) and ADP-HSET (green) provide insight on the shape of HSET, the "P (r)" (Y-axis) is plotted as a function of interatomic distance "r" (X-axis); Dummy-atom 3D reconstructions of (b) Apo-HSET and (c) ADP-HSET based on the P(r) function.

Overall, SAXS demonstrated that FL HSET is a folded, elongated homodimer in its Apo and ADP-bound forms, with a total length of 365-375 Å in its longest dimension. The low-resolution envelopes generated for apo- and ADP-HSET hint at structural differences between the two forms.

3.3 Image Formation in Single Particle Transmission Electron Microscopy

The characterisation of Apo and ADP-HSET described in earlier sections provided me with evidence that the samples were of sufficient quality to proceed with structural analysis by electron microscopy (EM). Single particle EM is a structural technique that images macromolecular samples at concentrations as low as 0.05 μM (Passmore and Russo 2016). It encompasses two main techniques- negative stain EM and cryo EM. In both cases, the specimen to be imaged is a thin film (below ~100 nm) (Glaeser and Taylor 1978) of macromolecular sample applied onto a surface known as a grid. The preparation of the sample in either EM technique will be discussed in the next sections. In this section, I will guide the reader through the physics of imaging a thin-film specimen by electron microscopy, as a prelude to my results.

3.3.1 From object to image

As mentioned in section 3.2.3 (*Page 65*), the ratio of the size of the object of interest and the wavelength of electromagnetic radiation used to irradiate it, dictates the type of information that can be obtained from a radiation-scattering experiment. Obtaining images that capture the atomic detail of objects in the order of a few hundred nanometres such as macromolecular complexes would theoretically require electromagnetic radiation whose wavelengths are less than a hundredth of a nanometre (Penczek 2010). A beam of electrons accelerated to 200-300 keV approximately possesses a wavelength of 0.002 nm, which is more than capable of capturing the structural detail within proteins or protein complexes. An electron microscope is a complex system of electromagnetic lenses and apertures whose electric and magnetic fields can be changed to control their ability to deflect and focus an electron beam (Orlova and Saibil 2011). The simplified schematic of the electron microscope in fig 3.18 shows that the lens-aperture systems can be broadly divided into three sections- the condenser, the objective, and the projector. Of these, the sections that make a crucial contribution to image formation are the condenser and the objective, which are located immediately before and after the specimen respectively. The

condenser system serves to collimate the electron beam before it illuminates the sample, while the electrons exiting the specimen are focussed by the objective lens to form the image at the image plane (fig 3.18). The projector serves to magnify the image that is produced by the objective, to enable visualisation on a screen or detector system.

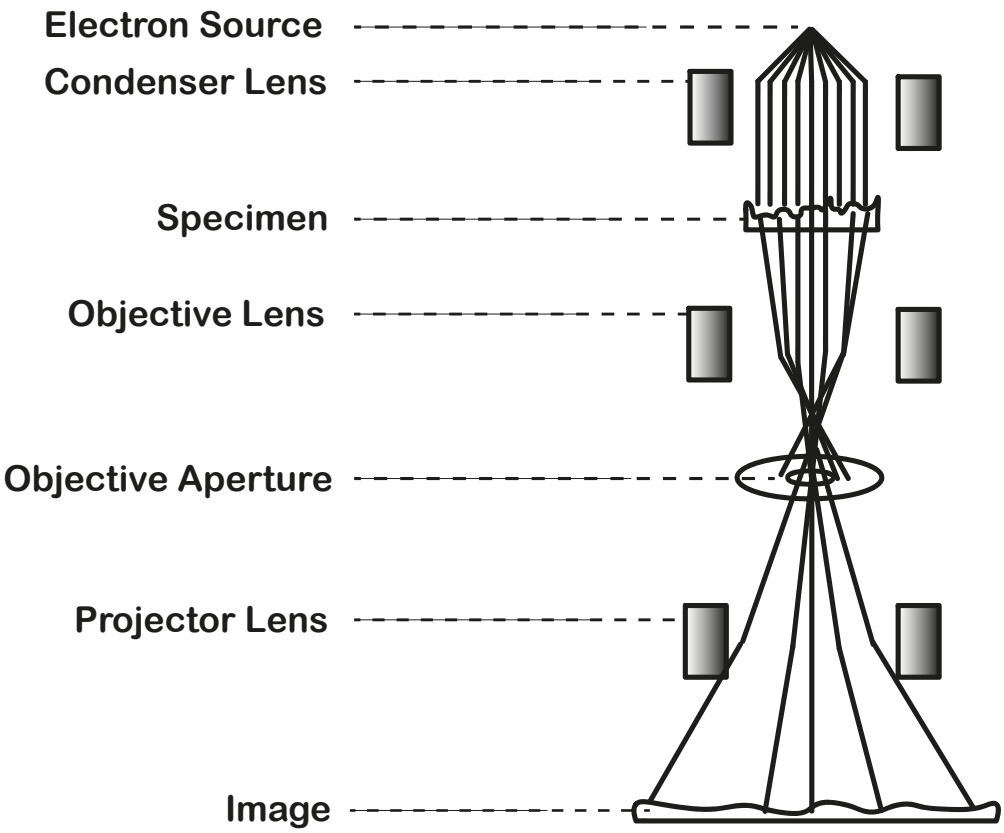


Figure 3.18 Schematic of Transmission Electron Microscope (adapted from Orlova and Saibil, 2011)

A typical transmission electron microscope comprises a source of electrons (electron gun) followed by a system of electromagnetic lenses and apertures that enable the collimation, deflection and focussing of the electron beam to form an image of the thin-film specimen.

The interaction between the specimen and the incident electrons can produce a variety of outcomes such as scattering in the forward direction, elastically or inelastically, at low angles or angles that are high enough to be blocked by the

aperture, as well as backwards scattering. In addition, the incident electrons can cause ionisation of the sample to generate secondary electrons, X-ray emissions or even bond rearrangements (Glaeser and Taylor 1978). However, only the elastically scattered electrons carry useful information about the features of the sample, that manifest as details in the image formed by the intersection of these electrons at the image plane (fig 3.19). The inelastically scattered electrons are not informative and are a source of noise in the images. Furthermore, the scattering only accounts for 10% of the electrons that are incident on the sample, implying that unscattered electrons also participate in image formation. In addition to the difference in number between the different types of electrons forming the image, the wave nature of electrons also influences the transfer of information from the object to the detector (De Broglie, 1924). In other words, an electron micrograph of a macromolecular sample is the result of the interference between electron waves that were scattered by the sample. Thus, the contrast generated in the image is a function of the particle-like and wave-like characteristics of the electrons interacting with the sample.

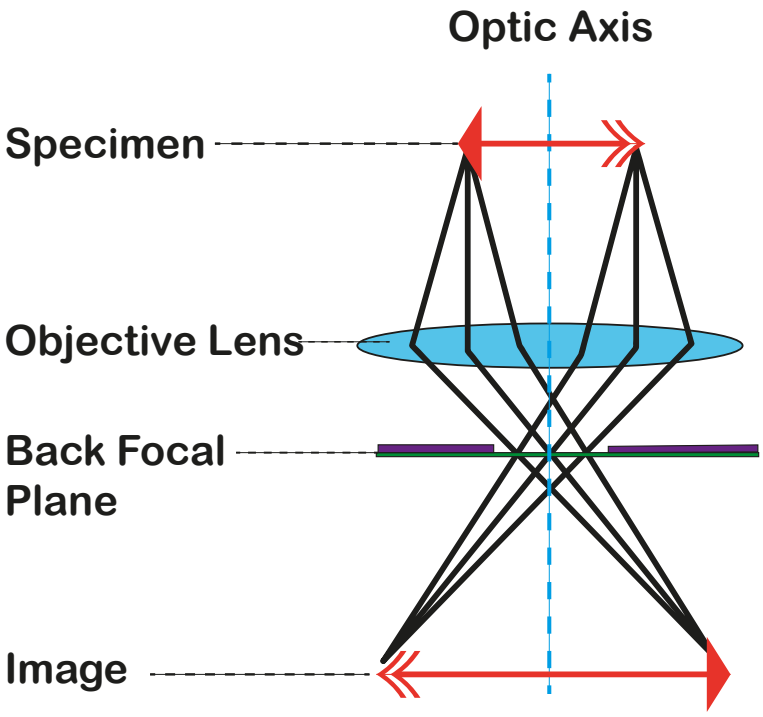


Figure 3.19 Image Formation by Objective Lens

Simplified ray diagram shows the focussing of the scattered rays arising from the object (red arrow, top) through the back focal plane to form the image (red arrow, bottom) at the image plane.

3.3.2 Contrast- the essence of EM

In the words of Helen Saibil and Elena Orlova, "an image represents variations in intensity caused by regional variations in specimen transmission" (Orlova and Saibil 2011). The ability to discern individual protein molecules in an image requires a difference in signal intensities between the molecules and the background. This variation in intensity is referred to as contrast and can be mathematically expressed as the ratio of the difference in intensities mentioned above to the average intensity of the entire image (Eq. 3.6) (Spence, 2003).

$$C_{im} = \frac{\rho_{mol} - \rho_{bkgd}}{\rho_{avg}} \qquad (3.6)$$

Where C_{im}- contrast in the image,

ρ_{mol} – Signal intensity from the molecule,

ρ_{bkgd} – Signal intensity from the background and ρ_{avg} – average signal intensity of the image.

3.3.2.1 Amplitude Contrast

In the simplest case, these variations arise from the absorption as well as back scattering of electrons from some regions in the specimen, compared to the effective transmission of other regions (fig 3.20). This amounts to a difference in the number of electrons reaching the detector from various parts of the specimen and is referred to as amplitude contrast.

However, biological macromolecules do not represent the above scenario. Composed of light atoms such as hydrogen, carbon, nitrogen and oxygen, these macromolecules are weak scatterers of electrons i.e., they are nearly electron transparent. As a result, majority of the electrons that illuminate the molecules are transmitted, suffering nothing more than a minor deflection in their paths. Yet,

one can still detect vague hints of molecular features in the image. This contrast is the result of the path difference between the unscattered and scattered electrons that causes a time-lag, or phase difference, between them, thereby contributing to phase contrast.

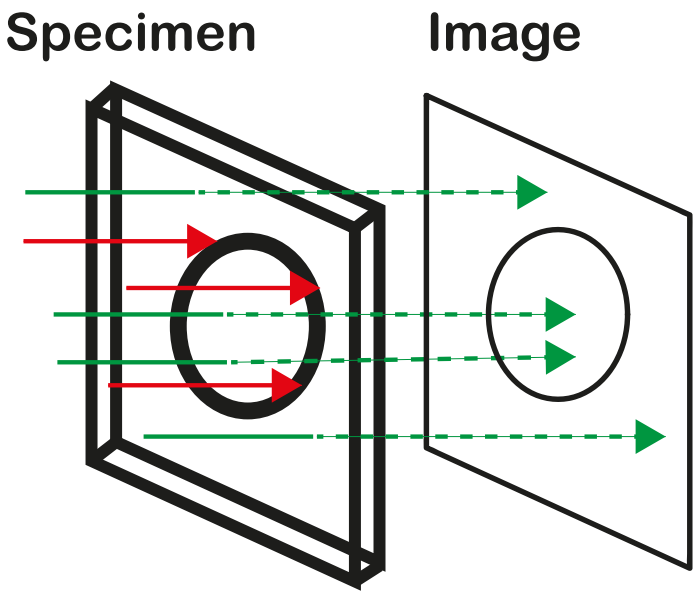


Figure 3.20 Amplitude Contrast

Formation of an image due to the absorption (red) of electrons by the dense region, like the ring in the specimen, combined with the transmission (green) by the light regions. This creates amplitude contrast in the image.

3.3.2.2 Phase Contrast

As shown in fig 3.21, phase variations between interfering waves manifest in the amplitude of the resultant wave i.e., the phase contrast is converted to amplitude contrast (Erickson and Klug, 1970). For instance, a 0° phase difference between interfering waves results in a wave whose amplitude is the scalar sum of the individual amplitudes (fig 3.21 a), while the scenario of a 180° phase difference leads to a scalar difference of amplitudes (fig 3.21 b). The phase shift caused by the interaction of the electrons with the specimen can be deduced mathematically as follows. If ψ_u denotes the unscattered wave function and ϕ denotes the total

phase shift that results from specimen interaction, then the scattered wave function ψ_s can be represented by Eq 3.7:

$$\psi_s = \psi_u \,.\, e^{i\phi} \qquad (3.7)$$

$$\psi_s = \psi_u \,(1 + i\phi) \qquad (3.8)$$

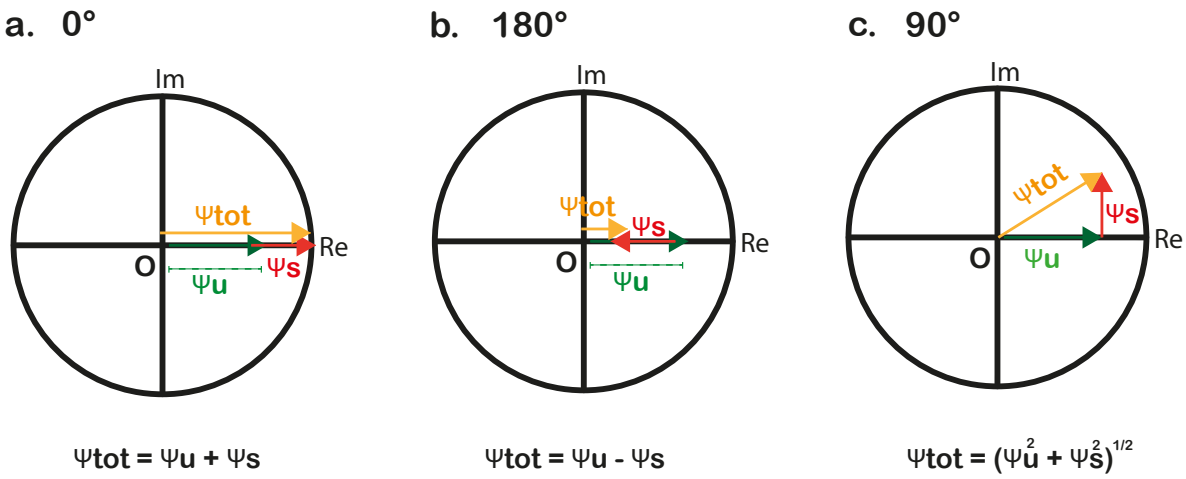


Figure 3.21 Phase Contrast

The interference between the scattered "ψ_s" and unscattered "ψ_u" electron waves yields resultant waves "ψ_{tot}" of different amplitudes depending on a phase difference of (a) 0º, (b) 180º and (c) 90º between the waves.

The exponential term can be expanded into a Taylor polynomial, whose higher order terms can be neglected due to the weak scattering contributed by the biological specimen. As a result, the scattered wave is a complex number whose real part is the unscattered wave and imaginary part is the phase-shifted version of the wave (Eq 3.8). This implies that they are mutually perpendicular in an argand plane or vector space. The implication of this weak-phase approximation is that the scattered wave is shown to undergo a 90° phase shift compared to the unscattered wave (Orlova and Saibil 2011). The interfered wave resulting from a

90° phase difference gives rise to a relatively small modulation in intensity (fig 3.21 c), thereby resulting in the weak contrast observed for such samples.

3.3.3 Contrast Enhancement

For reasons discussed above, biological macromolecules are termed weak-phase objects and show poor contrast in TEM images. Thus, the aim of single particle EM is to improve their visibility to reliably extract structural information from them. There are many ways that this can be achieved.

3.3.3.1 A beneficial defect

One of the most common ways of overcoming the poor phase contrast issue makes use of the inherent limitations of the objective lens system- its sphericity (Scherzer, 1936). The focus of a lens is defined as the single point of intersection of all rays that pass through the lens, based on the mathematical description of a focus associated with a parabolic surface (fig 3.22 a). However, electromagnetic lenses cannot be designed with this property and are unable to focus all rays at a unique focus (fig 3.22 b). This deviation from parabolic behaviour is known as spherical aberration. As a result, the high-angle scattered electron beams are focussed closer to the lens than the low-angle beams (fig 3.22 b).

Spherical aberration increases the path difference and therefore, the phase difference between the electron waves at the image plane, as shown in fig 3.22 b. For a scattering angle θ, the phase shift caused by spherical aberration ϕ_s can be expressed as follows:

$$\phi_s = \frac{1}{4} C_s \, \theta^4 \qquad (3.9)$$

where Cs is the coefficient of spherical aberration. The fourth power of scattering angle in Eq 3.9 indicates that this is a higher order aberration, which can only

contribute to a significant difference in the phase contrast of the image at high resolutions of 5 Å and beyond (Wade, 1992).

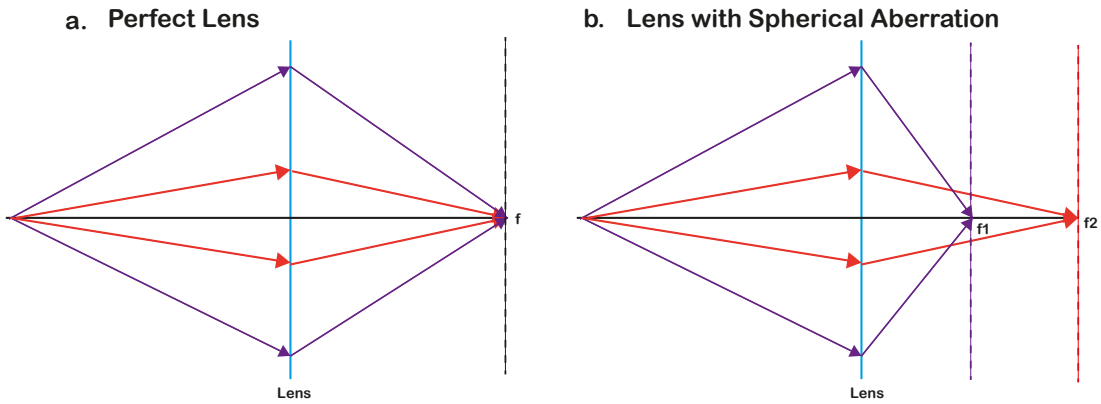


Figure 3.22 Spherical Aberration in Lenses

(a) A perfect lens would focus rays coming from different angles into a single focal point on the image plane; (b) Lens systems in typical electron microscopes display a spherical aberration which results in the differential focussing of rays coming from varying angles along the optical axis, which is a defect known as spherical aberration.

3.3.3.2 Imaging out of focus

In contrast to higher order aberrations, a more immediate and effective cause of phase shifts between the scattered and unscattered electrons involves deliberately operating the microscope out of focus (Orlova and Saibil 2011). Increasing or decreasing the current to the objective lens allows the electrons to be focussed closer to the lens or further away from it respectively, as shown in fig 3.23. The former situation, typically referred to as an over-focussed lens (fig 3.23 a), leads to the electrons reaching the detector after their paths have intersected, while the latter results in an under-focussed lens (fig 3.23 c) wherein the electrons reaching the detector prior to the intersection (Spence, 2003). The resultant deflection in the path of the electrons can be used to increase the path difference and therefore, the phase difference between them as well.

Traditionally, cryo-electron micrographs are collected under focus, where the protein particles appear black compared to a lighter background.

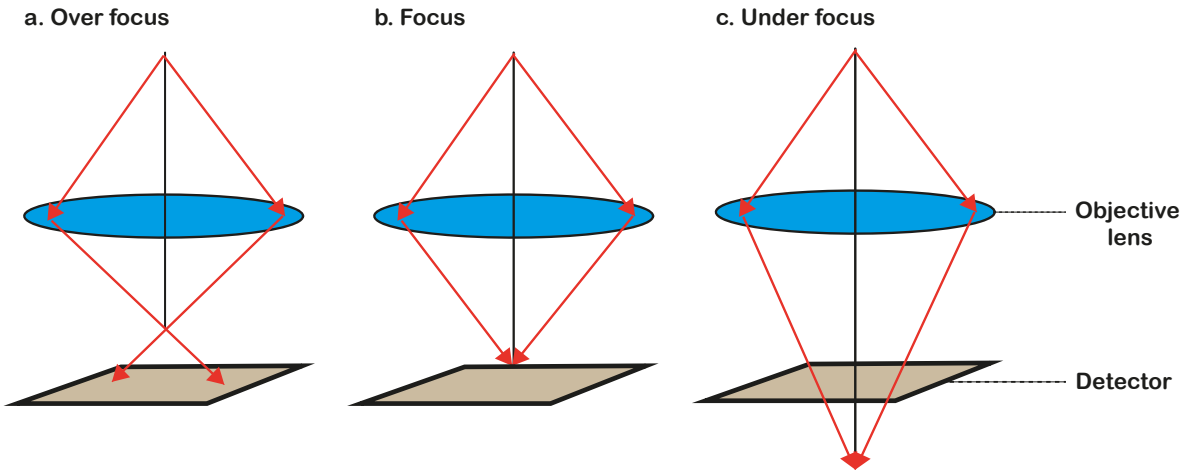


Figure 3.23 Applied Defocus of the Objective Lens

Changing the power of the objective lens can lead to an (a) over focussed lens wherein the electrons reach the detector after being focussed, (b) focussed lens and an (c) under focussed lens where the electrons reach the detector prior to being focussed.

Cryo EM typically employs defocii between -0.5 and -3 µm where the minus sign is used to denote an under-focussed lens, while negative stain EM applies defocii of -1 to -2 µm. If the defocus applied is denoted as Δz, then the phase shift ϕ_d due to the defocus can be mathematically represented as follows:

$$\phi_d = \frac{1}{2}\Delta z\, \theta^2 \qquad (3.10)$$

The combined effects of the phase shifts contributed by spherical aberration and defocussing the objective lens results in varying levels of contrast for different spatial frequencies. Specifically, both ϕ_s and ϕ_d are arguments of a sine function shown in Eq 3.11 which describes the transfer of contrast from the object to the image i.e., a Contrast Transfer Function (CTF) (Frank, 2006). In addition, a third contribution is the amplitude contrast that results from the minimal amount of absorption and back-scattering of electrons by the specimen (Fan,

Zhao et al. 2017), which can be accounted for by an additional term "-ϕ_a" (Eq 3.11).

$$CTF \approx \sin\left[\left(\frac{2\pi}{\lambda}\right)\left\{\frac{1}{4}C_s\,\theta^4 + \frac{1}{2}\Delta z\,\theta^2 - \phi_a\right\}\right] \qquad (3.11)$$

The graph of the CTF in fig 3.24 a illustrates the sinusoidal variation of contrast transfer with spatial frequency which is denoted as the reciprocal of resolution. The negative Y-intercept shows the minor amount of amplitude contrast inherent to a cryo EM image, and the sinusoid is shown to oscillate more rapidly towards higher spatial frequencies.

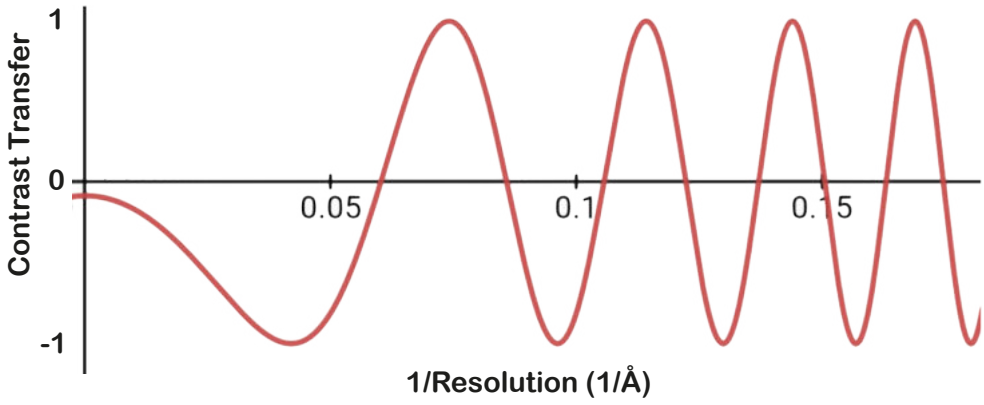


Figure 3.24 Contrast Transfer Function (CTF)

Graphical representation of the observed contrast in the TEM image of a biological macromolecule arising from operating the electron microscope under focus as well as higher order defects such as spherical aberration cause a sinusoidal variation of contrast in the image which rapidly oscillates at higher spatial frequencies; the level of contrast (Y-axis) is represented as a function of the spatial frequency which is the reciprocal of resolution (X-axis); the contrast level undergoes periodic reversal and shows the complete loss of information at certain spatial frequencies.

One of the major disadvantages of this CTF is the complete loss of information at the spatial frequencies where the curve has a value of zero i.e., zero-crossings

(fig 3.24). To overcome this, the sample is imaged at a range of defocii which results in sampling zeroes at different spatial frequencies, thereby enabling the recovery of information at nearly all frequencies. Another issue in the use of defocus is the loss of contrast at low frequencies such as 0.025 $Å^{-1}$ or less in fig 3.24. This implies a loss of detail at resolutions at and below 40 Å, compromising the ability to visually identify individual particles for structural analysis, especially for particles with a small size (Orlova and Saibil 2011).

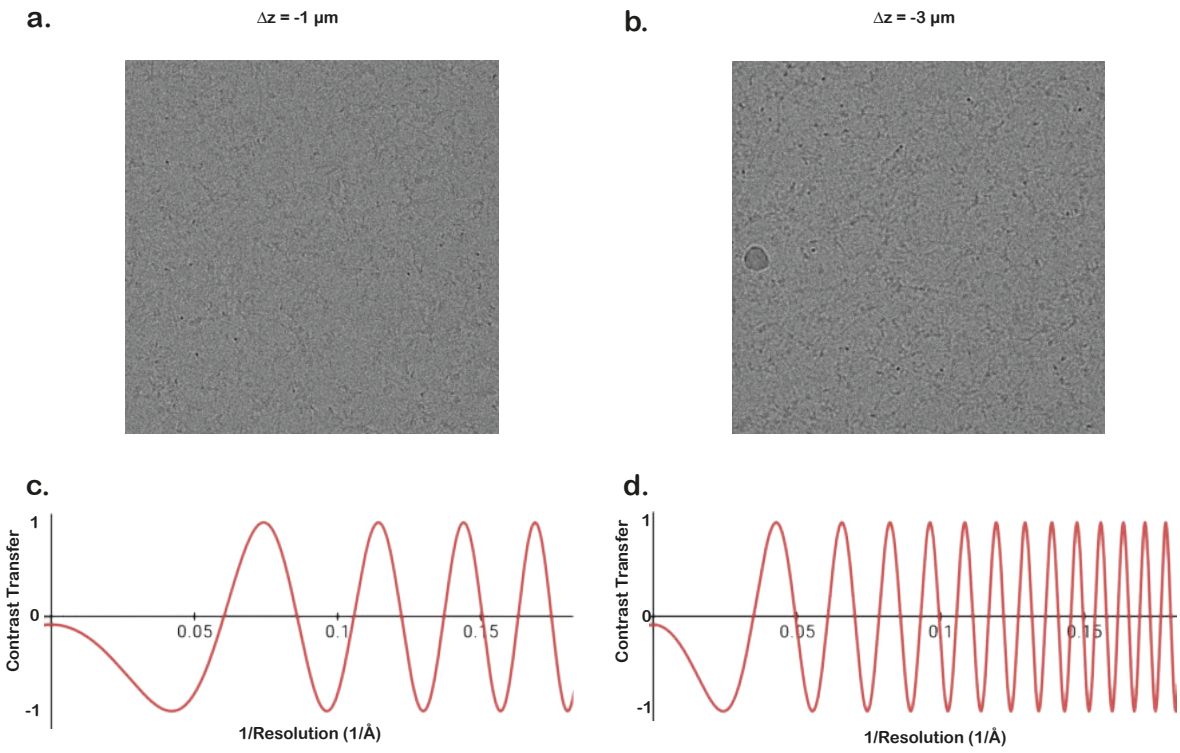


Figure 3.25 Effects of Defocus on CTF

The level of defocus alters the transfer of contrast from specimen to image, as demonstrated by the cryo-electron micrographs of ADP-HSET on Ultra-Au-foil grids acquired on the Glacios with a Cs of 2.2 mm and an amplitude contrast of 0.1, by applying defocii of (a) 1 µm and (b) 3 µm under focus; (c,d) the respective CTF curves obtained in each case, highlighting a more rapid oscillation of contrast with the application of a greater defocus.

Fig 3.25 shows the effect of an applied defocus on the CTF on a cryo EM dataset of ADP-HSET. An applied defocus of -3 µm shows a clear accentuation of contrast, improving the visibility of particles (fig 3.25 b) in comparison to the -1

μm of applied defocus in fig 3.25 a. However, the CTF curve captures the fact that the application of a greater defocus hastens the frequency of oscillations in fig 3.25 d, compared to fig 3.25 c. While this boosts the mid-range of spatial frequencies (~15-20 Å), it also severely limits the high-resolution details. This is the reason for EM datasets typically not being acquired at defocii over 3 μm.

3.3.3.3 Phase Plates

As mentioned above, operating the microscope further from focus trades information at high spatial frequencies for visibility, while imaging closer to focus does the exact opposite. An ideal scenario would involve generating sufficient contrast at all spatial frequencies while imaging at focus. This is achieved using a phase plate- a device that increases the phase difference between the scattered and unscattered electrons by ~90° (Nagayama and Danev 2008), theoretically resulting in perfect transfer of contrast in focus. Different types of phase plates have been synthesized for electron microscopy (Unwin 1971, Nagayama and Danev 2008, Schwartz, Axelrod et al. 2019), the most recent and commercially used variety being the Volta Phase Plate (VPP) (Danev, Buijsse et al. 2014). The VPP is a circular disc made of a thin carbon film which is placed in the back focal plane replacing the objective aperture. The illumination of a spot on the VPP by the central unscattered beam of electrons charges the film, resulting in the development of a beam-induced surface potential that retards the scattered electrons to varying degrees. In mathematical terms, the VPP includes a fourth term ϕ_p to Eq 3.11 (Fan, Zhao et al. 2017), which has the effect of increasing the contrast transferred across the lower spatial frequencies to a constant non-oscillating value, until the higher order effects appear from a resolution of 10 Å (0.01 $Å^{-1}$) to resume the oscillatory nature of the CTF, as illustrated in fig 3.26 b. The graphs of the theoretical CTFs in fig 3.26 demonstrate the staggering difference between the typical contrast generated using a regular objective aperture (fig 3.26 a) and a VPP (fig 3.26 b).

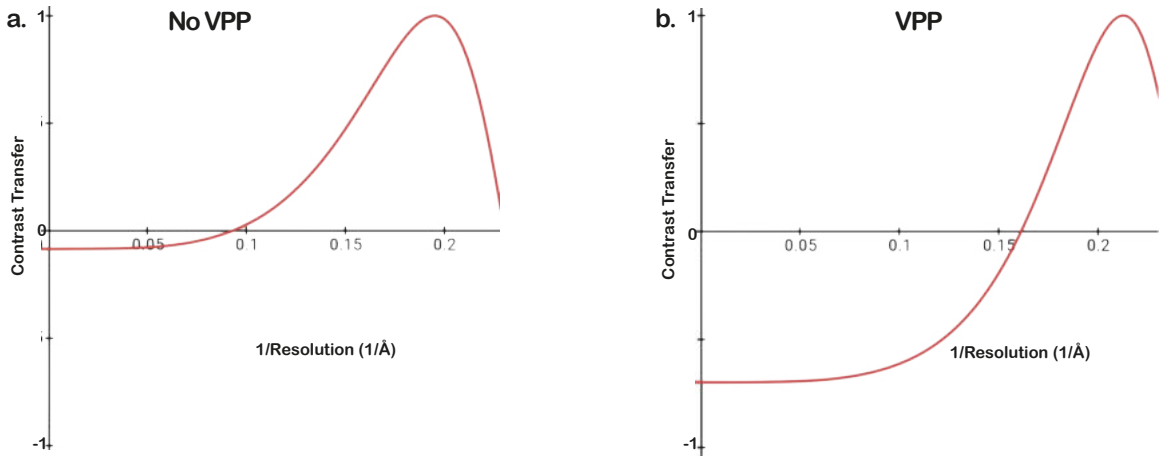


Figure 3.26 Effect of the Volta Phase Plate on In-Focus Data Acquisition

Theoretical CTF curves from a 200 kV microscope with a Cs of 2.2 mm using (a) an objective aperture of 100 µm and (b) a VPP, demonstrate the drastic difference in amplitude contrast contributed by the phase plate at lower spatial frequencies.

In practice, as it is difficult to accurately focus the microscope, the in-focus acquisition method is not employed. Instead, the VPP is also operated under focus to obtain an oscillating CTF. However, due to the enhanced contrast generated by the Volta potential, the defocii employed are typically between 0.5 and 0.7 µm (Danev, Tegunov et al. 2017), which is significantly closer to focus compared to the 1-3 µm range employed without a VPP.

3.4 Negative Stain Electron Microscopy- Data Processing

Negative stain EM is a high-contrast technique performed at room temperature, that provides a rapid means of qualitatively analysing the size, shape, and heterogeneity in a sample. The staining procedure involves embedding the sample in a shell of heavy-metal salt solution such as uranyl acetate (Ohi, Li et al. 2004). As heavy metals are excellent scatterers of electrons, this results in a reversal of contrast in the images with the protein molecules appearing white on a black background.

The overall pipeline I had adopted for setting up a data collection run proceeded in a similar fashion for both negative stain EM and cryo EM. First, the grid was surveyed by acquiring several images of consecutive areas at a magnification of around 210x and tiling them together to create a grid-map i.e., an atlas of the grid. This was used to evaluate the level of staining in negative stain EM or hydration in cryo EM and pick the squares that are well-stained or well-hydrated respectively. Following this, the stage was moved to grid-squares of interest to image the sample by the low-dose method using three modes- search, focus and exposure (Chen, Sachse et al. 2008). Initially, the Z-height of the stage was adjusted in search-mode that used a magnification of around 11,000x. The height was adjusted to achieve eucentricity, which is the condition wherein the specimen would remain stationary despite tilting the stage. The beam was then focussed away from the area to be imaged to prevent radiation damage, using the focus-mode magnification of 150,000x. Subsequently, the foil hole was imaged by defocussing the objective lens by -1 to -2 µm at the exposure-mode magnification of 50,000x, giving rise to a pixel size of 1.73 Å. This procedure was repeated for all selected foil holes within a square as well as every desired grid-square to accumulate a dataset.

3.4.1 Visual Characterisation of FL HSET variants

My first application of negative stain EM was to visually inspect the purified samples of His_6-FLHSET, $Strep_2$-FLHSET-I and $Strep_2$-FLHSET-II and understand the appearance of HSET molecules in electron micrographs. This

had the added benefit of evaluating the efficiency of the three purification strategies. For this experiment, I prepared grids from each of the three samples that were diluted to identical concentrations of 0.2 µM (see section 2.3.1, *Page 41*). The images were collected manually on the TF20 microscope using the EM-Menu software. Fig 3.27 a, b, and c show the best micrographs obtained for each sample.

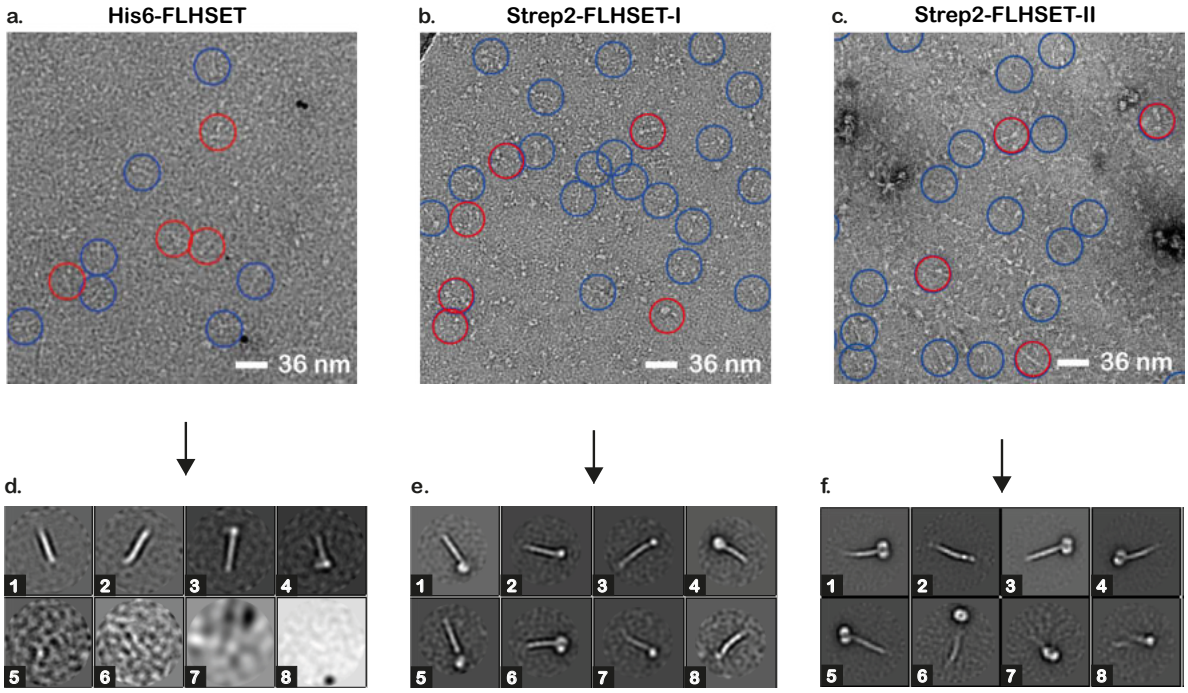


Figure 3.27 Visual Characterisation of FL HSET by Negative stain EM

(a,b,c) Representative negatively stained micrographs of Apo-HSET samples from the three purification strategies; manually picked HSET particles are highlighted in blue circles, off-target molecular species are highlighted in red circles. (d,e,f) 2D class averages of the particles from the respective datasets in RELION 3.1 provide the first visual confirmation of the existence of HSET in the samples.

Despite being at the same final concentration prior to application on the grids, each micrograph shows a significant difference in the particle density and contrast. From these micrographs, I identified and manually picked particles (blue, fig 3.27) that appeared similar to a ball (motor region) attached to a stick

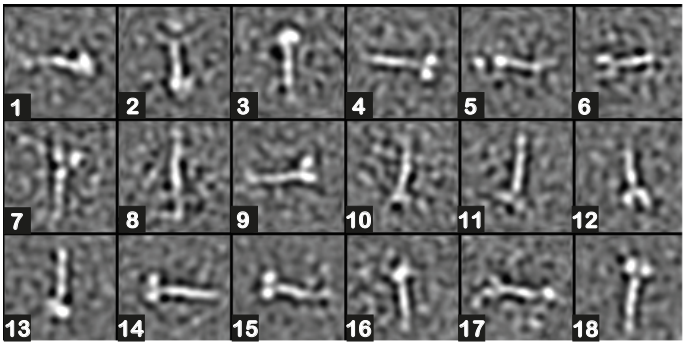
(stalk), matching the model described by early EM of a kinesin-14 (Hirokawa and Noda 2008). I used a picking diameter of 300 Å to encircle the particles in order to encompass the maximum length up to which I could observe signal associated with the particle. The small size and narrow, elongated shape of the particles made this task non-trivial despite attempts at increasing the scale of the images. Moreover, one can get distracted by a significant number of particles that either resulted from overlap between two or more HSET particles or aggregated HSET assemblies, termed as off-target species (red, fig 3.27). The poor contrast in the images of His_6-FLHSET as well as the crowded distribution in $Strep_2$-FLHSET-II further compounded the problem of recognising and avoiding off-target particles in both samples. Nevertheless, I managed to compile a dataset of 6000 manually picked particles per sample that I proceeded to classify in 2D.

The intent behind 2D class averaging was threefold, the first was to identify and isolate images of off-target species to clean the dataset. The second involved the identification homogeneous subsets of particles based on similarities in out-of-plane orientation, which correspond to individual views in 3D. The third was to align and average the images within each class to reinforce the signal from each particle and create a class average that provides enhanced visibility of the particles (van Heel and Frank, 1981). Using RELION's algorithm, I was able to get a 2D glimpse of FL HSET at better contrast. The figures 3.27 d, e, and f display the class-averages in descending order of particle count- typically ranging from 1000 images in the top-left to 50 in the bottom-right, which manifests as a similar descending trend in signal to noise ratio of the averages. While the number and quality of the class averages varies between datasets, the hint of a dumbbell-shaped motor domain attached to a long stalk was the first visual proof of the overall structure of HSET. The stalk tends to curve and a bending between the stalk and the motor domains can be observed (fig 3.27 f). The success of a 2D classification is significantly affected by the contrast and level of noise in the molecular images (Penczek, Radermacher et al. 1992). It is thus not surprising that His_6-FLHSET, with poor contrast, had a lower yield of successful class averages that showed fewer features compared to the $Strep_2$-tagged samples. In addition, RELION's algorithm results in an uneven distribution of particles among the classes, creating a few densely populated (top row) and several scarcely

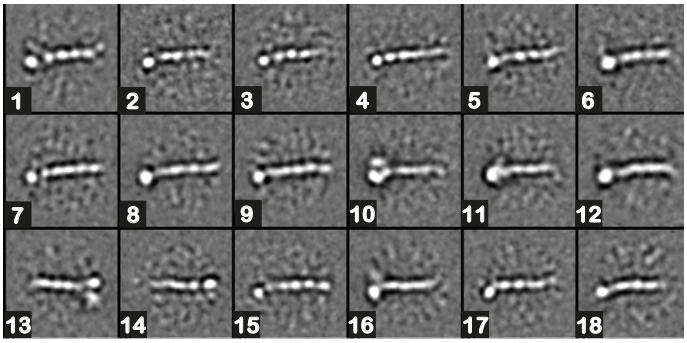
populated classes. This has the dual effect of compromising the total yield of classes as well enabling intra-class heterogeneity, as will be explained further in chapter 4.3 (*Page 171*). Thus, while RELION's classification established that HSET could be visualised, the appearance of the classes did not contain sufficient detail to properly evaluate the success of particle identification during the manual picking procedure.

Therefore, I experimented with a different classification algorithm called ISAC2, from the SPHIRE suite. In contrast to RELION, ISAC implements a more thorough approach to aligning and classifying images to tackle both the problems that plague RELION's classes, which is also explained in chapter 4.3 (*Page 171*). As a result, I obtained a greater number of classes in which HSET could be distinctly recognised for all three datasets (fig 3.28), including the low-contrast His_6-FLHSET. The ISAC classes also revealed better recovery of the full length of HSET in each case. The clarity as well as the length of HSET displayed in the class averages improved progressively from the His_6-FLHSET to $Strep_2$-FLHSET-II, as a function of improving contrast. The accuracy of alignments in the case of $Strep_2$-FLHSET-II was high enough to reveal the bend in the neck region connecting the stalk to the motor domain and what might be a preferential proximity of the stalk to one motor domain over the other (classes 1, 7, 8 and 15, fig 3.28 c). Furthermore, one can observe the start of the N-terminal Cargo-Binding Domain (CBD) on more than one occasion in the same dataset as indicated by the arrows in classes 4, 6 and 9, in fig 3.28 c. Overall, ISAC established the success of all three purification strategies through the unambiguous visualisation of HSET in all datasets and demonstrated that the manual picking was largely devoid of off-target species. The success of the algorithm is evident from the recovery of classes that show the first glimpse of all three FLHSET domains. Moreover, one can also observe multiple orientations of HSET based on the change in relative positions of the motor domains, implying an ability to proceed to a 3D analysis of the sample. Based on the level of detail in the classes, I concluded that $Strep_2$-FLHSET-II would be the most suitable for further structural analysis.

a. His6-FLHSET



b. Strep2-FLHSET-I



c. Strep2-FLHSET-II

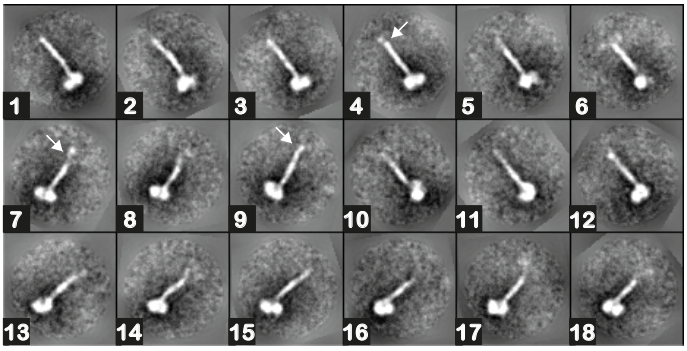


Figure 3.28 ISAC2 Serves as an Effective Classification Tool for FL HSET

Successful 2D class averages of Apo-HSET obtained by classification of the manually picked particles from the datasets corresponding to the three purification strategies- (a) His_6-FLHSET, (b) $Strep_2$-FLHSET-I and (c) $Strep_2$-FLHSET-II, using ISAC2 algorithm in the SPHIRE suite. Arrows in classes point at the N-terminal Cargo-binding domain.

3.4.2 Low-resolution structure elucidation of FL HSET

After I established a strategy to consistently obtain FL HSET in high yield and purity, I attempted to solve the structure of $Strep_2$-FLHSET-II in its Apo as well as

ADP-bound forms, using negative stain EM. My aim in doing so was to visualise FL HSET in 3D and investigate any differences in the structural and conformational characteristics of its Apo and ADP-bound forms. Therefore, I continued to improve the particle density of both forms of $Strep_2$-FLHSET-II on grids by increasing the concentration of protein as well as ensuring a minimum requirement of glycerol and salt, as established by mass photometry, (section 3.2.5.3, *Page 73*) while diluting the sample. I proceeded to set up and carry out automated data collections on the TF20 microscope using the EM-Tools acquisition software. The data-processing detailed below presented the opportunity to explore and optimise several parameters in the pipeline that would enable me to reliably extract information about a target as challenging as HSET.

3.4.2.1 Particle picking

The negative stain EM grids of Apo and ADP-HSET yielded datasets of 600 micrographs each, which showed identical levels of contrast and particle density. The samples were broadly similar in appearance with no detailed differences visible at this stage of the processing. From the different pickers available, I opted for CRYOLO v1.5.6 from the SPHIRE suite, which uses an approach based on neural networks to increase the accuracy of particle picking. The flow-diagram in fig 3.29 delineates my approach to train a pre-existing general picking model for negative stain datasets in CRYOLO, from a few manually picked particles. In the 12 micrographs that constituted the training set, I picked around 10-20 isolated particles per micrograph, sampling as many distinct views of HSET as possible (fig 3.29 a). Particle identification was aided by the boost in contrast from CRYOLO's low-pass filter (fig 3.29 b and c), enabling the model to pick 100,000 particle coordinates of Apo-HSET and 75,000 of ADP-HSET.

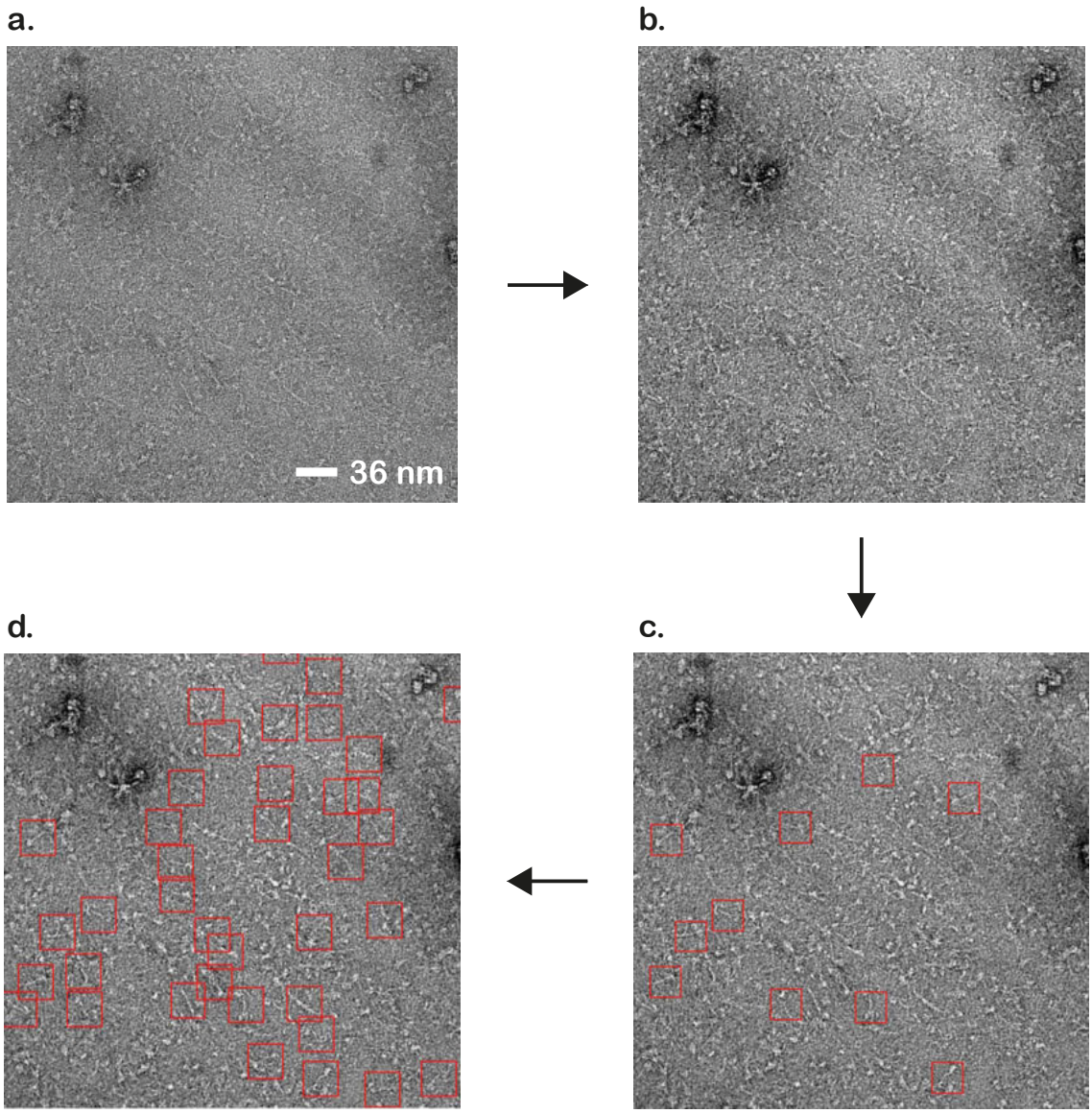


Figure 3.29 CRYOLO Pipeline for Picking Particles of FL HSET

Flow diagram shows the steps taken to create and trained model for particle-picking through CRYOLO v1.5.6; both Apo and ADP-HSET use the same picking model. (a) Unfiltered negative stain electron micrograph of Apo-HSET is subjected to (b) CRYOLO's low-pass filter to improve contrast in the image and aid (c) manual picking of a few particles by enclosing within red boxes of size 300 Å to comprise a training set; the training set is used to train a pre-existing model to recognise HSET particles, (D) the newly trained model is made to auto-pick more particles from all micrographs.

Furthermore, the degree of similarity between the datasets of Apo and ADP-HSET implied that the same picking model could be used for both datasets,

thereby reducing the time spent on manually picking particles and training models. This visually confirms the fact that Apo-HSET is just as stable and usable for EM studies as ADP-HSET, contrary to the predictions of TSA (section 3.2.2.2, *Page 63*), and consistent with the findings of the Kratky plot in SAXS (section 3.2.6.3, *Page 80*). The auto-picked coordinates were extracted in RELION with a box size of 440 Å after down sampling by a factor of two for faster downstream processing, resulting in a 128-pixel boxes with a pixel size of 3.46 Å.

3.4.2.2 2D Classification

My next goal was to classify the extracted HSET particles in 2D to obtain a homogeneous stack of particles for a 3D analysis. To achieve this, I had to optimise the classification procedure based on two factors- the efficiency of the available algorithms and the tuneable parameters in the most efficient algorithm. Initially, I tested classifiers from three different software suites on 30,000 particles extracted from a small subset of Apo-HSET micrographs that were chosen at random. The classifiers differed significantly from each other as discussed in detail in chapter 4.3 (*Page 171*).

Based on fig 3.30, the classes of $Strep_2$-FLHSET-II from ISAC provide the most complete and clear picture of the dimer in comparison to the other classifiers, due to its stringent and exhaustive approach (Yang, Fang et al. 2012). Moreover, ISAC also shows the ability to finely sample angular rotations about the axis of the stalk which are revealed in the arrangement of the classes (fig 3.30 a). However, it is precisely because of its stringent and exhaustive nature that it proved to be the most time-consuming and computationally expensive technique of the three. The time required to run an ISAC job as well as subsequent 3D jobs increased exponentially with the size of the dataset. Attempts at GPU-based acceleration of ISAC jobs were plagued with inconsistencies between the results obtained from non-GPU (MPI-CPU) and GPU-based runs for the same datasets. Thus, inconsistent results, time-consuming runs, and an inflexible file format that made it tedious to migrate between different data processing suites made ISAC an unproductive resource. Therefore, I had to stop using ISAC and SPHIRE's entire processing package.

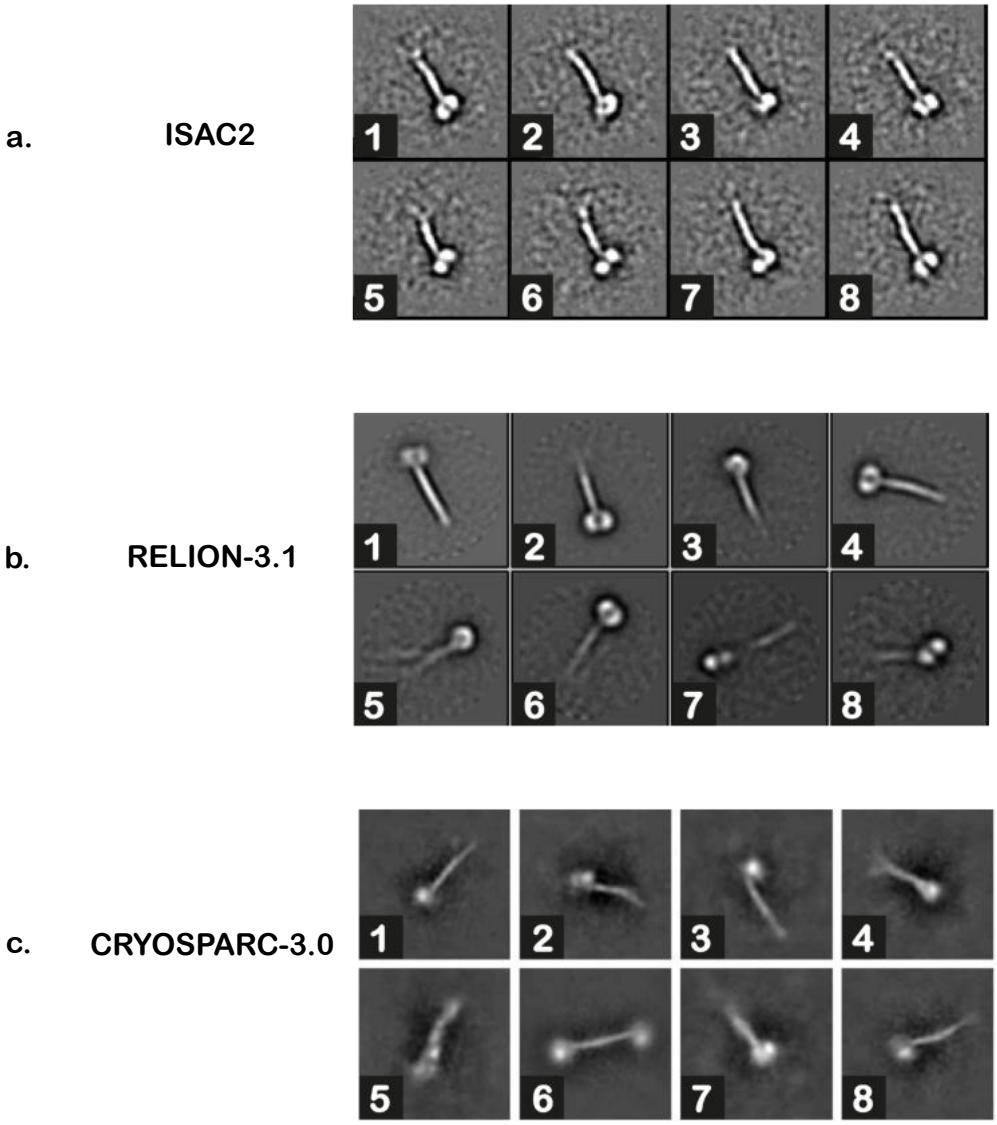


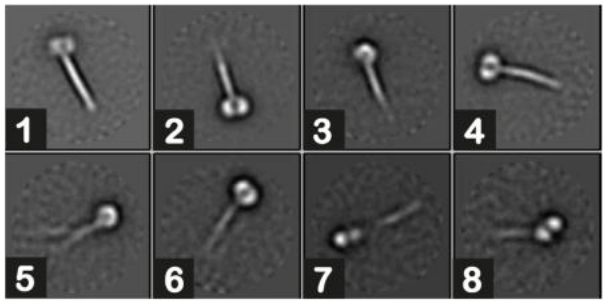
Figure 3.30 Comparing Classifiers

2D averages showing the highest detail from respective reference-free classifications of 30,000 Apo-HSET particles in (a) ISAC2, (b) RELION v3.1 and (c) CryoSPARC v3.0.

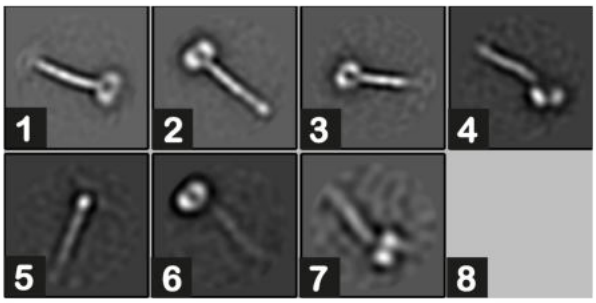
Among the three classifiers, cryoSPARC had by far the shortest runtime of 12 minutes for 30,000 particles, and yielded classes that managed to align a significant extent of the coiled coil domain (fig 3.30 c). Initially, I considered cryoSPARC as the ideal alternative to ISAC. However, the loss of information at

the motor domains towards the periphery of quite a few classes was concerning and made it difficult to distinguish between out-of-plane views. In comparison, RELION, though not the best in terms of visual detail in the class averages (fig 3.30 b), managed to provide a few different orientations adopted by the motor domains. Moreover, unlike cryoSPARC, RELION provided the option of changing angular and translational sampling parameters, which was another important factor in my choice of classifiers. Whilst the default classification parameters for both cryoSPARC and RELION can effectively align and center a compact and globular target, they struggled to align HSET which is mostly narrow and elongated. RELION appeared to attempt centering the globular motor domains which crops the stalk and compromises the recovery of information from the N-terminal end. To ensure the accurate alignment and centering, as well as the adequate recovery of the full length of HSET, I had to significantly modify in-plane angular and translational sampling parameters, as shown in fig 3.31.

a. 6-degree angular sampling +
5-pixel translational search range



b. 6-degree angular sampling +
10-pixel translational search range



c. 3-degree angular sampling +
10-pixel translational search range

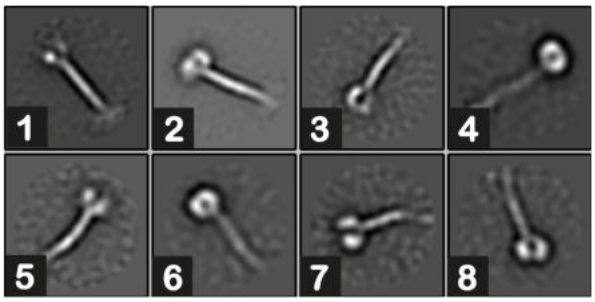


Figure 3.31 Alignment Optimisation in RELION

Representative 2D classes from reference-free classifications of 30,000 particles using different in-plane angular sampling frequencies and translational ranges for image alignment as follows- (a) 6-degree angular and 5-pixel translational, (b) 6-degree angular and 10-pixel translational, (c) 3-degree angular and 10-pixel translational; all cases used single-pixel steps for the translational searches.

For instance, a twofold increase in translational sampling range between fig 3.31 a and b resulted in classes showing a greater extent of the stalk, and an additional halving of the in-plane angular sampling frequency in fig 3.31 c highlighted its curvature. While a further reduction of angular sampling and increase in translational range by a factor of two might have improved the classes

further, the trade-off would have been an increase in runtime beyond a feasible limit. For this reason, I proceeded to classify the entire datasets of both apo- and ADP-HSET using a 3-degree angular sampling frequency and 10-pixel translational search range with a single-pixel search step.

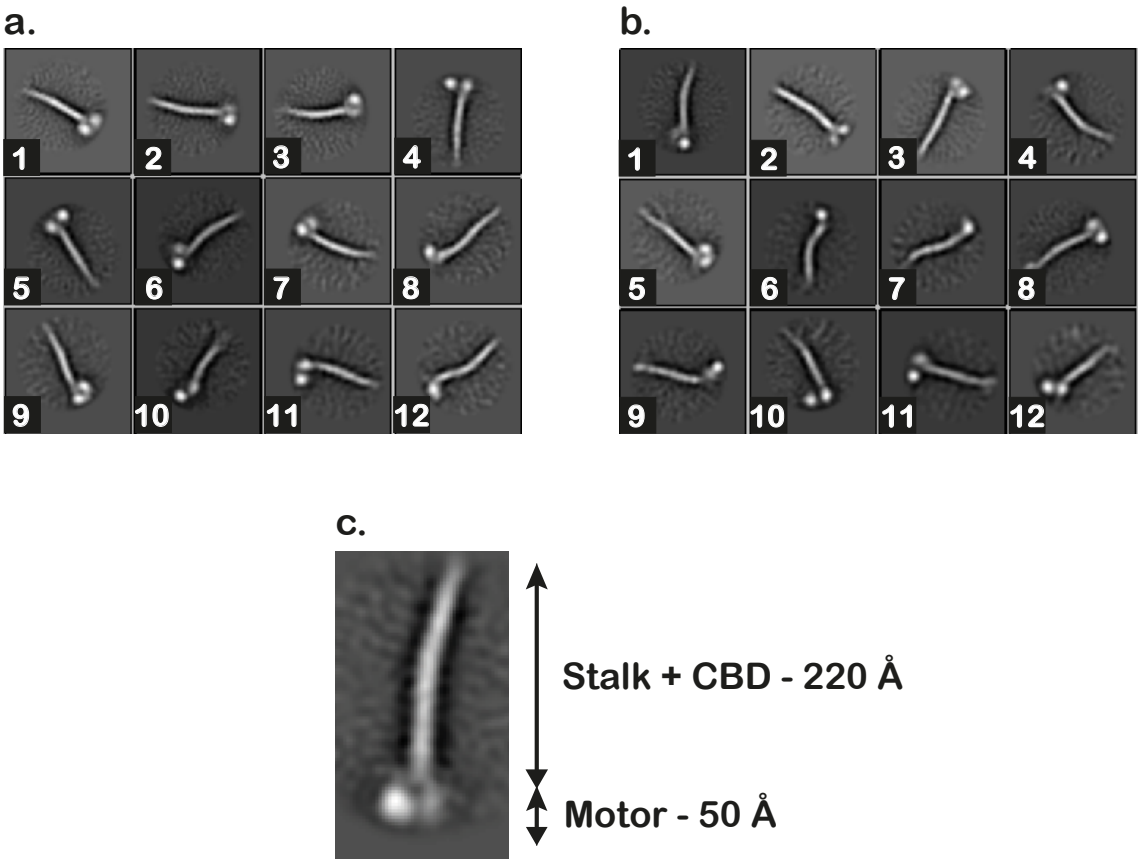


Figure 3.32 Reference-free 2D Class Averages of Apo and ADP-FL HSET

Reference-free 2D classification of entire negative stain EM datasets of (a) Apo HSET and (b) ADP-HSET using optimised parameters in RELION, classes reveal a large extent of HSET, features such as the curvature of the stalk and multiple views; (c) magnified view of an isolated class showing the measurements of the domains of HSET that are visible in the classes.

After a few rounds of selection and reclassification to clean-up the data, the resultant classes of both datasets showed an improved appearance of dimeric HSET up to 270 Å in length from top to bottom (fig 3.32 c). The classes also

appear to sample the rotation of HSET about the axis through the stalk, with views such as class 1 in fig 3.32 a that show the two motor domains positioned on each side of the stalk axis (front-views) and those like class 6 in fig 3.32 b that show both motor domains on the same side of the stalk, tending to eclipse each other (side-views). The asymmetry in the motor-neck connection was evident in the front-views of both forms of HSET (class 4, fig 3.32 a) (class 10, fig 3.32 b), some of which also showed a tilt of the motor domain to create an acute angle with the stalk (class 9, fig 3.32 a) (class 5, fig 3.32 b). This was an indication that HSET possessed a kink in its neck region between the motor and stalk domains, which was shown to occur in another member of the kinesin-14 family by single molecule studies (Endow and Higuchi 2000). One can also notice the curvature of the stalk in both forms of HSET (classes 2 and 6, fig 3.32 a) (classes 6 and 8, fig 3.32 b). The measurement of 220 Å corresponding to the stalk slightly exceeds the predicted value of 200 Å from the Alphafold model of HSET, suggesting the recovery of some portion of the N-terminal CBD (fig 3.32 c). However, one cannot discern the interface between the two regions at this resolution. Overall, one can clearly observe a significant number of unique views arising from the rotation of the HSET molecule about the stalk, which implies the availability of sufficient information to build its 3D ab-initio reconstruction.

3.4.2.3 3D Analysis

Having sampled a considerable number of out-of-plane views of FL HSET, I decided to use this information to build an ab-initio 3D reconstruction of the two forms. Rather than using all 34,000 Apo-HSET and 16,000 ADP-HSET particles that were obtained from the 2D analysis above, I used a subset of around 2,000 particles corresponding to unique views such as 1, 2, 8, and 11 in fig 3.32 a and 1, 8, 9, and 10 in fig 3.32 b. The reconstructions were generated without imposing symmetry, making use of the default parameters in RELION that are mentioned in the methods (see section 2.3.3.2, *Page 42*). The resulting models for both Apo and ADP-HSET, displayed in fig 3.33 a and b respectively, show a similar appearance and an overall resolution of 15 Å. A notable feature in the models is the overall length of ~300 Å, which is longer than the 2D class-averages by around 30 Å. The models encompass more of the stalk than the 2D classes and terminate with a hook-like structure, that might be a hint of the CBD. Before

extending this analysis to a greater number of particles by 3D classification, I used the ab-initio reference to extract more information about the views that were sampled in the population of Apo and ADP-HSET.

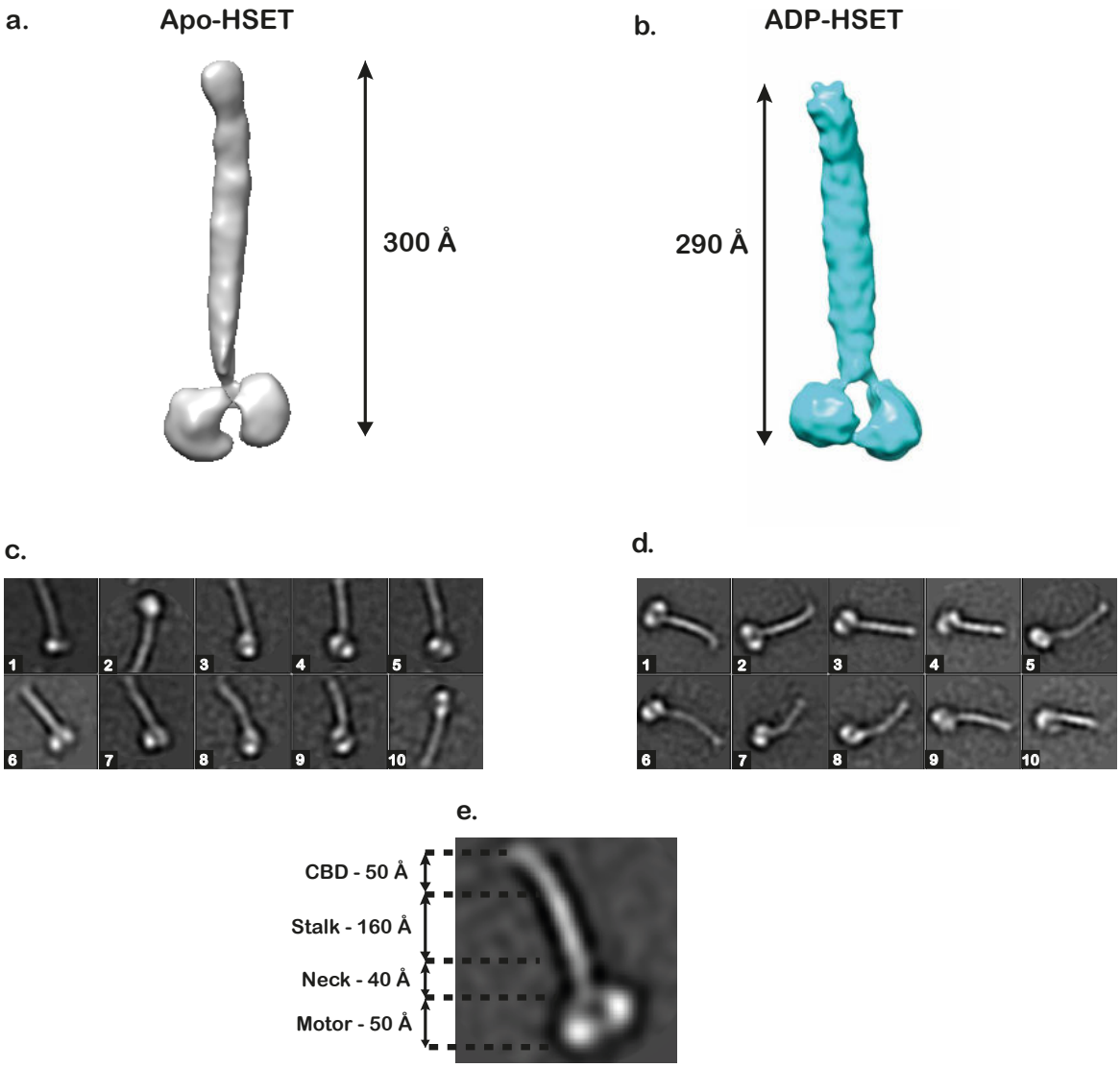


Figure 3.33 Reference-based 2D Classification of Apo and ADP-FL HSET

3D ab-initio reconstructions of (a) Apo HSET and (b) ADP-HSET at resolutions of 15 Å reveal close to 300 Å of the protein; models were used as references to align particles in 3D space and reclassify them in 2D without the application of in-plane alignments; the reference-based 2D classes were thus obtained for (c) Apo and (d) ADP-HSET, classes reveal further rotational sampling, (e) magnified version of isolated class provides additional insight into the domain architecture of HSET in comparison to the reference-free class averages.

I followed the strategy of reference-based 2D classification (fig 3.33 c and d) proposed by the group of Jan Löwe (Lee, Merkel et al. 2020), developed to salvage unique views of elongated proteins not visible in a reference-free approach. Based on the images in fig 3.33 c and d, the range of views represented in the classes is consistent with finer sampling about the axis of the stalk and improved recovery of the N-terminal of curved HSET, measuring 300 Å in total. On close inspection of the stalk, one can notice a bright density in the first 100-120 Å starting from the motor domains followed by a distal relatively dull appearance, that could correspond to the CBD (fig 3.33 e). The end point of the bright density could be the stalk-CBD interface, implying a length of around 200 Å for the stalk-neck region of HSET, which agrees with the Alphafold prediction. Aside from increasing the number of particles that could be used for 3D classification by approximately 5,000- 8,000, the reference-based approach also succeeded in unravelling previously unobserved detail in Apo and ADP-HSET.

I had now obtained datasets with 42,000 particles of Apo-HSET and 21,000 of ADP-HSET, which I classified into 3 and 2 classes respectively in 3D, using the respective ab-initio references and the optimised alignment parameters from the 2D classification. In the case of Apo-HSET, the particles were evenly distributed between two of the classes (fig 3.34 c and d), with the third one nearly unpopulated. On the other hand, ADP-HSET showed 19,000 particles in one of the classes (fig 3.34 e) and assigned the remaining 2,000 to the second class in which HSET's domains were improperly defined. Unlike the ab-initio models, the 3D classification showed significant differences between Apo and ADP-HSET. As fig 3.34 c, d, and e show, two distinct sub-populations of Apo-HSET have been recovered as opposed to a single population of ADP-HSET.

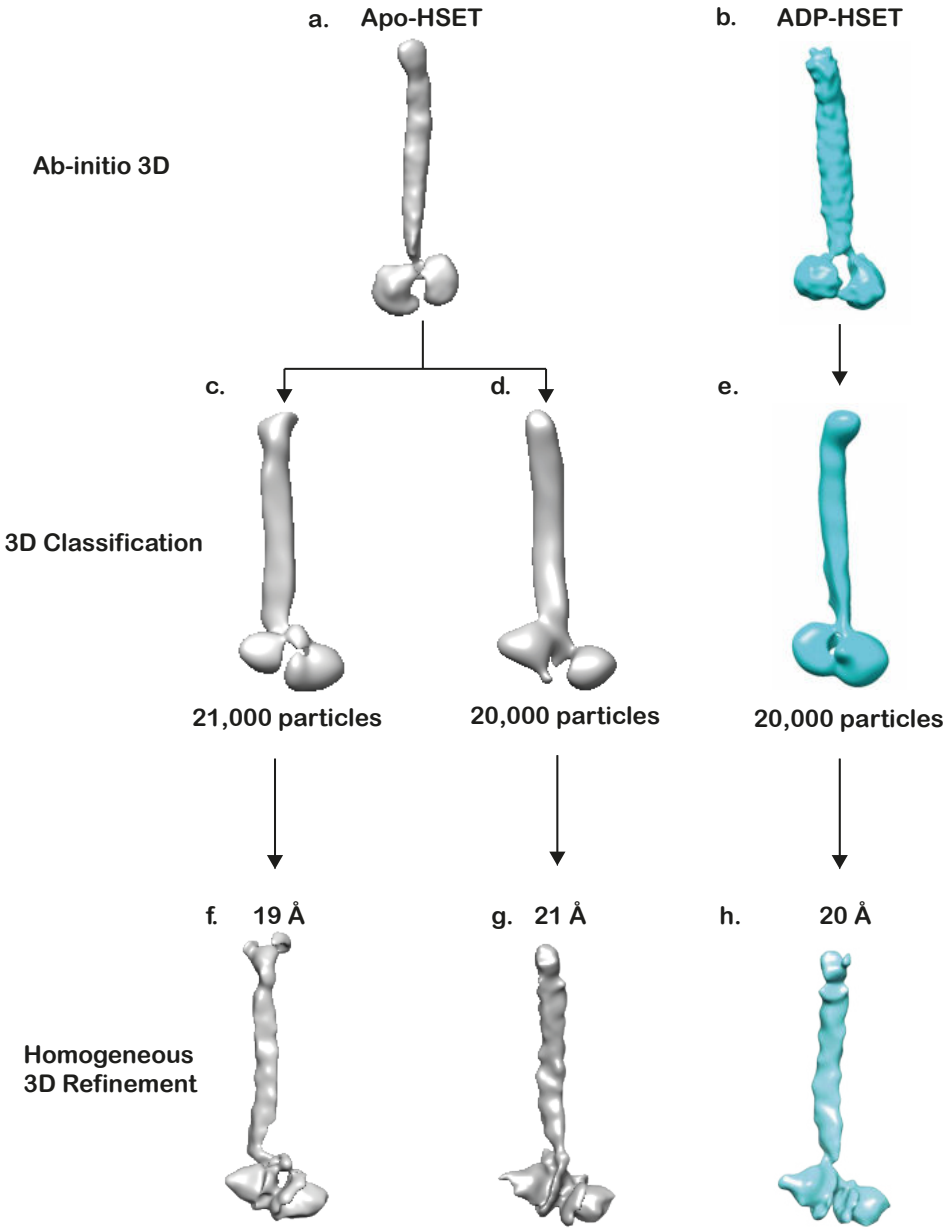


Figure 3.34 3D Classification and Refinement of Apo and ADP-FL HSET

(a,b) 3D ab-initio models of Apo-HSET and ADP-HSET from RELION 3.1. (c,d,e) Unsupervised 3D classifications of the Apo-dataset reveal two equally abundant conformational states of Apo-HSET; ADP-HSET shows a single 3D class, which appears similar to one of the conformers of Apo-HSET. (f,g) the 3D classes of Apo-HSET are refined in CryoSPARC v3.0 to a resolution of 19 and 21 Å; the refined maps highlight the unique arrangements of the motor domains and the motor-neck connections. (h) the map of ADP-HSET is refined to 20 Å; the single state resembles one of the conformers of Apo-HSET.

All classes are 300 Å long which suggests the inclusion of density from the CBD as well, albeit without a clear distinction of its connection to the stalk. Although the resolution isn't sufficient to ascertain the exact structural differences between the classes, one can still observe the single similarity they possess- structural asymmetry. This asymmetry can be observed in the orientation of the motor domains, with one of the motor domains positioned closer to the stalk than the other (fig 3.34 c, d, and e). I proceeded to refine these classes of both forms of HSET, as described in fig 3.34.

While RELION is unable refine the classes beyond the starting resolution, cryoSPARC showed a significant improvement in the level of detail. The resolution quoted by cryoSPARC for each of the refined maps ranged between 13 and 15 Å. However, I decided to estimate the resolution by successively low pass filtering the maps and checking for a change in the level of detail in them. The filters had no effect on the models at values below 20 Å. After a finer adjustment, I was able to obtain a more realistic resolution estimate of 19-21 Å for the maps (fig 3.34 f, g and h), which is a typical resolution obtained from negative stain analyses (Ohi, Li et al. 2004). The three refined maps shown in fig 3.35 display a level of structural detail which enables the interpretation of similarities as well as differences between the two forms of HSET. For instance, the motor domains in either conformer of Apo-HSET can be seen to display differential modes of attachment to the stalk (fig 3.34 f and g). This is accompanied by a visible change in the neck connecting the stalk to the motor domains from an L-shaped kink in fig 3.35 a to a relatively straight structure in fig 3.35 b. For this reason, I decided to name the conformers type K (Kinked) and type S (Straight). Conversely, the region comprising the rest of the stalk and the N-terminal CBD in the type K state appears straight, while an evident curvature is seen in the type S state.

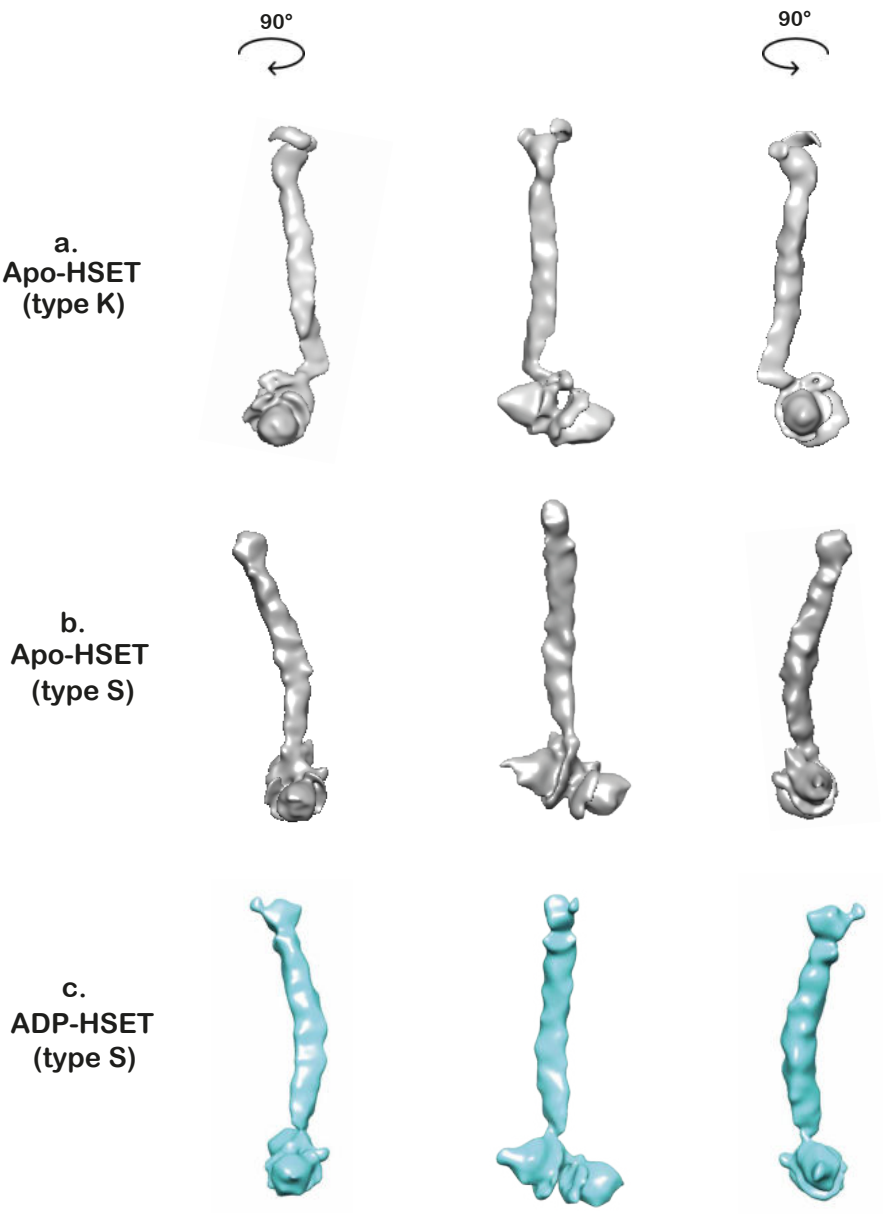


Figure 3.35 Final 3D maps of Negative stained Apo and ADP-FL HSET

Refined 3D models of (a, b) Apo and (c) ADP-HSET in three orthogonal directions highlight differences in the orientation of motor domains and the kink in the neck region that connects the stalk to the motor domain.

The apparent coordination observed between the structural features that define each arrangement of Apo-HSET is the first sign of the inter-domain crosstalk that plays a major role in coordinating the mechanical cycle of HSET. For this reason, I believe that the arrangements I have isolated are in fact, distinct conformational

states of Apo-HSET. Moreover, the unmistakeable similarity in the appearance of the motor, neck, and stalk of fig 3.35 b and c suggests that majority of the ADP-HSET in my sample have adopted a type S conformation. While the conformers listed here were the only ones observed in my data during the 3D classification, the occurrence of a few particles in the third class of Apo-HSET as well as the second class of ADP-HSET mentioned earlier suggest that there could indeed be more unique conformational states. The implications of these conformers in the mechanical cycle of HSET can be fully understood by further resolving the two forms of HSET using cryo EM.

3.5 Cryo Electron Microscopy- Sample Preparation

Cryo EM, unlike negative stain EM, is devoid of any staining and is performed at cryogenic temperatures (-170 to -196 °C) from sample preparation to data collection (Dubochet, Adrian et al. 1988). The visibility of protein molecules on the grid is dictated by the phases of the electrons rather than their amplitudes (Orlova and Saibil 2011). In other words, the contrast in the image depends on a more subtle effect of the time-delay between the scattered electrons reaching the detector rather than the pronounced effect of a difference in numbers observed in negative stain (see section 3.4, *Page 97*). Due to this inherently low contrast, cryo EM requires a higher concentration of particles and the collection of thousands of micrographs when compared to negative stain EM in order to visualize the structural details within the molecules, albeit with a higher potential resolution.

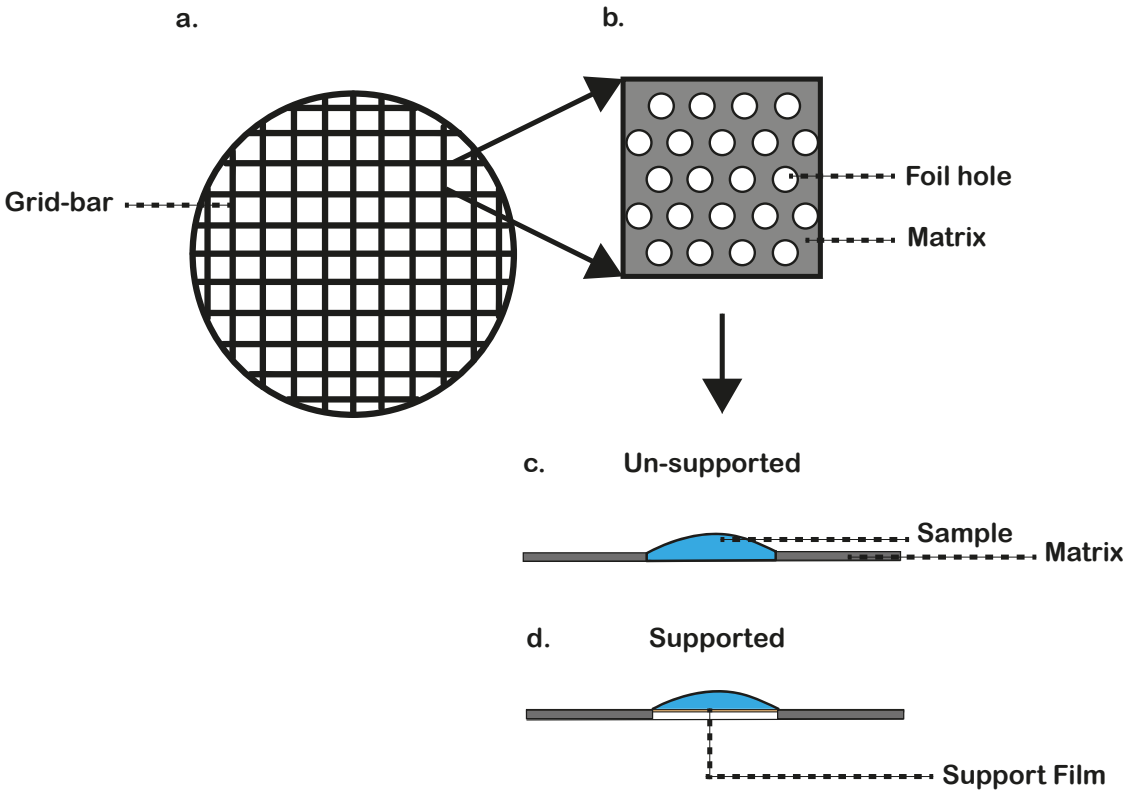


Figure 3.36 Holey Grids for Negative stain EM and Cryo EM

(a) A holey TEM grid is a flat disc that contains metallic grid-bars to support (b) an inert matrix/foil with holes in it of equal size and even spacing; cross-sectional view of an unblotted sample occupying the holes in an (c) un-supported and (d) supported grid.

The making of grids of a sample is often the rate-limiting step in solving its structure by cryo EM (Passmore and Russo 2016). A grid (fig 3.36 a) is a flat, circular disc with a diameter of around 3 mm, that is made of a mesh of metallic bars that support an inert matrix, otherwise known as a foil. The foil has holes that are either of the same size and evenly spaced as shown in fig 3.36 b or of varying size and non-uniform spacing. The grids I used were of the former variety, known as holey grids, while the latter are lacey grids. In either case, the holes serve as a path for the transmission of electrons and the sample is imaged over them (Sgro and Costa 2018), as illustrated in fig 3.36 c. In some cases, a thin film of material such as carbon, molecular graphene, or graphene oxide is manually layered on top of the grid to cover the holes with a support (fig 3.36 d). Support films are typically used to improve particle distribution in the holes by promoting the adsorption of the molecular sample, whereas un-supported grids lack any means of controlling the distribution of sample over the holes (Orlova and Saibil 2011).

A successful grid is the result of several physical and chemical factors that influence the stability of the sample as it is forced into a thin film (fig 3.37). The success of the physical parameters such as blot force, blot time, wait time, and type of grid being used is typically evaluated by the abundance of hydrated squares on a grid-map, the maximization of the hydrated holes in each square and an ice layer over the holes that provides reasonable contrast in the micrographs. On the other hand, the consistent presence of uniformly distributed particles at a high density with no heterogeneity or aggregation in every hole is dictated by the chemical parameters such as the concentration of salt and glycerol in the buffer (Thompson, Walker et al. 2016). These concern the content and quantity of the buffer(s) that are used in diluting the protein sample, which have been investigated in detail in the section on mass photometry (see section 3.2.5, *Page 70*). In this chapter, I will present the extensive tests that I performed on the various physical and chemical conditions mentioned above. I performed these tests on unsupported as well as supported grids (fig 3.37) but prioritised the former as with unsupported grids there is no signal from a support that could contribute to the background and deplete the contrast. This was especially

important when investigating a small and elongated target molecule such as HSET, for which is challenging to obtain sufficient contrast to begin with.

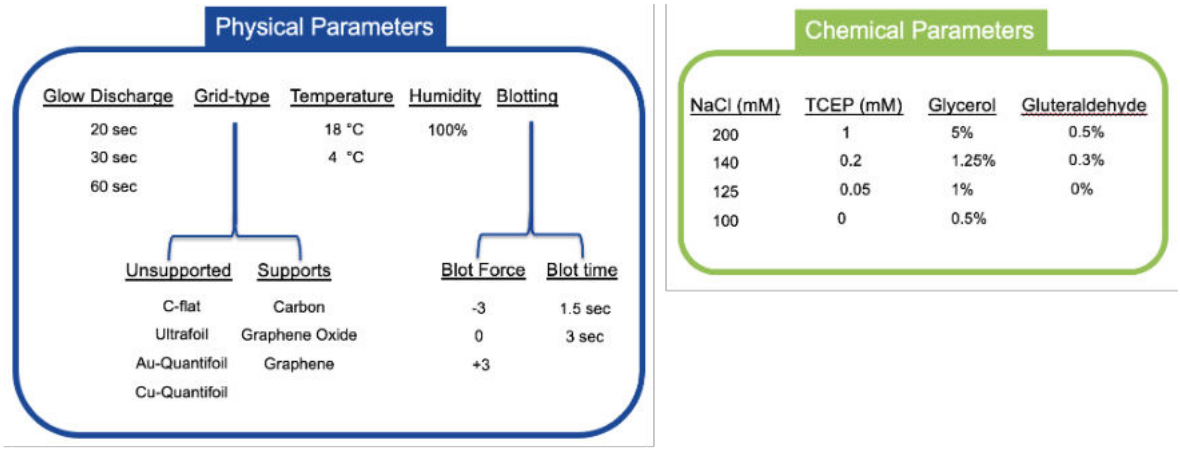


Figure 3.37 Optimisation of Grids for Cryo EM of HSET

Several physical and chemical parameters were tested to obtain the perfect grids for cryo EM of Apo- and ADP-HSET

3.5.1 Blot-force

The blot force is the force with which the liquid in the sample is blotted away from the surface of the grid in a device called the Vitrobot. The FEI Vitrobot IV comes with a scale for applied blot forces ranging from negative to positive values, which controls distance of separation between the blot-pads during the blotting step (Sader, Matadeen et al. 2020). A blot force of 0 is defined by the blot-pads just touching and letting a sliver of light through them, and each step corresponding to one unit of blot force is a 25 μm movement of the pads. An increase in positive values brings the pads closer together, while an increase in negative values moves the pads further ways from each other (Sader, Matadeen et al. 2020). One can achieve a similar effect by either varying the blot force applied for a specific

time, or by changing the time for which a constant force is applied. To minimize the number of variables, I decided to only vary the blot force and use a standard blotting time of 3 seconds for all grids unless explicitly mentioned. I tested blot force magnitudes of +3, 0 and -3 on the Vitrobot scale. The effects shown in fig 3.38 a, b, and c for a Copper Quantifoil grid highlight a reduction in the thickness of the ice with decreasing blot force from +3 to -3. This trend is also visible in fig 3.38 d and e, for other unsupported grid types that I tested as well. Judging by the gradient of ice thickness on the grid-map as well as the number of available hydrated squares, I decided to deploy a blot force of -3 for all subsequent unsupported grids.

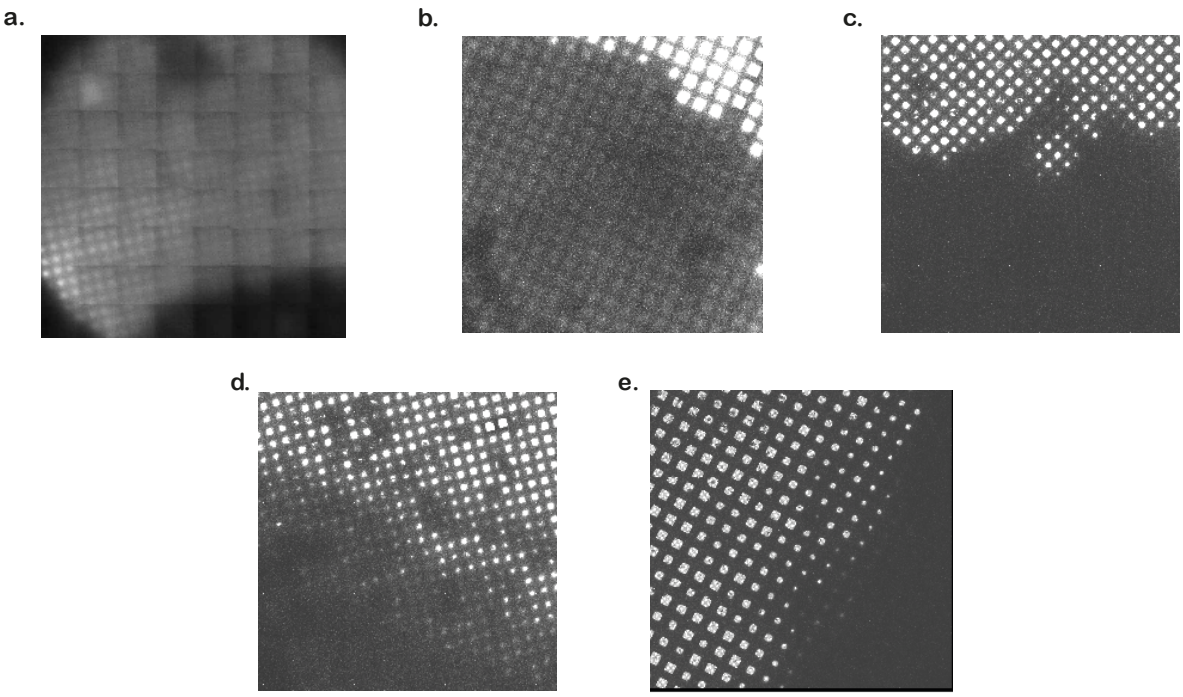


Figure 3.38 Effects of Different Blot Forces on Unsupported Grid-types

Copper Quantifoil R 1.2/1.3 grids of Apo-HSET prepared by applying a blot-force of (a) +3, (b) 0 and (c) -3 show the ice thickness over the grid decreasing with increasingly negative blot force; blot force of -3 was also tested on (d) C-flat 4-Cu-T-50 and (e) Ultra-Au-foil grids.

3.5.2 Grid-type

In fig 3.38 c, d, and e, one can observe a sharper gradient with significantly fewer usable squares on the Quantifoil grid compared to the C-flat and the Ultra-Au-foil grids, therefore the Quantifoil grids were not pursued further. Of the two other grid types, I prioritised C-flat grids because they were easily available and more affordable compared to Ultra -Au foil.

The C-flat grids contain a carbon matrix supported by copper grid-bars, like Quantifoil grids. The C-flat grids are different in the significantly thinner and flatter carbon matrix which is intended to achieve thinner ice over the grid (fig 3.38 d), leading to improved contrast.

3.5.3 Protein Concentration

Fig 3.39 a and b show a significant difference in particle density in the micrographs upon a twofold dilution of Apo-HSET, as well as a density-dependent reduction in contrast. Moreover, many of the particles that are visible on the undiluted and twofold diluted grids do not possess distinguishable features of HSET. This could be the non-specific higher order oligomeric clusters that occur as HSET has tendency to aggregate at this concentration range, as shown in DLS (see section 3.2.4.2, *Page 68*). The reduction in contrast is exaggerated by a significant reduction in particle density between dilution factors of two (fig 3.39 b) and three (fig 3.39 c).

In an attempt to control the distribution of particles with dilutions, I decided to introduce a small percentage of the cross-linking agent glutaraldehyde, which reacts with the amino groups of the N-terminus and the side chains of lysines and forms intermolecular cross-links in the sample (Hayat, 1986). The improvement in particle density observed in a threefold diluted HSET sample after the incorporation of glutaraldehyde (fig 3.39 d) is significant, in comparison to the glutaraldehyde-free version of the same sample (fig 3.39 c). However, the grid

showed aggregated protein in a significant proportion of the holes, which dominated the grids with further dilution of HSET, as shown in fig 3.39 e and f. I thus found myself at an impasse due to the chemical variable of particle distribution, whilst having optimised the physical variables of blot-force and grid-type. Therefore, I decided to explore Ultra-Au-foil grids, which have gold (Au) grid bars and a gold matrix. Considering the Ultra-Au-foil grids showed similar ice thickness as C-flat grids with a blot force of -3 (fig 3.38 e), I decided to test them as an alternative to C-flat grids.

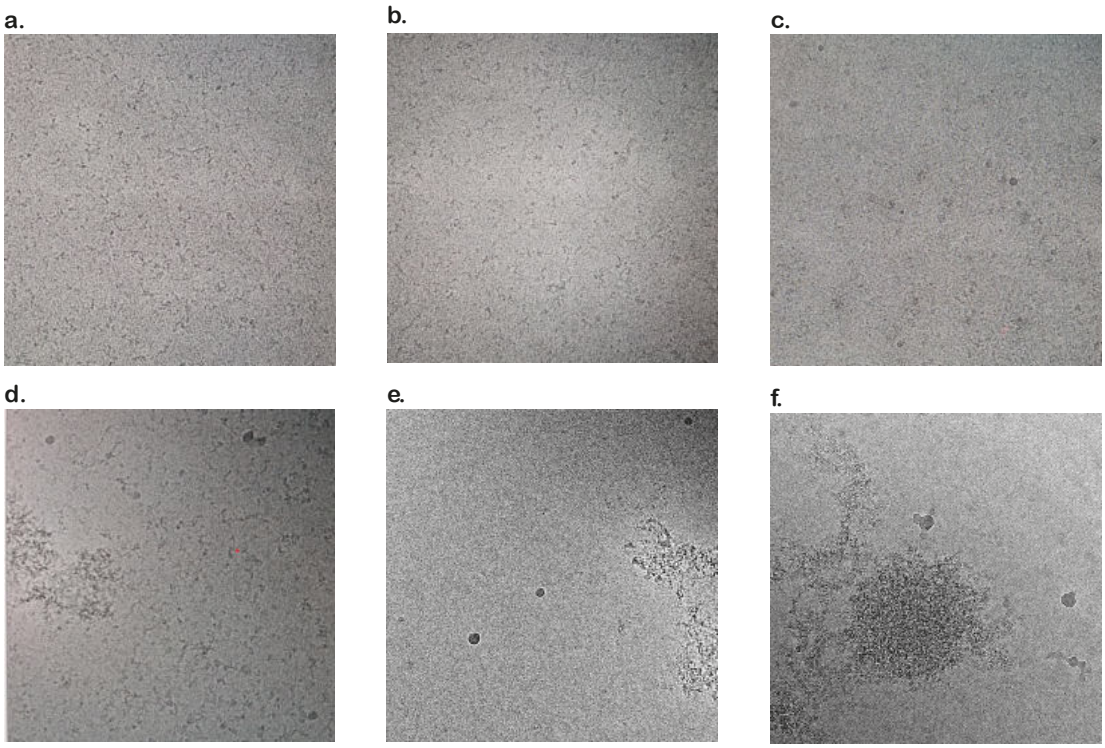


Figure 3.39 Effects of Protein Dilution on C-flat Grids

Micrographs obtained from C-flat grids containing Apo-HSET in the (a) undiluted- 0.8 mg/ml, (b) two-fold diluted- 0.4 mg/ml and (c) three-fold diluted- 0.26 mg/ml concentrations show a drastic difference in particle density for minute changes in concentration. Minor amounts of the crosslinker glutaraldehyde were incorporated to address this issue; effects of glutaraldehyde with dilution of HSET is demonstrated for the (d) three-fold and (e) six-fold and (f) ten-fold dilutions of HSET.

3.5.4 Glycerol Content

Fig 3.40 b shows a large increase in particle density on Ultra-Au-foil grids when compared to the C-flat grids with the same HSET concentration. I reasoned that this was likely due to the difference in the material of the matrix in the grids. C-flat and Quantifoil grids have a carbon matrix to which protein molecules tend to bind. In contrast, Ultra-Au-foil grids have a gold matrix to which protein molecules are less likely to bind. Therefore, more of the HSET molecules stay in the holes of Ultra-Au-foil grids, giving rise to a better particle density.

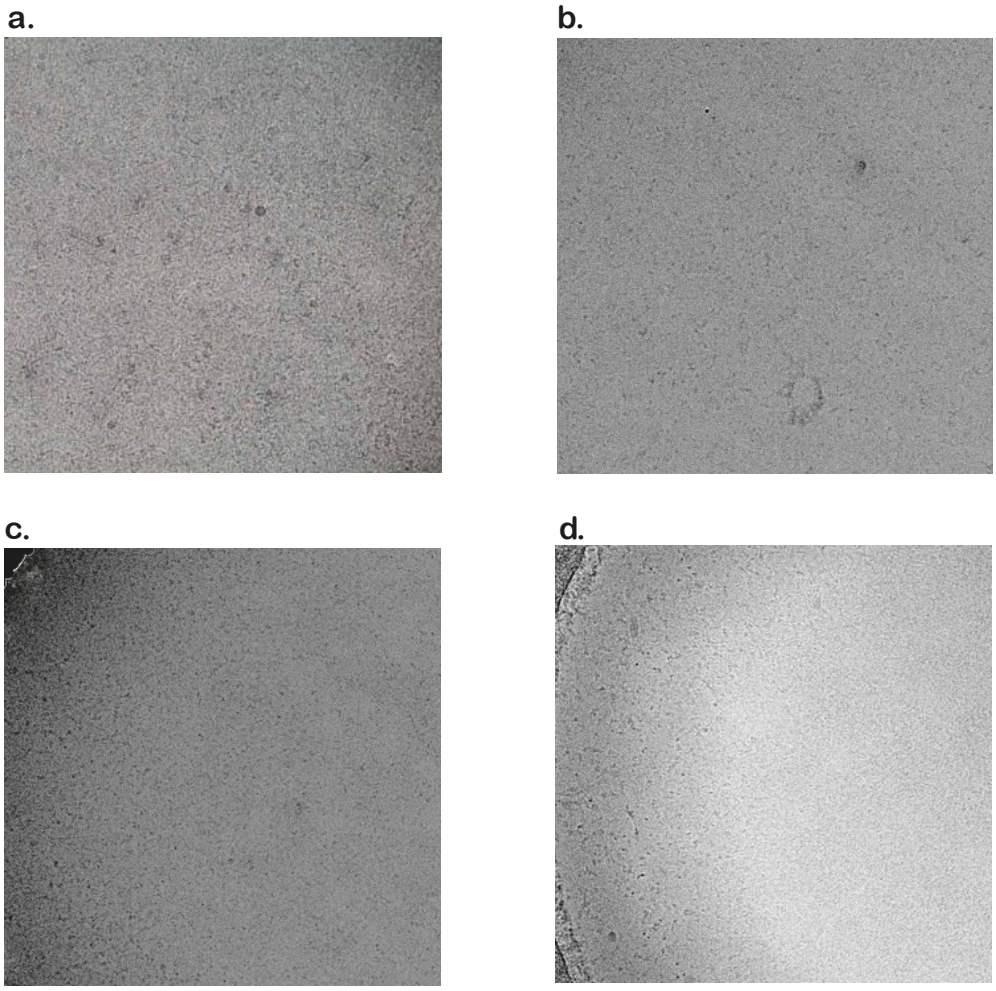


Figure 3.40 Exploration of Ultra-Au-Foil

Cryo-grids of Apo-HSET diluted to 0.26 mg/ml on (a) C-flat and (b) Ultra-Au-foil grids demonstrate a big difference in particle density and contrast for the same concentration; One of the challenges faced with Ultra-Au-foil grids was the thinning of ice towards the center of the holes and thickening near the edges (c,d) with further dilution, which clustered the particles towards the periphery.

The increased particle density in the case of Ultra-Au-foil grids presented the necessity to explore further dilutions of HSET to prevent particle overlap. To limit the loss of contrast, I decided to trial further dilutions while also reducing the glycerol content as much as possible within the limits identified by mass photometry (see section 3.2.5.3, *Page 73*). I was able to decrease the HSET concentration to 0.06 mg/ml. Below this concentration, I noticed a rise in aggregation, despite a favourable level of glycerol and salt in the buffer. Moreover, I observed a thinning of the ice near the centre of the holes, which clustered the particles closer to the edges as shown in fig 3.40 c and fig 3.40 d. Thus, striking a balance between the dilution factor and the contrast level to achieve a uniform distribution of particles over the holes in unsupported grids proved difficult. In contrast, I was able to achieve a better degree of control over particle density on carbon support film using a protein concentration that was six times lower during negative stain EM. Therefore, I decided to transition from unsupported to supported grids.

3.5.5 Support Films

An advantage of using supported grids is the ability to extrapolate the concentration of the sample from negative stain EM experiments. Although the concentration may have to be fine-tuned by a factor of two or three, the availability of a starting protein concentration and buffer conditions, has the potential to eliminate some of the optimisation necessary when using unsupported grids. However, this was not the case for the blotting parameters. As figure 3.41 a and b show, the result of a blot force of +3 applied for 3 seconds is completely different for an unsupported Cu Quantifoil grid compared to a supported one. The fact that one can see any squares that aren't shrouded in ice on the supported grid is in stark contrast to the behaviour of the ice on the unsupported grid for the same blot force. Furthermore, rather than a global

gradient of ice thickness, the hydrated squares are marked by individual gradients from the centre to the grid bars, akin to a fried egg as shown in fig 3.41 c.

The hydrated squares are also scattered in an apparently random fashion compared to the unsupported grid, with variations in ice thickness visible in the overall shape of the square as well as the darkness of the centre of the square (fig 3.41 d). This is a stark contrast to the hydration in the unsupported grids that affected the entire square (fig 3.41 c). While the pattern of hydration in the unsupported grids could be considered advantageous as it appears to maximise the usable holes in a square, the disadvantage is that entire squares can be discarded due to thin ice. The greater extent of blotting observed for the same blot force could be a result of the increased thickness of the grid due to the support layer. Consequently, the contact between the blotting pads and the supported grid would be greater than the case of the unsupported grid. This is further supported by the dry grid resulting from a greater blot force in fig 3.41 e.

A difference between the blotting of unsupported and supported grids was the incubation of the sample on the grid prior to blotting. While the incubation time on the grid was minimised for unsupported grids, it was necessary to promote adsorption of the particles onto the support of the supported grids. Fig 3.41 f illustrates the difference in particle distribution for a 10 second incubation time for the two types of grids. While the particle density is ideal for support, the incubation tends to cause aggregation of the protein on an unsupported grid as it increases the frequency of collisions between particles as well as that of the particles with the air-water interface (D'Imprima, Floris et al. 2019). While this effect could also occur in a supported environment, a minimal duration is required to promote adsorption of the particles onto the support film. To prevent any destabilisation to the sample by temporarily warming it up during the incubation, I decided to lower the temperature inside the Vitrobot to 4 °C, matching the temperature of the sample.

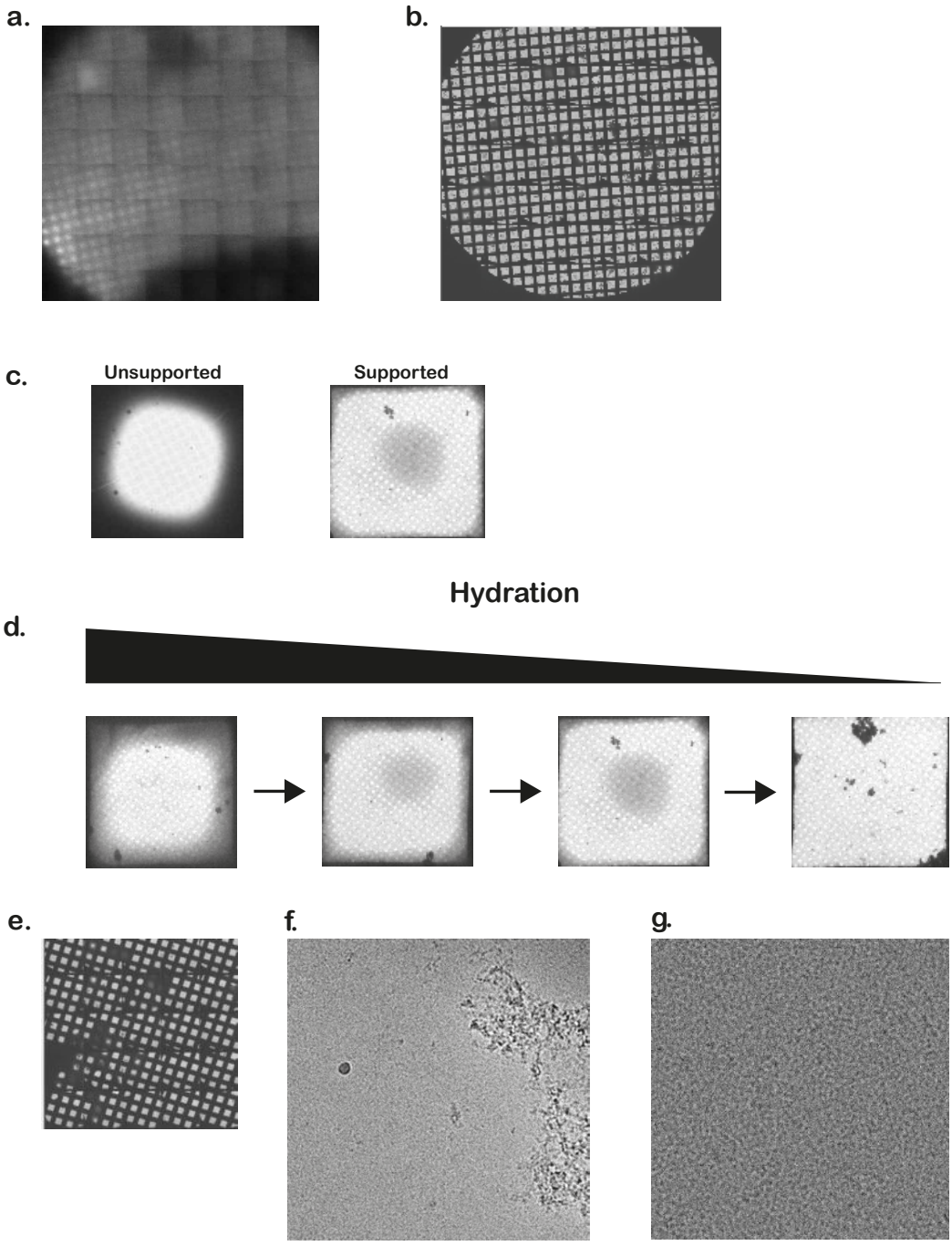


Figure 3.41 Working on Carbon Support

Carbon-supported Copper Quantifoil grids were preferred due to the ease of extrapolating chemical parameters from negative stain EM; A comparison between the ice film obtained across the Quantifoil grid that is (a) unsupported and (b) supported using the same blot force of +3 for 3 seconds; (c) the difference in the pattern of hydration per square between the two types of grids; (d) the manifestation of ice thickness across squares in a supported grid ranging from thickest (left) to thinnest (right); (e) extremely dry supported grid obtained using a blot force of 0; (f) The difference in particle distribution between an unsupported grid (left) and a supported grid (right) after a 10-second on-grid incubation at 4 °C prior to blotting.

Once the level of hydration and the ice thickness were optimised, I encountered possibly the greatest disadvantage of using a support film- a high background. This was compounded by the fact that HSET preferred thicker ice than desirable, i.e., within the periphery of the square shown in fig 3.41 c. To enhance the contrast, the grids were screened on the Glacios using a Volta Phase Plate (Danev, Buijsse et al. 2014). As shown in fig 3.42, the difference with and without a VPP was significant.

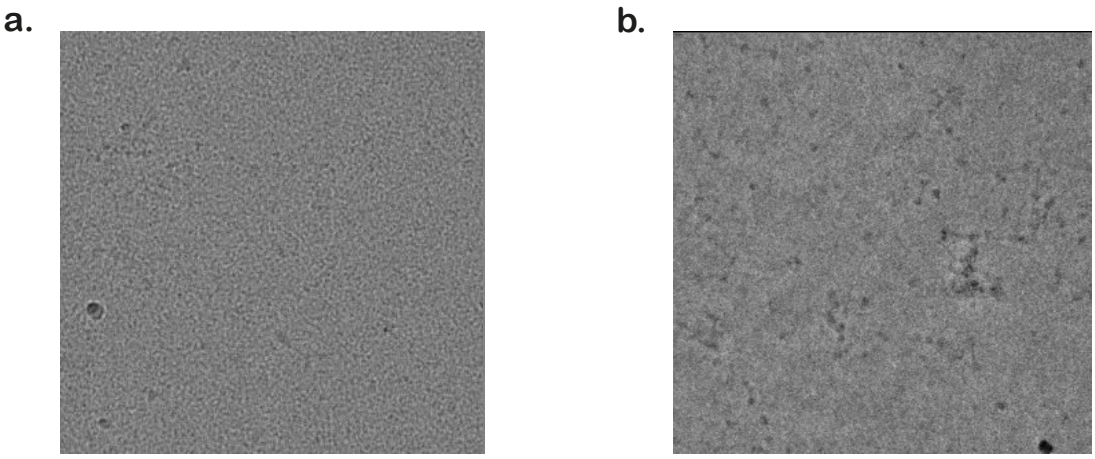


Figure 3.42 The Effect of Enhanced Phase Contrast

HSET particles on a successful Copper Quantifoil grid with carbon support imaged on the Glacios with a (a) 100 µm objective aperture at a defocus of -3 µm and (b) Volta Phase plate (VPP) at a defocus of -0.6 µm; the VPP boosts phase contrast to enhance the visibility of the particles against the background of the support film.

Despite the benefit of the enhanced phase contrast, the use of a VPP is generally discouraged by the EM community due to complications in set-up, acquisition and processing of images that are discussed in detail in chapter 4.4 (*Page 176*). To avoid relying on the phase plate, I tried using supports of molecular graphene and graphene oxide (GO) (Fan and Sun 2022) which tend to have a lower background. However, my attempts at using these supports were plagued by several issues. The foremost was the high degree of hydrophobicity of the grids which repelled most of the sample during sample application in the Vitrobot. This was observed for grids that underwent no pre-treatment prior to sample

application (based on manufacturer's instructions), as well as for those that were glow-discharged for 60 seconds or plasma-cleaned in a mixture of argon and air. Additionally, the batch of GO grids I used, contained several holes that were improperly bridged by the support film. After managing to successfully apply a small volume of sample onto molecular graphene and graphene oxide grids respectively, the few regions that showed proper bridging by the support film in the case of both grids, did not show any improvement in the visibility of HSET when compared to a carbon-supported grid. For this reason, the pursuit of alternative supports to carbon was abandoned.

3.6 Cryo Electron Microscopy- Data Collection and Processing

The size of a weak phase object plays an important role in its visibility in cryo-electron micrographs. Reliable identification of particles in the micrographs requires a significant amount of contrast at lower spatial frequencies. However, as shown in fig 3.25 (*Page 94*), acquiring images under focus boosts the intermediate frequencies rather than the lower range due to the sinusoidal nature of the CTF curve. With a size of 150 kDa and a mostly elongated shape, the total contrast generated by HSET is very weak, rendering HSET close to invisible in cryo-images on supported grids, as shown earlier in fig 3.42 a. Thus, to improve the visibility of HSET in cryo EM, I used a VPP to image HSET on carbon-supported grids. In the previous chapter, fig 3.42 illustrated the need for a phase plate to image my grids. Evidently, the boost in contrast generated by the VPP could play a crucial role in the success of the recognition, alignment, and classification of molecular images (von Loeffelholz, Papai et al. 2018). The improved signal to noise ratio of HSET particles contributed by the enhanced phase contrast would be beneficial in resolving the internal detail of the HSET molecule by cryo EM.

In setting up an acquisition using a VPP, the device is typically inserted in place of the objective aperture in the back-focal plane (see section 3.3.3.3, *Page 95*). The VPP is aligned to the optical axis to achieve on-plane illumination conditions (von Loeffelholz and Klaholz 2021). This condition is realised when the beam irradiates a single spot on the phase plate, leading to a localised build-up of potential over time which shifts the phases of the electron beam that passes through it (Danev and Baumeister 2016). This phase shift develops over the duration of the exposure of the spot and is moved to a neighbouring spot just short of saturating at a maximal value of phase shift. Each spot is referred to as a position of which there are 76 in total on a VPP (Danev, Buijsse et al. 2014). Hence, in addition to the acquisition parameters mentioned earlier, one must also define the activation time and the number of exposures per position on the VPP. These parameters are determined from an activation plot, obtained by exposing a position on the phase plate for an extended duration and monitoring the phase development in the resultant images.

3.6.1 The Glacios Sessions

The in-house Glacios microscope provided my first experience with VPP data collections and led to the first visualisation of FL HSET in high contrast in cryo EM images. The data from the Glacios were collected at a rate of 60-70 movies an hour, which meant that a weekend of Glacios time would yield 3000-4000 micrographs.

3.6.1.1 Live processing and pre-processing

The overall principle behind defocussed VPP-based data acquisition is described in section 3.3.3.3 (*Page 95*). In addition to the use of a nominal level of defocus, another factor plays a role in the VPP data- the phase shift resulting from the VPP is not always 90° i.e., π/2 rad. Instead, it develops with the exposure-dependent build-up of potential over time and thus attains a continuum of values between 0 and π rad as shown by the development plot or activation plot in fig 3.43. The fig 3.43 also shows that the level of contrast obtained is maximal at π/2 rad and deteriorates the further away from this value, on either side of it. The acceptable values of phase shifts were shown to range between 0.2π - 0.8π rad, according to the developers (Danev, Tegunov et al. 2017). The effects of the phase shifts applied to the data through the course of development of a single VPP position are described in the images adjoining the plot in fig 3.43.

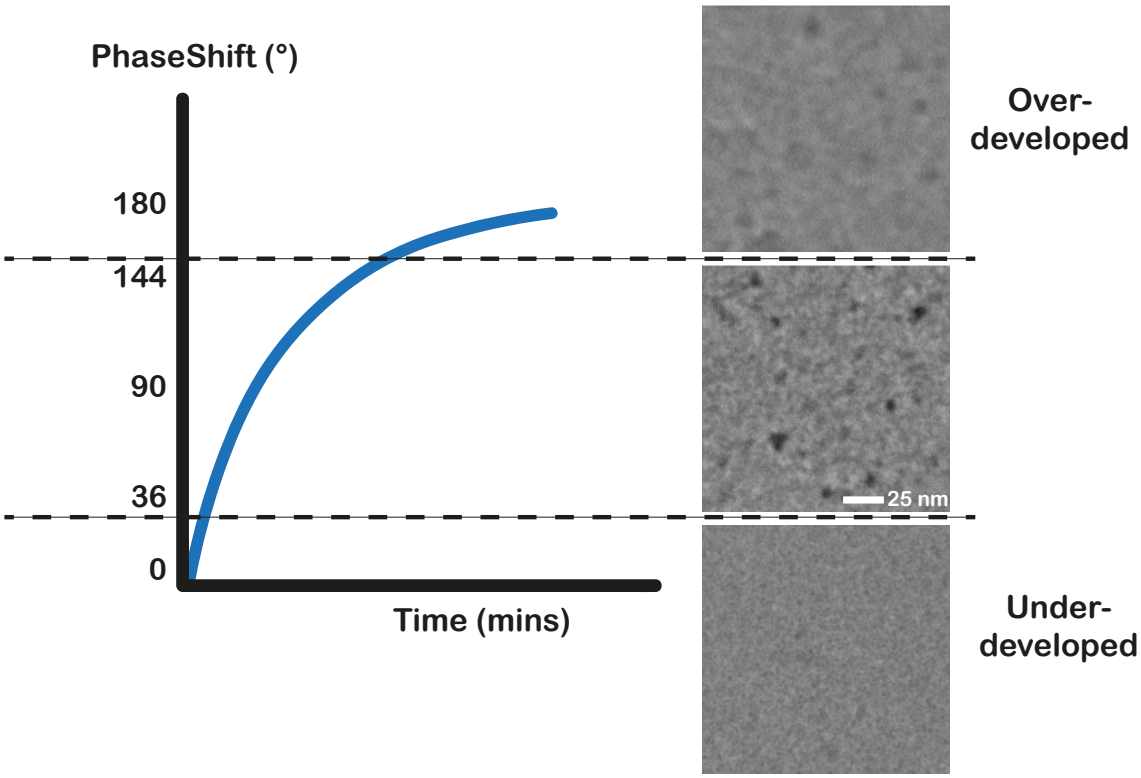


Figure 3.43 The Different Regimes of Phase Contrast

An example of a phase development profile which is obtained by the continuous exposure of the VPP over the carbon matrix of a Quantifoil grid; different stages of phase development are depicted by representative micrographs from an ADP-HSET dataset demonstrating the effects of low (under-developed), desirable, and high (over-developed) phase shifts on the quality of information in the image; data was collected on the Glacios.

The image corresponding to the under-developed phase plate in fig 3.43 was acquired at a phase shift of 0.25π rad, while the over-developed image corresponds to a phase shift of 0.8π rad. Considering the additional contrast needed for a particle as small as HSET, I chose to start at a phase shift of 0.3π rad. Based on the activation plot of the Glacios VPP shown in fig 3.44 a, a 40-second activation time would be required to reach a phase shift of 0.3π rad. This allowed a time frame of nearly 5 minutes of acquisition to reach a phase shift of 0.8π rad, which implied a rate of 50 exposures per position. However, considering the time spent on focussing the beam between exposures during which the VPP would still be developing, I decided to use 46 exposures for each position. On-the-fly pre-processing of the acquired movies in cryoSPARC live allowed the real-

time monitoring of the phase development through the exposures (fig 3.44 b). The live monitor showed that each position began with a phase shift of around 0.18-0.2π rad, which was lower than predicted based on the activation plot. More unexpectedly, a significant proportion of the phase shifts per position reached values of 0.75π rad or higher well within the 5-minute lifespan of each position (fig 3.44 b). Some positions would reach a maximal value of 0.9π rad after which the phase shifts would start over from 0.5π rad to trace a shorter curve before heading to the next position (fig 3.44 b). This was a clear sign of an unanticipated over-development of the VPP, further supported by a peak in the histogram of the phase shifts centred at 0.8π rad (fig 3.44 d).

To avoid this, I eliminated the activation time and pre-emptively halved the number of exposures per position before continuing with the acquisition. The former had the effect of reducing the starting phase shifts per position further still, to values of nearly zero rad (fig 3.44 c). The shorter exposures reduced the number of maximal phase shifts for the first couple of positions, which peaked at around 0.7-0.75π rad. However, these values gradually increased with time and tended to surpass 0.9π rad, showing more than one development curve per position (fig 3.44 c). The histogram of the data quickly took the shape of a bimodal distribution, with one peak centred around low phase shifts of 0-0.2π rad while the other was consistently at 0.8π rad (fig 3.44 e).

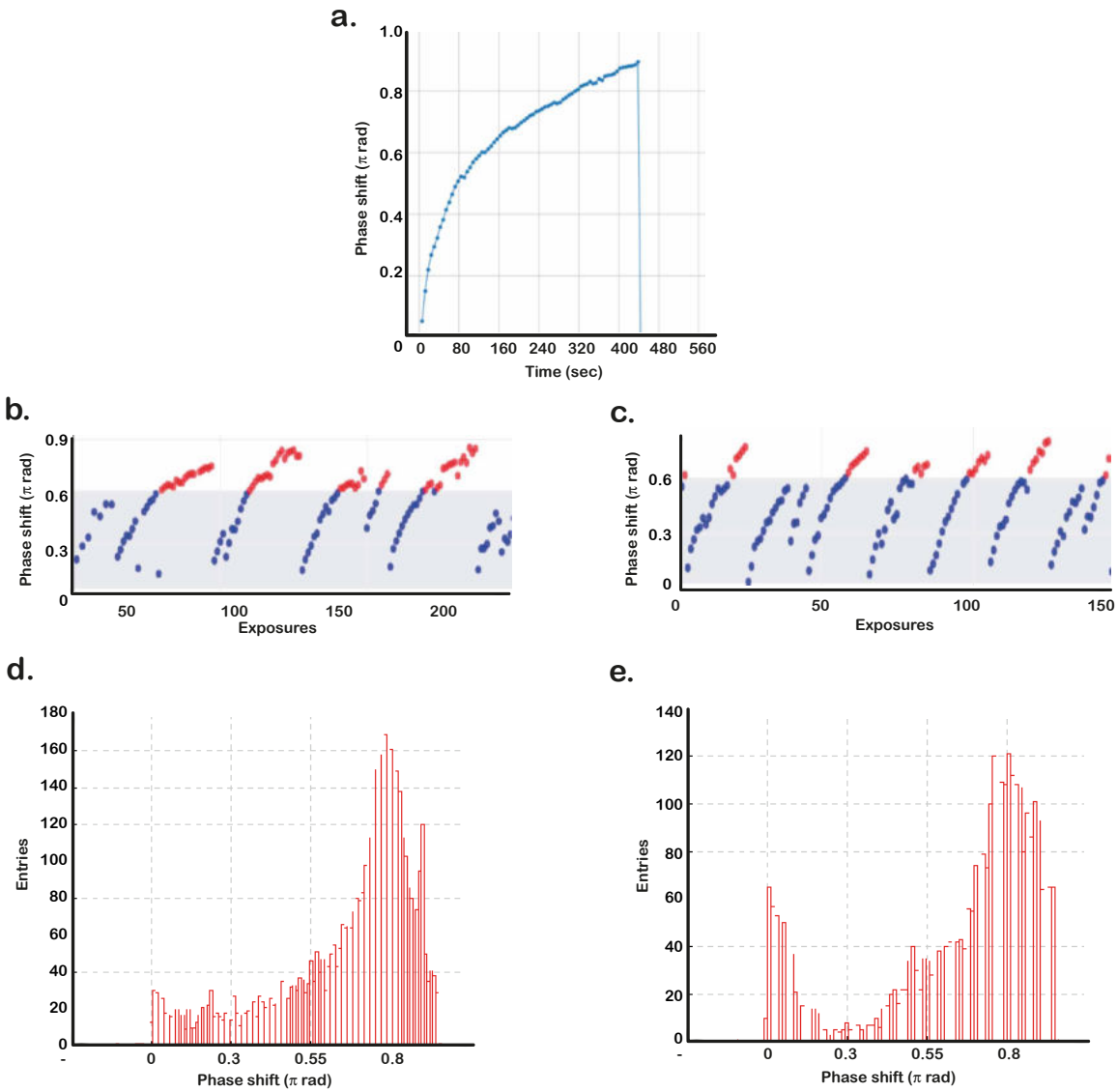


Figure 3.44 Distribution of Phase Shifts Through an Overnight Collection of ADP-HSET Data on the Glacios

(a) The activation plot of the VPP on the Glacios shows the development of phase shifts (Y-axis) over time (X-axis); On-the-fly pre-processing in cryoSPARC live shows the phase development (Y-axis) per acquired exposure (X-axis) and histogram from CTFFind4.1.14 shows the distribution of phase shifts across the entire dataset collected under the following conditions- (b,d) an activation time of 40 seconds and 46 exposures per position (c,e) no activation time and 24 exposures per position; data points in red in the live data have phase shift values above 0.62π rad; The data shown was acquired from an ADP-HSET grid.

To weed out these blurred images that contaminated the dataset, I identified markers in the metadata which could be used to customise filters to clean up the

dataset (table 3.2). On average, around 68% of the dataset was discarded after each session of data collected on ADP-HSET, 83% of which was due to the phase plate. For every 4,000 images collected, at least 2500 were discarded during the purging process. To compensate for this loss of data, I organised six sessions of collecting ADP-HSET datasets, which yielded a total of 15,000 movies.

Fault in CTF parameters	**% Of the entire dataset that was blurred**	
	ADP-HSET (6 sessions, 15,000 movies)	**Apo-HSET (1 session, 3,000 movies)**
Maximum estimated resolution > 10 Å	8	12
Defocus > 1 μm	18	20
Poorly defined thon rings	23	25
Phase Shift < 0.25π or > 0.7π rad	58	70
	Total- 68%	**Total- 80%**

Table 3.2 Narrowing Down the Causes of Blurred Images in Apo and ADP-HSET Datasets

(Left) the anomalies in metadata values that are attributed to the blurred appearance in many micrographs, and (right) the prevalence of the affected micrographs expressed as a percentage of the entire dataset in each case.

As a result of the focus on data collections of ADP-HSET grids, Apo-HSET grids were used for just one session that resulted in a dataset of 3,000 movies. However, as shown in fig 3.45 b, the over-development in the Apo-HSET dataset was significantly worse than ADP-HSET, with the histogram showing a peak centred around negative phase shifts in addition to a peak at 0.9π rad. Visually, most micrographs appeared blurred and shaky. Overall, the Apo-dataset suffered a loss of 80% of its images (table 3.2).

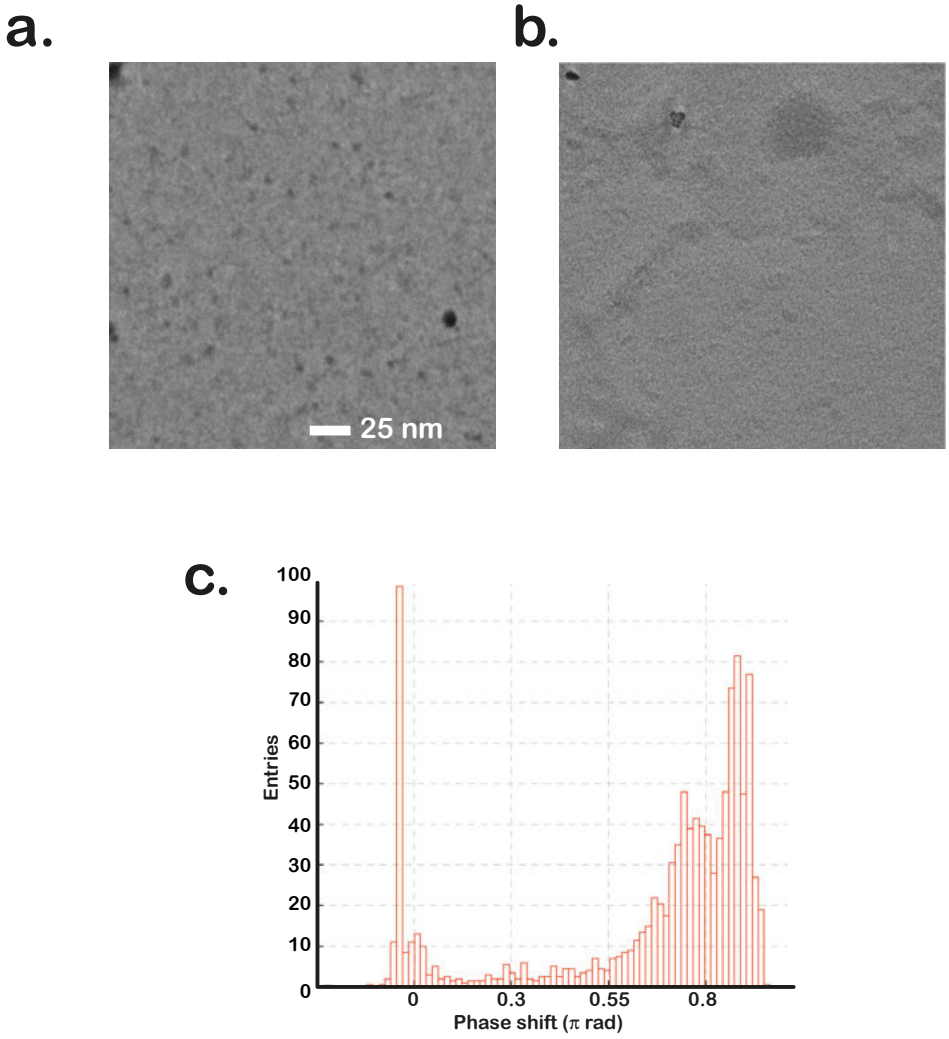


Figure 3.45 Processing an Apo-HSET Dataset

The dataset of Apo-HSET was plagued with abnormally excessive phase development; (a) a rarely occurrent unaffected micrograph shown beside a (b) typical blurred and drifty micrograph that dominated the dataset, (b) the histogram of phase distribution shows most of the micrographs in the undesirable regime of phase shifts as well as negative phase shifts.

3.6.1.2 Particle picking

Due to the inverted contrast and the significant difference in signal to noise ratio compared to negative stain EM, I had to generate a new picking model suited to my cryo EM dataset. My approach led me to compare the effectiveness of CRYOLO's low-pass filter (fig 3.46 b) and a neural network-based denoiser called JANNI (Just Another Noise2Noise Implementation) (Wagner, Merino et al. 2019) in enhancing the signal in the micrographs for automated picking. The JANNI model I employed for the Glacios dataset (fig 3.46 d) had been pre-trained on multiple cryo EM datasets by the developers, giving it the ability to enhance the signal of any target molecules against the background.

In contrast to negative stain EM which succeeded with a low-pass filter, a similar treatment of cryo EM images seemed to make the picking model go astray (fig 3.46 c) with a higher number of false positives and negatives than the JANNI model, which was successful (fig 3.46 e). However, like the negative stain data, I could use the same picking model for both forms of HSET. A total of 300,000 particles were obtained across all sessions of ADP-HSET and 50,000 for Apo-HSET. The particle coordinates were imported into RELIONv3.1 to extract particles as described in section 2.4.3.2 (*Page 45*).

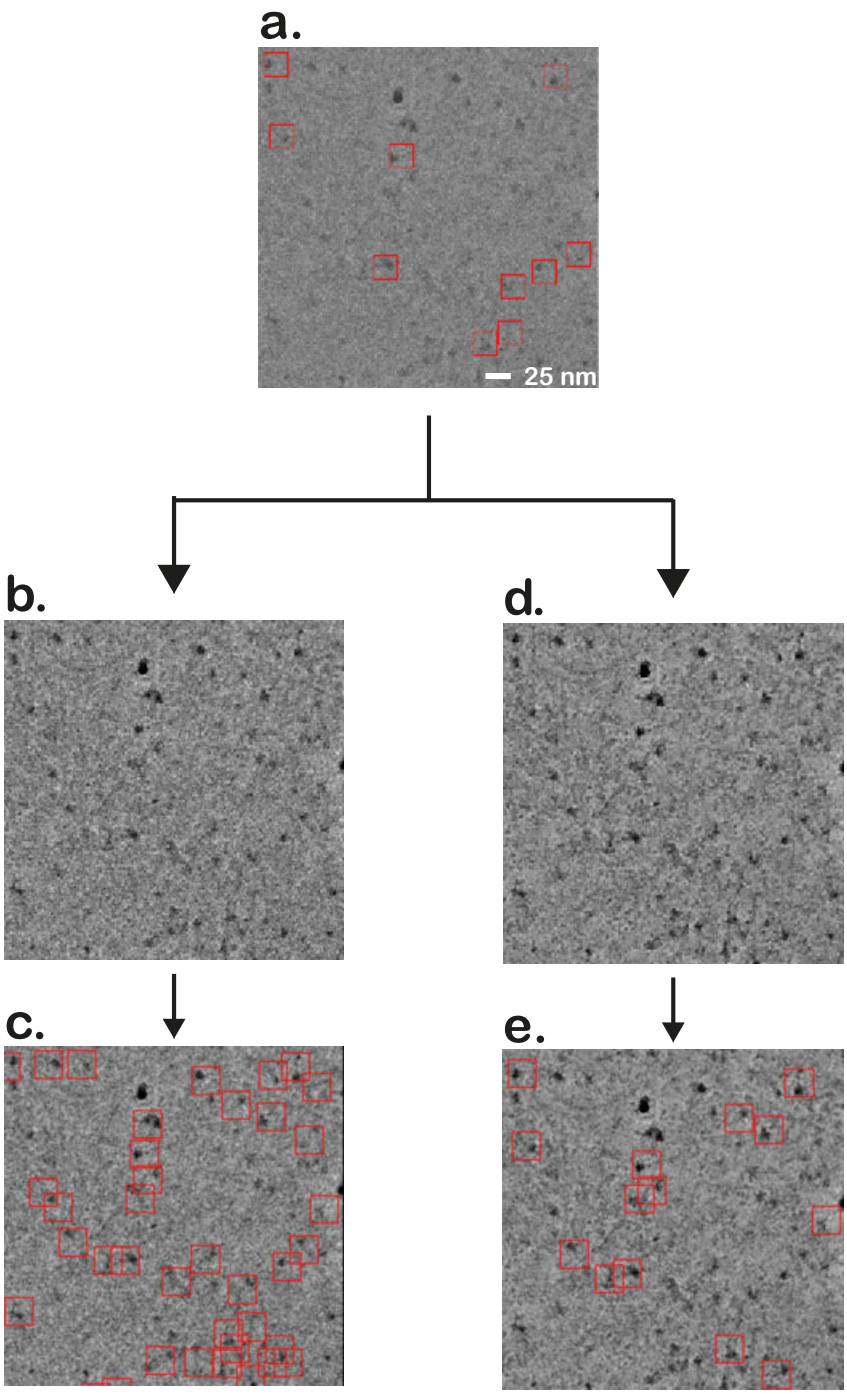


Figure 3.46 Particle Picking Pipeline in CRYOLO

The steps undertaken to obtain an optimal strategy for an optimal picking model are demonstrated on a sample micrograph from the training set. (a) An unfiltered micrograph with a few manually picked particles is denoised using the (b) default low-pass filter or (d) default JANNI denoising model; the particles are used to train pre-existing picking models, the trained models are allowed to auto-pick particles from the (c) low-pass filtered or (e) JANNI-denoised micrographs respectively.

3.6.1.3 2D Classification

For the 2D classification, I used the same alignment parameters that I had optimised for structural analysis by negative stain EM (see fig 3.31). While I classified the 50,000 particles of Apo-HSET in one run, I worked on the data from each session of ADP-HSET independently and processed each of them as soon as I acquired them. Thus, despite amassing 300,000 particles of ADP-HSET in total, the initial 2D classifications of each session only used 30,000-50,000 particles at a time. As displayed in fig 3.47, the classes obtained for both forms of HSET began to show the familiar elongated appearance with a globular shape at one terminus. However, the classes of Apo-HSET showed an unexpectedly high degree of blurring, despite the thorough purging of the dataset described earlier (fig 3.47 a). The blurring was persistent despite attempts at selecting and reclassifying the data into a greater number of classes. This led me to conclude that there was a problem with the data used for the classification rather than the classification itself. A further inspection of the metadata of the micrographs that yielded the particles revealed that the over-development issue was more severe than I initially thought. As shown previously for the ADP-HSET dataset (fig 3.44 b and c), the issue is typically recognised during the live monitoring of phase shifts by the presence of second smaller development curve per position of the VPP. This occurs due to the phase shifts crossing a value of π rad, which results in a similar CTF pattern to 0 rad due to the sinusoidal nature of the CTF which repeats itself every π rad. This could mean that many images that were recorded with a value of 0.3π for instance, actually had an over-developed phase angle of π + 0.3π rad. Simply put, the damage to the dataset is too deep-rooted for elimination by filtering based on the calculated metadata values. I therefore had to conclude that the data is no longer usable for analysis, and now turned my focus to processing the ADP-HSET dataset.

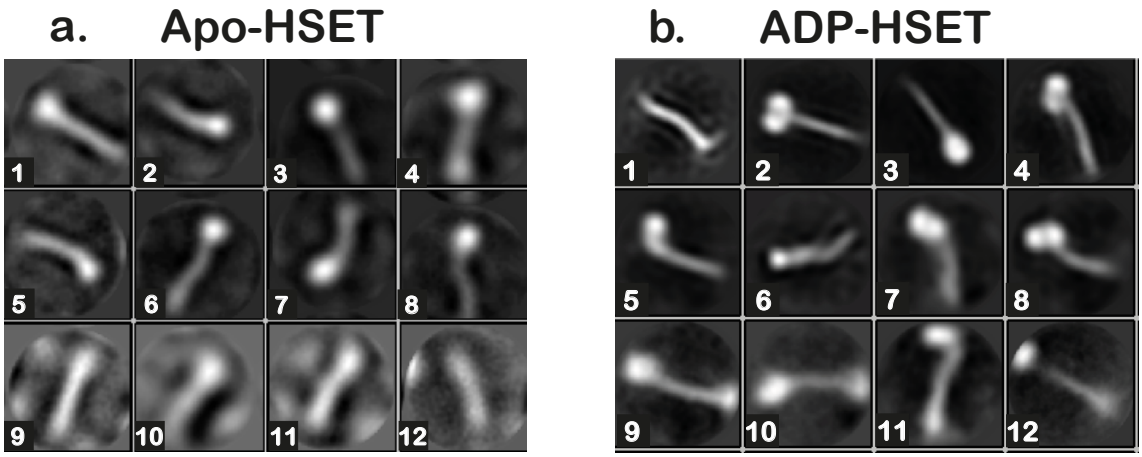


Figure 3.47 Reference-free 2D classification of Apo and ADP-HSET

Reference-free 2D class averages of (a) Apo-HSET appears blurred beyond repair, despite the removal of bad micrographs from the dataset and (b) ADP-HSET shows the shape that is reminiscent of the results of negative stain EM.

Unlike the Apo-HSET classes, the classes obtained from ADP-HSET (fig 3.47 b) resemble the characteristic shape that was observed in the negative stain dataset (see section 3.4.2.2, *Page 104*), thereby providing the first glimpse of FL HSET by cryo EM. One can notice both motor domains in some classes such as class 2,4 and 8. They also differ from each other in the tilt angles about the axis of the stalk (fig 3.47 b). The classes also sample views with greater tilt angles such as class 5 which appears almost perpendicular to class 2. Other classes suffer from fuzziness or a lack of discernible features, indicative of heterogeneity within the class that could arise from the misalignments of different views or the incorporation of false positives. The class-averages that result from such heterogeneity display recognisable kinesin features with a level of fuzziness superposed on them and are considered to be over-averaged.

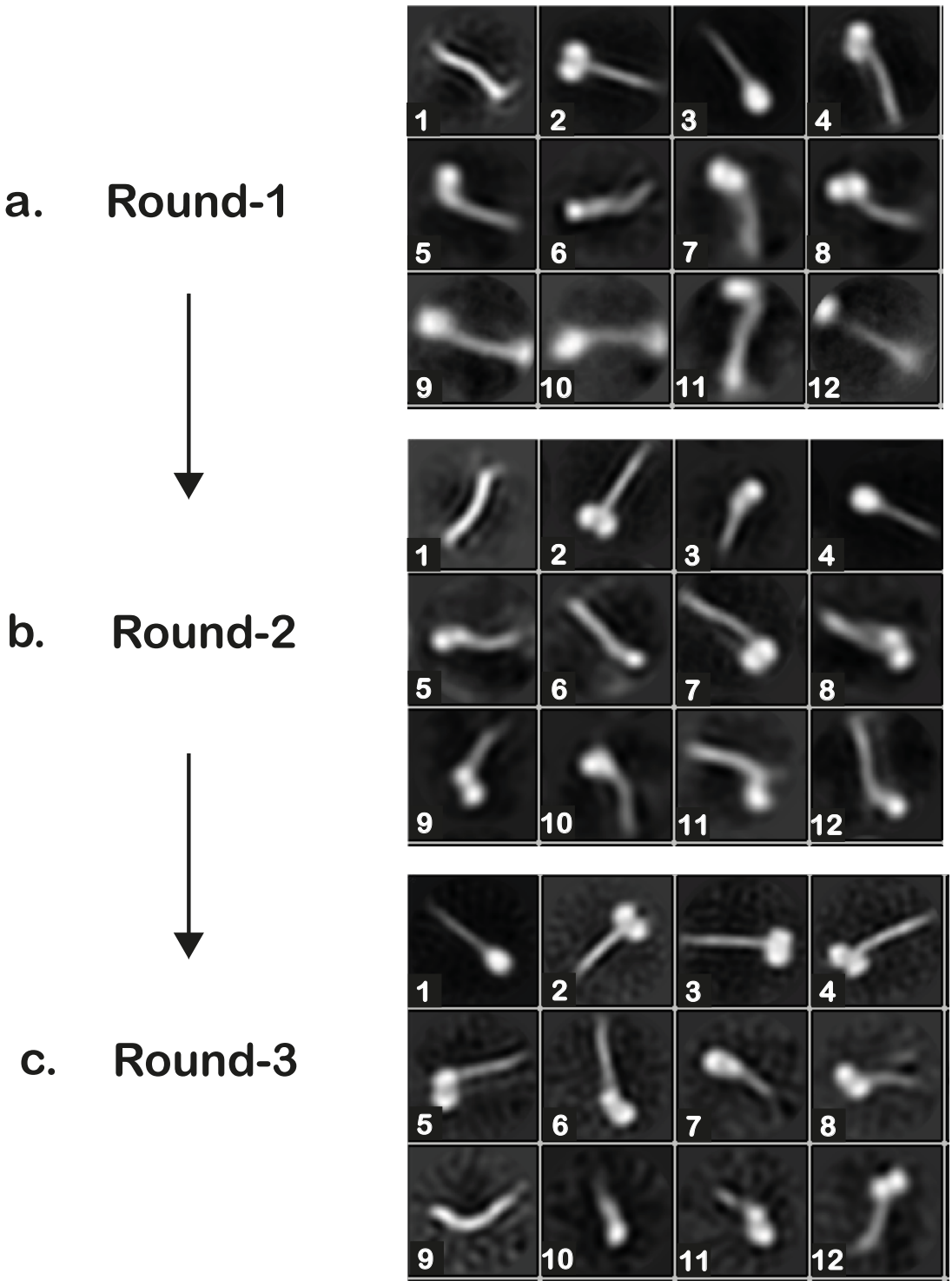


Figure 3.48 Purification by Classification

The extracted particles of ADP-HSET were subjected to three rounds of reference-free 2D classification to enhance the level of visible detail by achieving homogeneity; the best classes from each round are displayed as follows- (a) round 1, (b) round 2 and (c) round 3. The classes showing off-target features or poor signal-to-noise ratio at the end of each round were discarded and the remaining particles were reclassified into the same number of classes.

After using the first round of 2D classification to eliminate the obviously off-target classes, I decided to append the data from all six sessions into a single stack of particles, to tackle the over-averaging issue. My strategy involved subjecting the stack of particles to multiple rounds of selection followed by reclassification into a greater number of classes, allowing more room for the effective segregation of the particles (fig 3.48). I managed to remove the false positive images to a significant degree, as demonstrated by the progressive unblurring of the classes (fig 3.48 b and c). After maximising the number of classes with a high level of clarity (fig 3.48 c) and observing no further improvement with further classifications, I noticed another issue in the appearance of the classes.

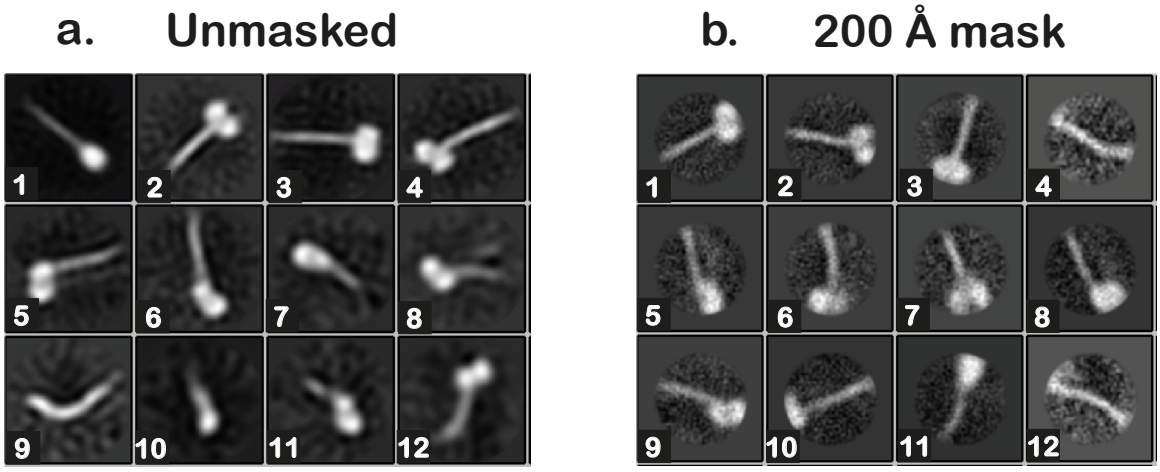


Figure 3.49 Testing the Effects of a Circular Mask on the 2D Classes of ADP-HSET

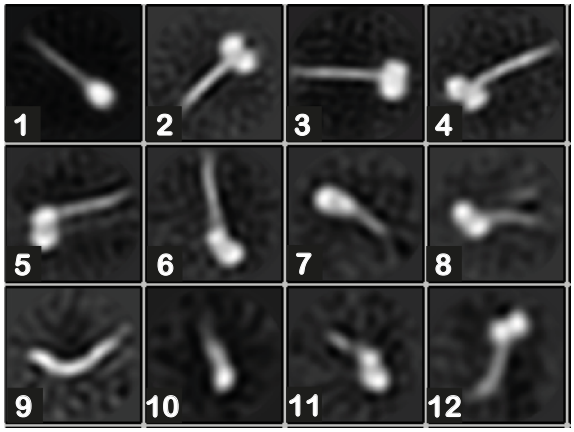
The application of a circular mask on the stack of particles during their 2D alignment and classification was undertaken to improve the quality of the 2D classes and view high resolution features; class averages of the (a) unmasked dataset are displayed alongside the classes obtained by (a) applying a circular mask of diameter 200 Å.

Unlike negative staining, classes obtained from cryo EM will eventually resolve higher resolution features due to the correction of the CTF of the movies. While the resolution of the data in the movies reached values as high as 5 Å, the classes I obtained in fig 3.48 c were yet to resolve features such as the coiled coil helices which would appear below 10 Å. Further rounds of selection and reclassification simply served as a never-ending sink for particles under the guise of an

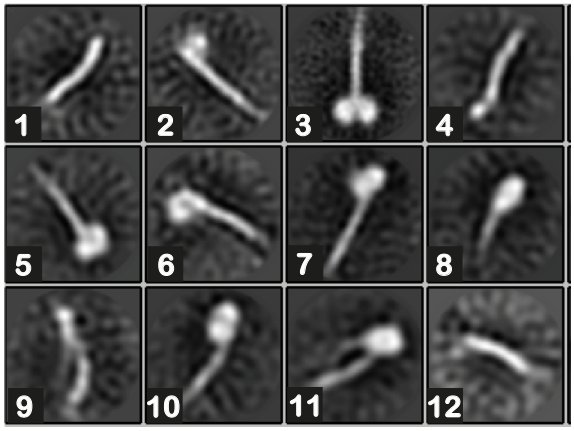
improvement in quality. Thus, my next goal was to resolve higher resolution features of HSET through approaches such as masking (fig 3.49) and filtering (fig 3.50) the particles.

The only effect of masking was to make the classes appear grainier thereby giving the illusion of improved resolution (fig 3.49 b). The high-pass filtering on the other hand, did lead to an increase in informative classes proportional to the resolution cut-off of the applied filter as shown in fig 3.50 b and c. Based on the progressively increasing difficulty in recognising an HSET particle in a high pass filtered image, I decided to avoid filtering the dataset beyond 150 Å. However, this was still not sufficient to resolve any secondary structure features within HSET. Drawing inspiration from the negative stain datasets that showed more detail in 3D than 2D, I proceeded with the generation of an ab-initio 3D model (fig 3.51).

a. No high pass



b. 200 Å high pass



c. 150 Å high pass

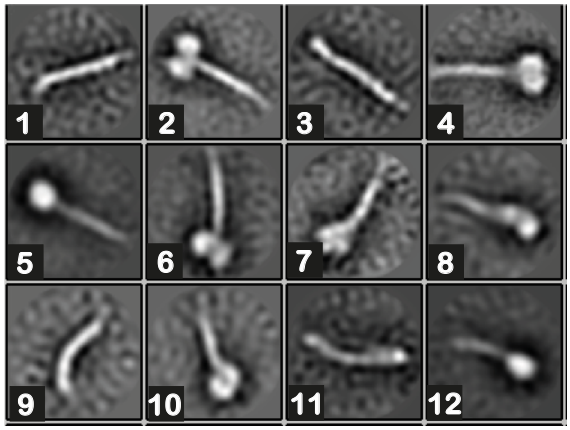


Figure 3.50 Testing the Effects of High-pass Filters on 2D Classes of ADP-HSET

To view higher resolution features in the classes, the particles were filtered to different cut-off spatial frequencies and reclassified into the same number of classes; class averages of the (a) unfiltered stack of particles are displayed alongside the averages obtained by filtering the stack to (b) a high frequency cut-off of 200 Å and (c) a high frequency cut-off of 150 Å.

3.6.1.4 3D Analysis

To improve the results that I obtained from 2D, I decided to explore the 3D analysis of the ADP-bound dataset. I selected 20,000 particles accounting for the best classes of ADP-HSET to generate an ab-initio 3D reconstruction (fig 3.51 a). While RELION quoted a resolution of 15 Å for the reconstruction, it clearly showed less detail than the negative stain models that had a resolution of 20 Å in section 3.4.2.3 (*Page 109*). The apparent resolution of the ab-initio map looked closer to 30 Å. Regardless, my next step was to attempt a 3D classification of the particles that remained after the 2D classification of each session. A total of 75,000 remained and were combined into a single stack of particles for the 3D classification, which failed to segregate them into more than one class and showed no additional detail. Hence, following the strategy I used for the negative stain datasets again, I used a reference-based 2D classification of the particles from this single class to evaluate their 3D alignments as shown in fig 3.51 b.

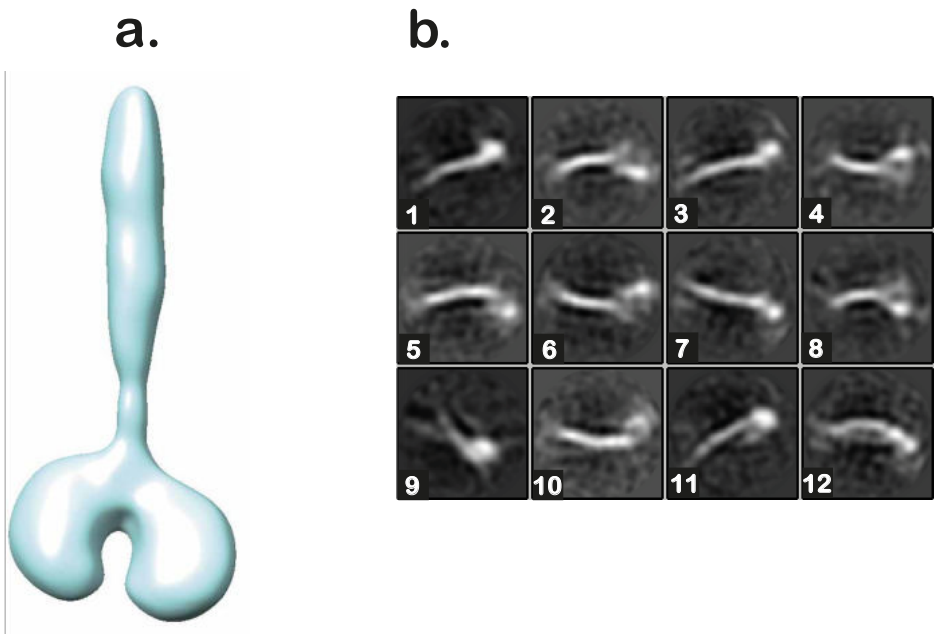


Figure 3.51 Reference-based 2D Classification of ADP-HSET

(a) An ab-initio 3D model of ADP-HSET generated from 20,000 particles belonging to the best 2D class averages was obtained at a quoted resolution of 15 Å; this model was used for reference-based classification of the dataset. (b) The reference-based 2D classes from 75,000 particles appeared scrambled and were less informative than the reference-free classes.

While some classes such as class 7 and 11 in fig 3.51 show the curvature in the coiled coil and appear to resemble side-views quite realistically, the rest of the classes have a scrambled and unreliable appearance. For instance, the presence of more than one stalk branching out from the motor domains in class 9, as well as the unrecognisable motor domains in classes 2 and 6, are evidence of improper alignment and by extension, a compromised classification. This was likely due to the feature-less ab-initio reference that was used, as well as the dominance of low-resolution features in the dataset as seen in the reference-free classes earlier. To tackle this problem, I decided to perform multiple 3D classifications, starting with the particles from a single session and refining the resultant 3D class using cryoSPARC's Homogeneous refinement tool (fig 3.52). I then appended the particles from another single session to this stack, performed another 3D classification using the successful 3D class from the previous run as a reference, and refined the resultant class. The aim of this strategy was to reduce the chance of faulty alignments by limiting the particles used for the 3D classification in each round, and use the resultant 3D class as a better resolved reference to reinforce the alignments of further classifications.

One can immediately notice some level of detail in the class from the first dataset (fig 3.52 b) that distinguishes it from the ab-initio reference (fig 3.52 a). While adding particles by combining the datasets doesn't change the resolution of the classes much, they do appear subtly distinct from each other in terms of curvature and angular orientations of the domains (fig 3.52 b, d and f). The unchanging resolution of the classes was reminiscent of the negative stain dataset, which only revealed more detail during the refinement in cryoSPARC. In a similar manner, there was an instant shift in resolution to 12 Å or below, with an improvement that was proportional to the number of particles included. However, the detail in the refined maps did not reflect their estimated resolution. The maps were still plagued by significant levels of low-resolution noise and missing density.

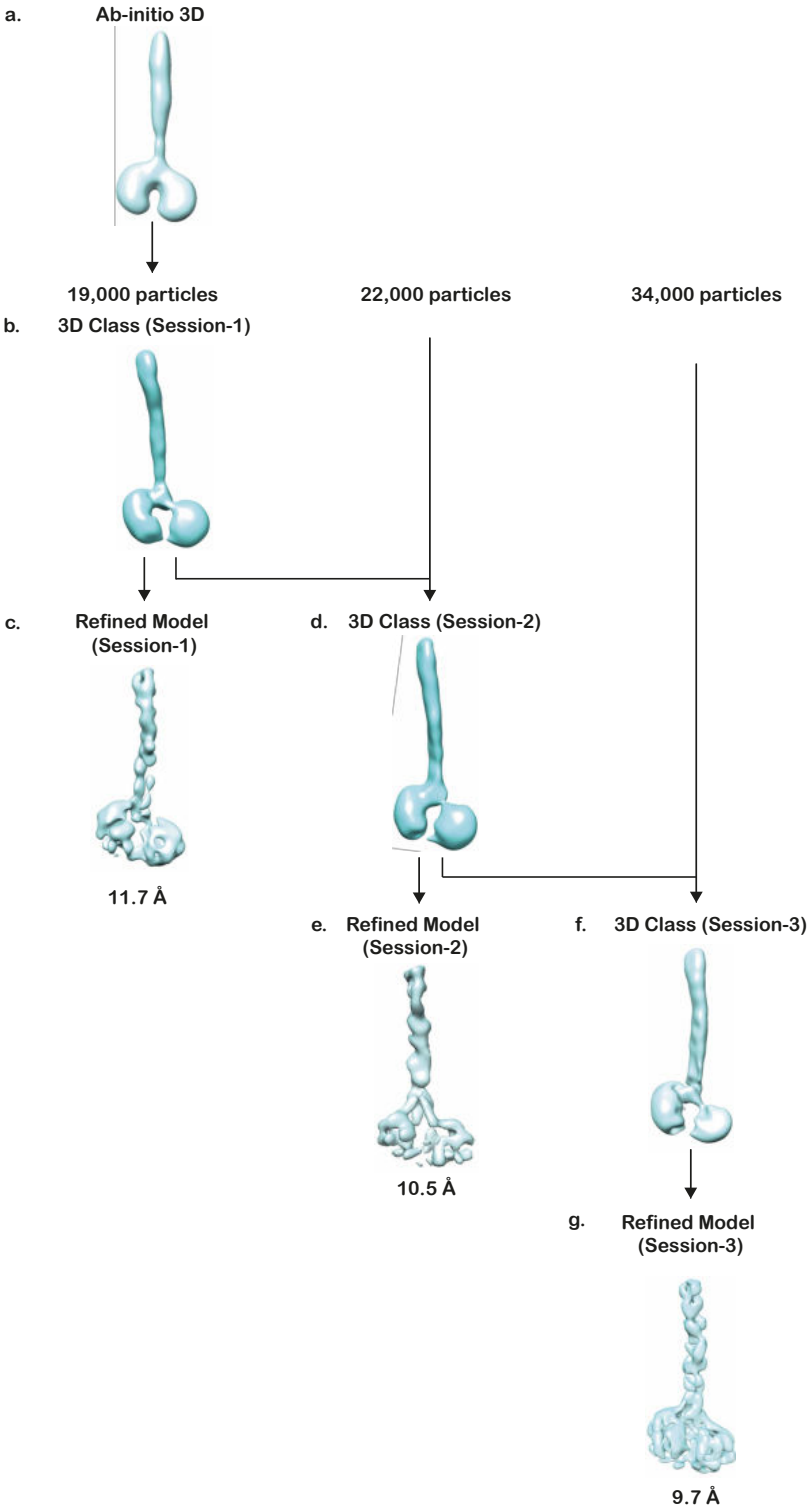


Figure 3.52 Refining the 3D Structure of ADP-HSET

The (a) ab-initio 3D model was used as a reference in the (b) 3D classification of a dataset containing 19,000 particles; the class obtained was used as a reference to classify a dataset of (d) 22,000 particles, which was subsequently used to classify (f) a 34,000-particle stack. The classes were respectively refined to resolutions of (c) 11.7 Å, (e) 10.5 Å and (g) 9.7 Å with suitable custom soft-edge masks using the Homogeneous Refinement Tool in cryoSPARC.

More importantly, the defining characteristics appeared to change entirely from one map to the next. For instance, both motor domains originate directly from the stalk in fig 3.52 c while fig 3.52 e shows branches extending from the stalk into the motor domains which appear to have gained more strands that are presumably helices, with a 1 Å leap in resolution. The stalk has changed from showing the twists of the coiled coil with a minor bit of discontinuity in the former map to a continuous mass of poorly defined density with a hook near the cargo-binding end and a marked partition near the neck linker. While the final map in fig 3.52 g shows a level of completeness that seems to correct the inconsistencies of the first two maps, it has still not managed to resolve the two strands of the coiled coil from each other despite being below 10 Å in resolution. Moreover, the features of the motor domains appear different from each other in the map and differ significantly from its predecessor in fig 3.52 e. Overall, the refinement in cryoSPARC has indeed provided a hint about the appearance of ADP-FL HSET in 3D and revealed a certain level of detail. Uncovering further information would require the collection of higher quality and quantity of data.

3.6.2 The Titan Krios Session

As the data processing from the Glacios stalled at mid to low-resolution ranges in 2D and 3D, I decided to obtain professional guidance in a facility where the performance of the VPP had been tried and tested- LONCEM. Unlike the Glacios which is made of two condenser lenses- C1 and C2, the Krios is equipped with an additional C3 lens, which enabled the illumination of a smaller area whilst maintaining a parallel beam. Using the Krios, I was able to collect three exposures in every hole as opposed to the single exposure per hole when using the Glacios, thereby increasing the acquisition rate to 90-100 exposures per hour. The data collection sessions typically lasted 6 days which yielded a total of 12,000 movies. The VPP activation plot from the Krios (fig 3.53 a) shows a phase development that is slightly slower than observed with the Glacios. The Krios VPP takes 72 seconds to reach 0.3π rad compared to the 40 seconds in the Glacios, and 6.5 minutes to reach 0.8π rad compared to the 5 minutes in the Glacios.

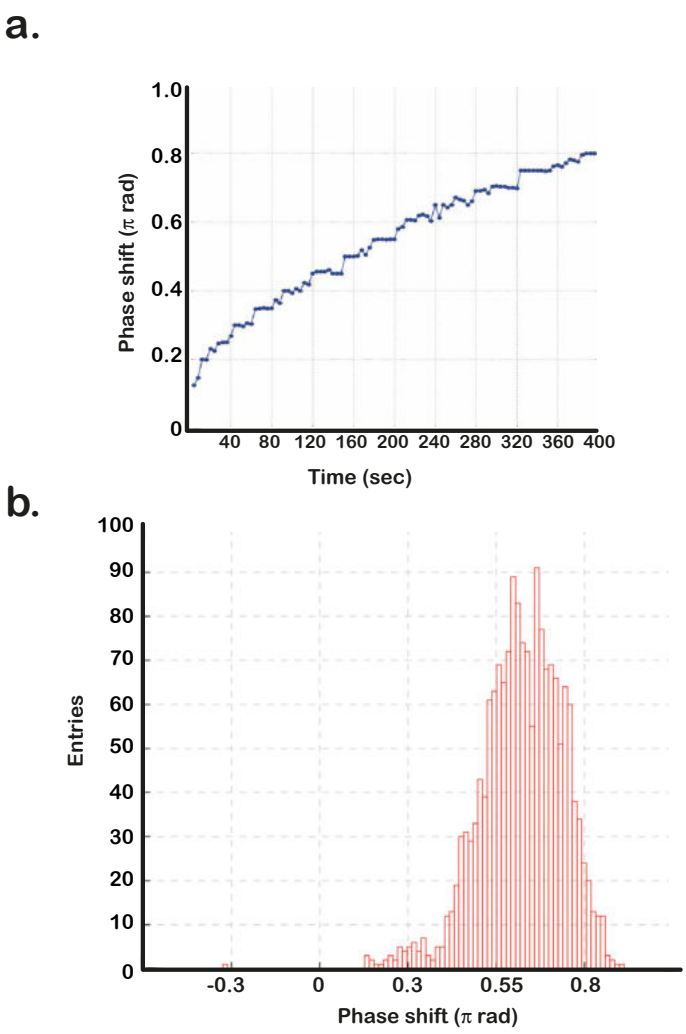


Figure 3.53 Phase Development in an ADP-HSET Dataset Collected on the Titan Krios

(a) The phase development of the VPP in the Titan Krios at LONCEM shows a longer activation time and the possibility to collect a greater number of exposures per position than the Glacios; (b) the histogram obtained after pre-processing the dataset shows a desirable distribution of phase shifts across the micrographs with most occurring in the range of 0.5-0.6π rad phase angles.

The slower phase development and shorter exposures enabled me to collect ~100 images per position of the VPP. The histogram of the phase shifts in fig 3.53 b revealed a unimodal distribution centred around 0.55-0.6π rad, which is close to the region of maximal contrast i.e., 0.5π rad. While there was a tendency for the VPP to over-develop in this case as well, it was significantly less pronounced than with the Glacios. Hence, unlike the Glacios data that was 70%

unusable, 70% of the Krios dataset remained usable after I cleaned up the dataset.

3.6.2.1 Particle picking

For particle picking, I had to create a new picking model due to the difference in pixel size and contrast in the images in comparison to the Glacios datasets. My training set comprised thirty-six micrographs whose phase shifts were around 0.45-0.55π rad i.e., the region of maximal phase contrast (Danev, Tegunov et al. 2017). Although I used JANNI to denoise my images, the default JANNI denoising model seemed less successful than in the case of the Glacios dataset (fig 3.54 b).

As can be seen in fig 3.54 c, the picker that used the default denoising model missed some target particles and picked quite a few false positives. I rationalised that the picker was misled by the denoising which did appear to enhance the visibility of the background as well to some extent. Therefore, I created a custom JANNI denoising model that was trained exclusively on a small subset of 20 movies from my Krios dataset. Fig 3.54 d shows a noticeable difference in enhancement of the target in the custom model as opposed to the default JANNI model which enhanced the visibility of the background as well. One can also notice that the image denoised with the default denoising model (fig 3.54 b) is grainier than the one denoised with custom model (fig 3.54 d).

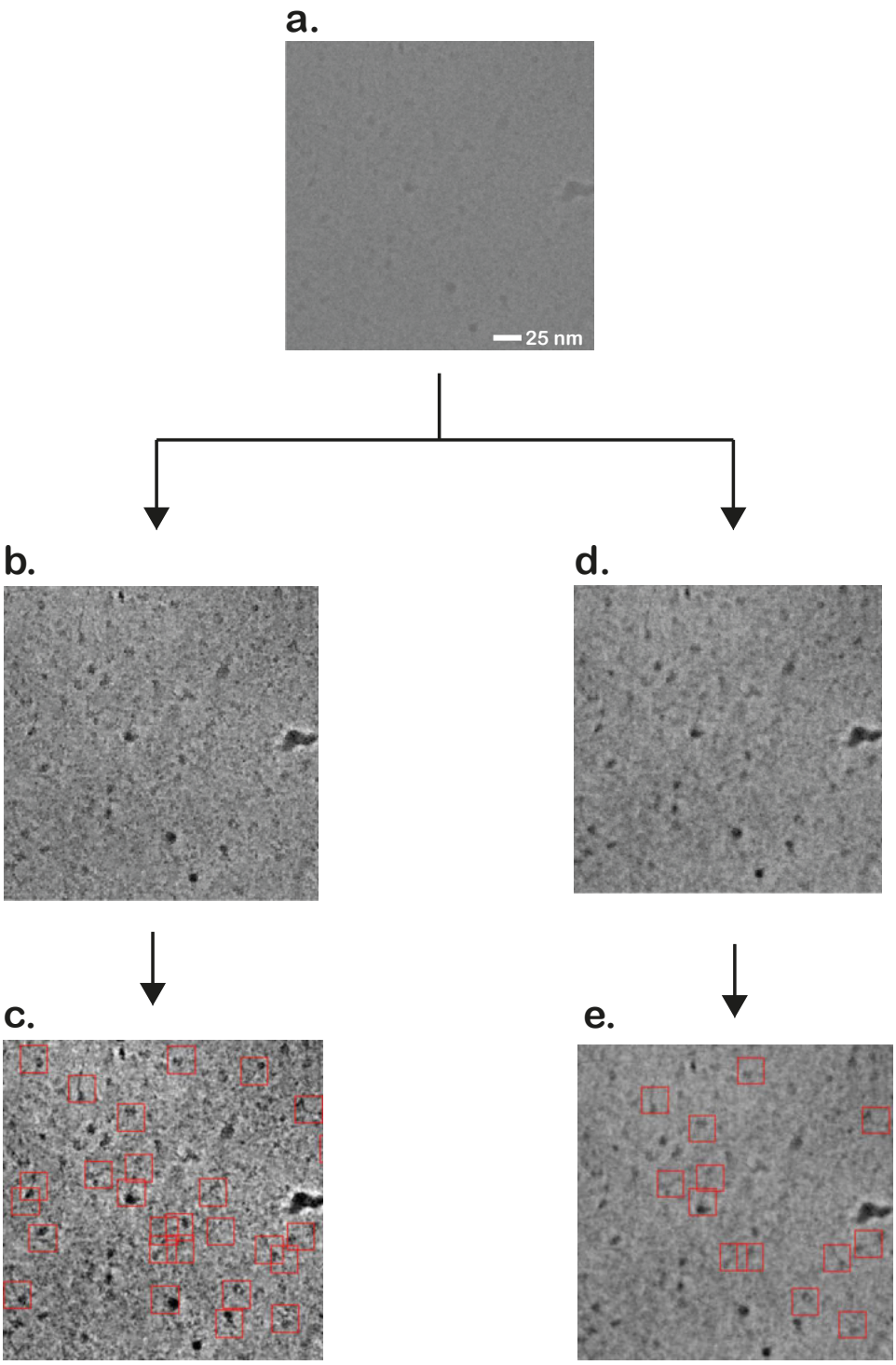


Figure 3.54 Testing Different Denoising Models to Aid Particle Picking in CRYOLO

(a) An unfiltered sample micrograph from the training set is denoised using (b) default JANNI denoiser and (d) custom denoising model to train a weighted-picking model; the model that was trained by (c) the default denoiser shows more coordinates and many off-target picks, while training by the (e) custom denoiser shows better accuracy.

Customising the denoiser meant creating a weighted picking model from scratch, rather than refining the pre-existing general model. Initially, this resulted in a poor yield of particles due to the absence of pre-existing weights in the model. Hence, I adopted the approach of iterative training in which the picked coordinates of the current model were manually corrected and used to train the next model (fig 3.55). This reinforced the correct choices made by the picker, resulting in an increased yield of true positives (fig 3.55 c). A grand total of 418,000 particles were picked from the entire dataset and extracted by applying a threefold down sampling rate in RELION.

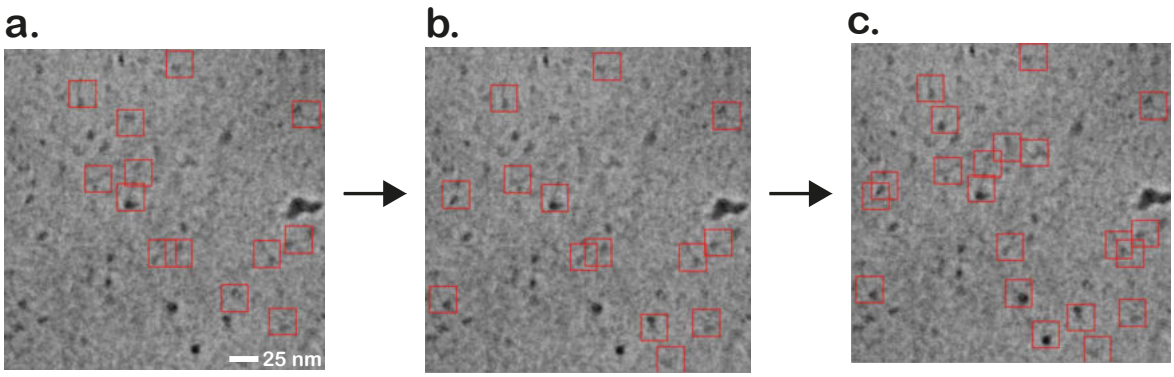


Figure 3.55 Iterative Optimisation of Picking Model in CRYOLO

Micrographs were denoised and picked using a customised model; the model was improved over three iterations by manually editing the coordinates picked by (a) the initial model to train (b) the second generation, which was subsequently used to train the (c) third generation.

3.6.2.2 2D Classification

Unlike the Glacios data which were split into their individual sessions for independent processing, the Krios dataset was processed in a more straightforward manner, as a single stack. The 418,000 extracted particles were classified into 200 initial classes to confirm the identity of the picked coordinates and eliminate the false positives. The classes that best resembled the previously established HSET shape are shown in fig 3.56 a. In contrast to the Glacios classes in fig 3.48 c, these classes already seem to appear considerably better defined and start to show a hint of detail near the neck linker in class 1. However, the accumulation of around 3,000 particles per class implied a high likelihood of

over-averaging, which I chose to tackle by selection and reclassification, similar to the Glacios dataset. In this case, the result was a success and showed a visible improvement in the detail of the class averages by the third round (fig 3.56 c).

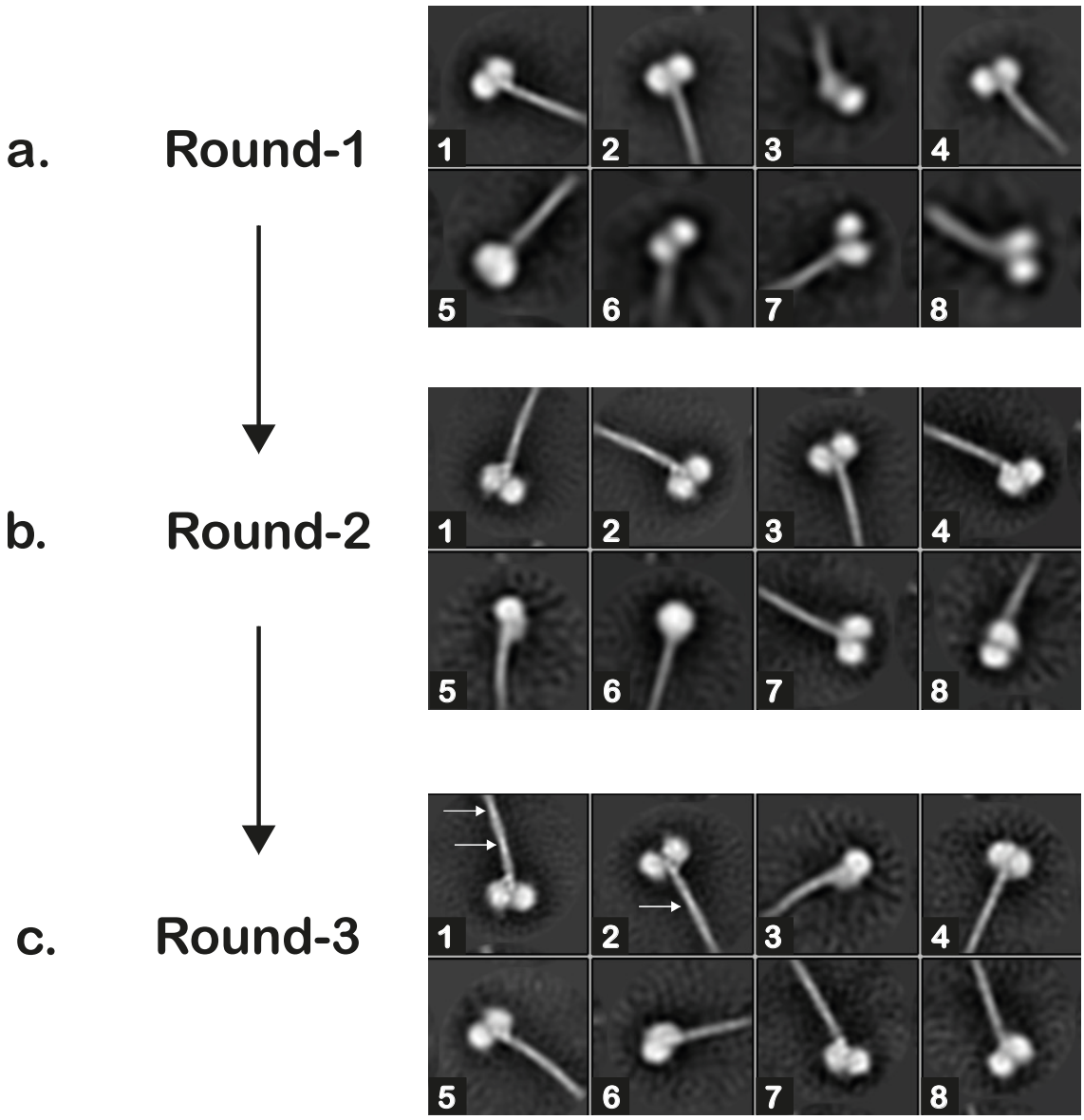


Figure 3.56 Treating the Over-Averaging Issue in 2D Classes of ADP-HSET

2D averages obtained after (a) the first round of classifying ADP-HSET appear blurred due to over-averaging; this issue was treated by iteratively selecting and reclassifying the particles from the best classes over two rounds. (b) the classes from the second round begin to appear sharper and those in the (c) third round start to show evidence of turns and crossovers in the coiled coil as alternative fuzzy and bright regions in the density respectively; crossovers are marked by white arrows.

One can clearly see the classes becoming progressively sharper from fig 3.56 a to c, and gradually beginning to reveal additional features in the motor domains rather than showing just the white dumbbells (class 1, 2 and 7, fig 3.56 c). In addition, the coiled coil shows an alternation between fuzzy regions and bright spots of density (class 1, fig 3.56 c), which are the first convincing glimpses of the twisting strands and their crossover points respectively (arrows, class 1, fig 3.56 c). The crossover points marked in fig 3.56 c were found to be 24.8 pixels i.e., 50 Å apart, implying the pitch of the coiled coil is around 100 Å. This estimate is close to the typical measurements of the pitch exhibited by coiled coil domains in proteins (Seo and Cohen, 1993), thereby validating the features observed in the 2D classes. Furthermore, this progress confirmed my suspicion about the resolution being limited by over-averaging of classes. However, the absence of information beyond two crossovers of the coiled coil, combined with the stagnation in the resolution of the classes beyond three rounds of classification, made me suspect the distal N-terminus of HSET.

My new hypothesis was based on the fact that the N-terminal CBD could contribute to poor alignment of the entire length of HSET, due to inherent flexibility in the region, lack of sufficient signal, and/or the inclusion of signal from neighbouring molecules within the corners of the box. I tested this by applying a circular mask around the dataset that restricts the information in the images up to a radius of 100 Å from the centre of mass of HSET, which is located closer to the motor domain end of the protein. The result as shown in fig 3.57 b, was the unequivocal appearance of secondary structure information in the motor domains as well as an improved definition of the strands of the coiled coil, implying a resolution higher than 10 Å. After two rounds of the masked classification, RELION quoted a resolution of 8.5 Å in the best classes (classes 1 and 2, fig 3.57 b) which did not improve in subsequent rounds. Attempts at further masking resulted in cropping out information in the motor domains or a grainier appearance of classes that signalled overfitting, with no real change in detail. Nevertheless, I had achieved the minimum level of detail required to begin interpreting the features visible in the classes.

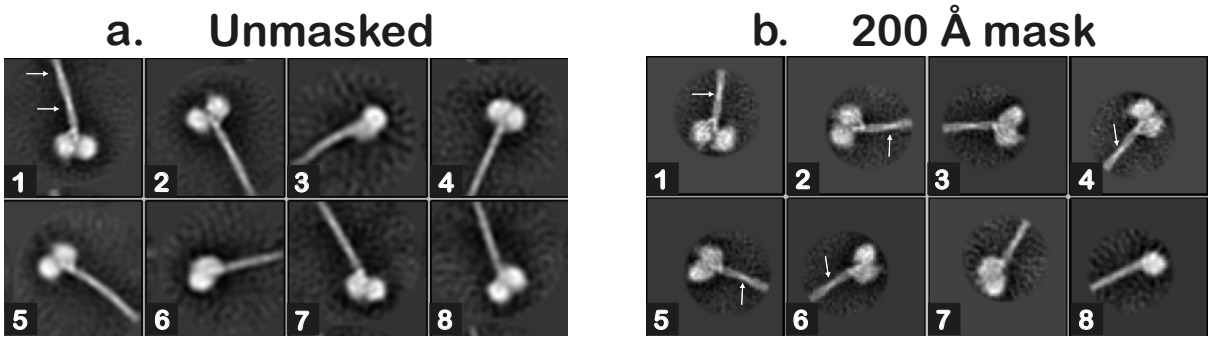


Figure 3.57 Unravelling Secondary Structure Detail by Using a Circular Mask

A circular mask of 200 Å diameter was applied during classification to limit the information and improve alignment of particles. The effect of masking is seen by comparing the appearance of classes (a) before and (b) after the application of the mask; the masked classes show secondary structure in the motor domain and the coiled coil up to a resolution of 8.5 Å; crossovers in the coiled coil are marked by white arrows.

The first observation was the clear and consistent display of the two strands in the coiled coil, with a greater level of clarity in the crossover points in the masked classes (arrows, fig 3.57 b) compared to the unmasked version (fig 3.57 a). Due to the diameter of the mask applied, the information is limited to the first crossover point in comparison to the unmasked class 1 of fig 3.57 a. The next important feature observed was the asymmetry between the two motor domains which was already evident in negative stain EM (see section 3.4.2.3, *Page 104*). One of the motors of the homodimer such as the left motor in class 1 and the top in class 2 consistently show a better resolution than the other, which might be indicative of reduced conformational variability. Furthermore, in the classes 3, 4, and 5, one can observe an apparent extra link between the stalk and the better resolved motor domains, which might be contributing to its stability. While one could notice the tilt between the two motor domains in classes 5-8, they were largely limited to small angles, in comparison to the diversity in views that was observed in the negative stain dataset (fig 3.32 b, *Page 108*). Moreover, any orthogonal views such as class 8 were not as well resolved as the previous classes. These observations suggested a bias in the particle stack towards front-facing views, which had not been detected in negative stain EM ADP-HSET on the same type

of grids. Nevertheless, I proceeded to generate an ab-initio 3D reconstruction using the high-resolution classes.

3.6.2.3 3D Analysis

The ab-initio 3D map was generated using 3,000 particles corresponding to the classes 1, 3, 6, 7, and 8. The map was obtained at a resolution of 9 Å as shown in fig 3.58 b, displaying the separation of the coiled coil strands, as well as a well-resolved motor domain on the right of the stalk in comparison to its less resolved partner on the left. Forward projections of the map in IMAGIC were used to measure the values for semi-helical pitch of the coiled coil, which was found to closely match the value obtained from the 2D classes in fig 3.56 c. Additionally, I attempted to determine the spacing between the strands of coiled coil, which was found to be 10 Å, as shown in the front view (fig 3.58 b). However, the orthogonal views in fig 3.58 a and c show a clear flattening and stretching of the density in this region that increase its width to 16 Å, indicating that the ab-initio model was affected by anisotropy. As a result, the front view was the only orientation of the ab initio map of ADP-HSET that proved to be informative about the coiled coil.

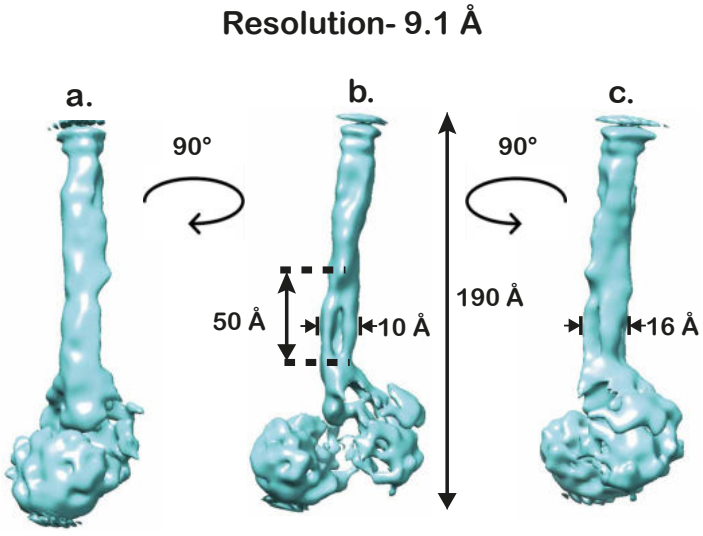


Figure 3.58 3D Ab-initio Reconstruction of ADP-HSET

3D initial model of ADP-HSET up to a length of 190 Å was obtained at a resolution of 9 Å; (b) front view highlights the separation of the individual strands of the coiled coil as well as the asymmetric motor domains; (a,c) the orthogonal side views show a flattening of density indicative of anisotropy due to a lack of information at those orientations.

The map was low pass filtered to 20 Å and used as a reference for a 3D classification of 150,000 particles into 2 classes, without the use of a custom soft-edge mask. This strategy is similar to the one used on the negative stain EM datasets (see section 3.4.2.3, *Page 109*) and resulted in two classes containing 70,000 and 80,000 particles respectively (fig 3.59 b and c). The class in fig 3.59 a which contained fewer particles, converged to the familiar HSET shape, while the latter in fig 3.59 b converged to a poorly resolved, incomplete structure. I thus proceeded with the refinement of the former class. However, I could not refine this 3D class with the parameters that I had used for 2D classification, namely a 3.7-degree angular sampling and 10-pixel translational search range with steps of size 1-pixel. The 18 Å resolution of the refined model was lower than that of the 9 Å ab-initio reference, which was also supported by the visible absence of any detail in the refined map (fig 3.59 c) in comparison to the ab-initio map (fig 3.58). Therefore, I carried out a stepwise optimisation of every available 3D alignment parameter to suit the resolution of details in the FL HSET molecule, as detailed in fig 3.60.

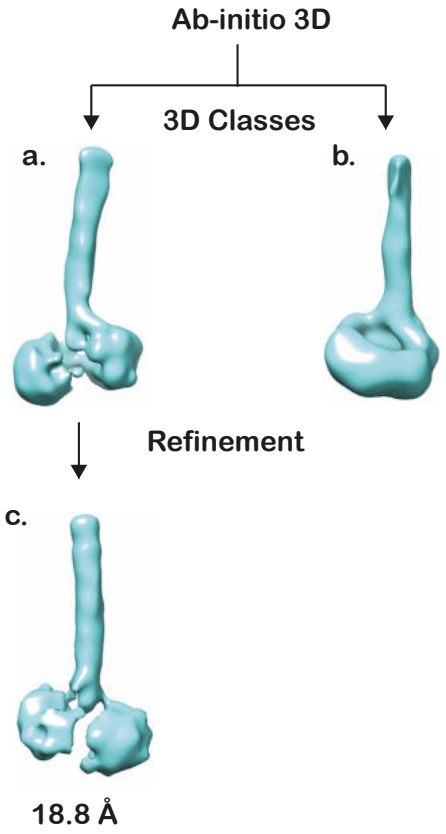


Figure 3.59 An Impasse in 3D Refinement

(a) 3D ab-initio model at a resolution of 9 Å generated from the masked classes is used as reference for a 3D classification of the dataset; (b) the first class appears to resemble the familiar shape of HSET while (c) the second class appears unusable. The first class is refined using fine angular sampling and large translational search range to produce (d) a model that appears to have suffered a loss in resolution when compared to the ab-initio model.

As the ab-initio modelling had been successful, I decided to use those parameters as a starting point i.e., a 15-degree angular sampling with 6-pixel search range and a step size of 2 pixels. The structure I obtained (fig 3.60 b) was quoted by RELION to have a resolution of around 19 Å and revealed significantly less detail than the 18 Å model of its predecessor. This led me to conclude that the alignment required finer sampling. I re-ran the classification with two-fold and four-fold reductions in the angular sampling shown in fig 3.60 c and d respectively, while leaving the translational parameters unchanged. The noticeable improvement of detail in the motor domains from a resolution of 19 Å fig 3.60 b to 15 Å in fig 3.60 d confirmed my hypothesis. In the next step, I attempted to halve the translational

search step from a step-size of 2 pixels (fig 3.60 e) to 1 pixel (fig 3.60 f), while retaining the optimised 3.7-degree angular sampling. Again, there was a visible improvement to a resolution of 13 Å. In a similar fashion, I proceeded to narrow down the optimal translational search range, which resulted in the largest increase in detail, reaching a resolution of 8.7 Å that was comparable to the 2D classes. While the general trend favoured a narrower search range, the quoted resolution as well as the level of detail in the map at a 2-pixel range (fig 3.60 i) showed negligible improvement compared to 5-pixels (fig 3.60 h) and was accompanied by a doubling of computational time. For this reason, I chose 5-pixels and thus arrived at my final set of parameters for the refinement of HSET.

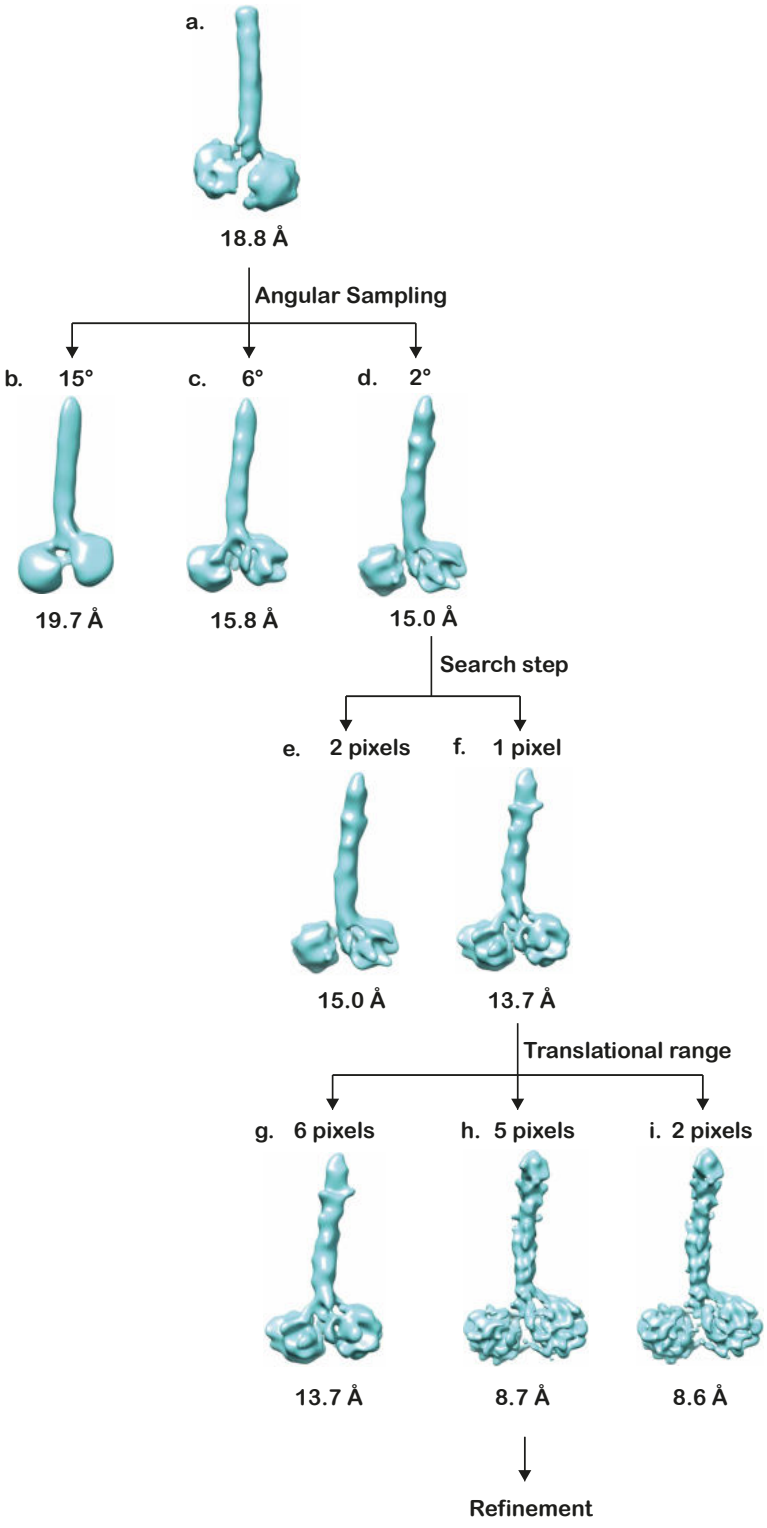


Figure 3.60 Optimising the Parameters for 3D Refinement

Flow-chart shows the pipeline of optimising all parameters to refine the (a) reference model, starting angular sampling ranges of (b) 15°, (c) 7.5° and (d) 3.7°, translational sampling sizes of (e) 2-pixels and (f) 1-pixel and search ranges of (g) 6 pixels, (h) 5 pixels and (i) 2 pixels.

Refinement using these optimised values yielded the first 8.5 Å structure of ADP-HSET up to a length of 190 Å (fig 3.61 b). The map shows two well resolved motor domains and a stalk that is quite noisy and thus not very informative. The details seen in the motor domains show some resemblance to the features observed in the 2D classes from fig 3.57 b with a resolution of 8.5 Å, which matches the value quoted for the 3D map by RELION v3.1. Another detail observed in the 2D classes that is reflected in the 3D map is the asymmetry between the motor domains, with the motor domain on the left being closer to the stalk than the one on the right (fig 3.61 b). As a result, the left motor is slightly elevated and is at a smaller angle to the stalk compared to the one on the right (fig 3.61 b). Additionally, the neck linker connecting the motors to the stalk has more than one connection with the left motor (arrows, fig 3.61 a) in comparison to the single connection on the right (fig 3.61 c), likely also contributing to the proximity of the left motor to the stalk.

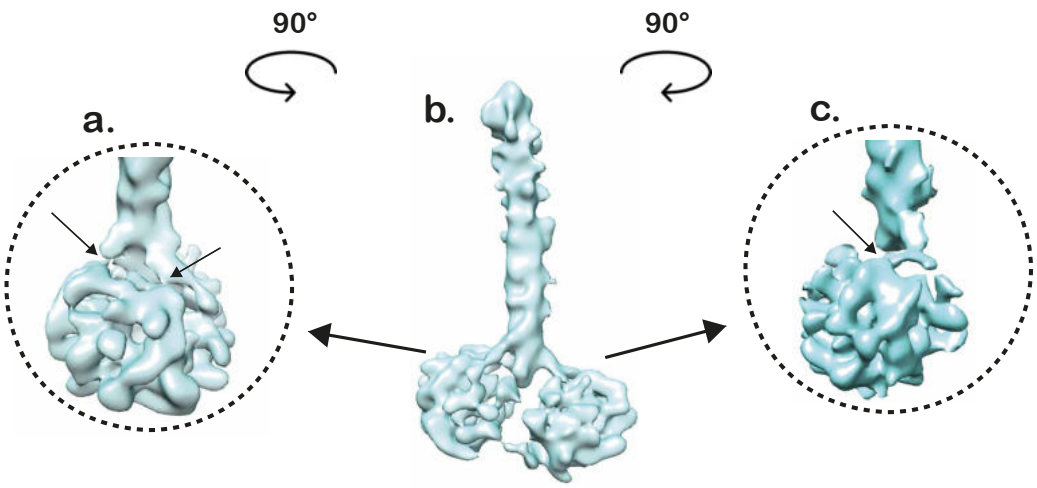


Figure 3.61 Final Refined Map of ADP-FL HSET

(a,c) Orthogonal side views showing magnified motor domains and (b) front view of final refined map of ADP-FL HSET up to 190 Å in length at resolution of 8.5 Å; arrows in the side views demarcate the connections to the stalk.

Unlike the asymmetric map, the atomic model of dimeric HSET obtained from Alphafold (Jumper, Evans et al. 2021) which is based on a crystal structure of dimeric Ncd, has inherited a twofold symmetry about the axis of the stalk (Sablin, Case et al. 1998). As a result, one of the two motor domains shows a poorer fit into the density than the other (fig 3.62 a). The quality of the fit is further compromised by the longer coiled coil helices in the atomic model and a negligible neck linker region, which is at odds with the distinctly visible strand of the neck linker in both motor domains of the map shown above in fig 3.61 a and c. Overall, 62% of the atoms in the Alphafold model fit inside the contour of the map. Attempting to fit the monomers individually to the map (fig 3.62 b) increases this value to 78% of the atoms (fig 3.62 c and d). The displacement of the monomers relative to each other brought about by their individual fitting is apparent in the stalk, where one of the strands is more excluded from the density than the other, indicating a tilt between them (fig 3.62 b). Additionally, one can notice the strands unable to effectively coil around each other, suggesting that a vertical displacement has moved the helical repeats out of register with one another. Both aspects of the stalk in the model do not reflect real detail and are merely used as qualitative descriptors of the change in fit. Realistic modelling of this region would require flexible fitting, which was not performed in this study. The possible implications of this map and the inferences drawn from the homology models will be discussed at length in chapter 4.5 (*Page 181*).

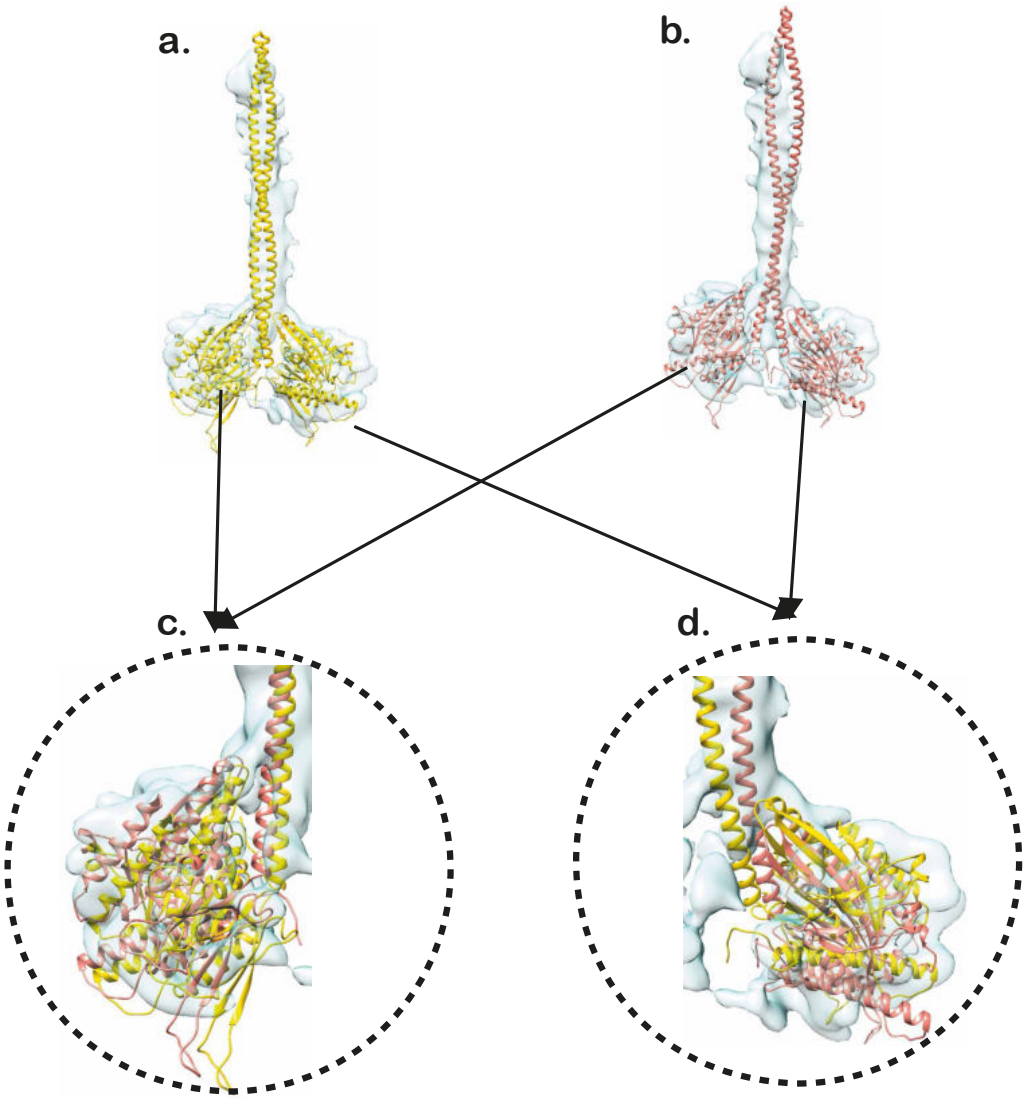


Figure 3.62 Homology Modelling Highlights Divergence from Crystal Structure

Refined cryo EM map of ADP-HSET is overlaid with (a) two monomeric chains that were fitted to individual motor domains and (b) a single dimeric Alphafold model of HSET; enlarged views shows the discrepancies between the two aligned models in the motor domains on the (c) left and (d) right of the stalk.

4. DISCUSSION

4.1 Evaluating the Stability of FL HSET

The feasibility of HSET as a target for structural investigation has always been uncertain. As mentioned in section 1.6 (*Page 30*), the crystallisation of FL HSET has been consistently unsuccessful, thereby limiting the structural understanding of the protein to its motor domain (Park, Ma et al. 2017). Moreover, while the motor domain was successfully crystallised in its ADP-bound state, the structures of Apo- and inhibitor-bound forms of HSET or any other kinesin-14 family member are yet to be solved. Therefore, a logical approach was to initiate the structure determination of FL HSET using of single particle cryo EM rather than X-ray crystallography. To obtain the best possible sample for cryo EM, I used the biochemical and biophysical characterisation techniques detailed in section 3.2 (*Page 58*). Specifically, the ability to ensure the structural and thermal stabilities of both forms of HSET would be important for structural studies. This chapter discusses the various steps taken to stabilise Apo and ADP-HSET.

4.1.1 Stabilising the structure of FL HSET

As outlined in section 3.2.2 (*Page 62*), thermal denaturation profiles, i.e., melting curves, describe the propensity of a protein to transition from a folded state to an unfolded state upon increasing temperatures. At temperatures less than 25 °C, the process of attaining a well-folded structure is generally more energetically favourable than unfolding (Niesen, Berglund et al. 2007). However, increasing temperatures gradually tip the scales in favour of the latter. The melting temperature thus describes the point at which half of the molecules in the sample are unfolded (Lo, Aulabaugh et al. 2004). The low melting temperature of Apo-HSET (T_m-Apo= 32 ºC) along with its broad and shallow melting curve (top, fig 3.9, *Page 64*) suggest that isolated Apo-HSET has a significant amount of conformational flexibility and may have a reasonable degree of structural instability. ADP-HSET on the other hand, has a much higher melting temperature (T_m-ADP= 50 ºC) and the peak in its first derivative plot is much narrower than

Apo-HSET (bottom, fig 3.9, *Page 64*). Therefore ADP-binding stabilises the motor domain of HSET significantly, which might be beneficial for structural EM studies. This raises the question- is Apo-HSET not a feasible target for structure determination?

This concern was answered by the SAXS experiment that was performed at a significantly lower temperature of 4°C. The Kratky analysis of the SAXS data is a qualitative means of deducing the degree of structure in a protein, based on the shape of the Kratky plot (fig 3.16, *Page 81*). To summarize the analysis of my results, the Kratky plot of Apo-HSET indicated a folded structure that displayed some level of inter-domain mobility. Moreover, the plots for Apo and ADP-HSET overlapped completely, suggesting that the difference in structural stability between the two forms of HSET was indistinguishable. This was in stark contrast to the behaviour of the two samples at physiological temperatures during DSF. Thus, SAXS established that Apo-HSET can be pursued by cryo EM, provided it was maintained at 4°C during storage and sample preparation.

4.1.2 Minimising the polydispersity of HSET samples

In addition to having a folded structure, another metric to evaluate the suitability of a sample for cryo EM is the dispersity of the sample due to aggregation or denaturation. This was studied in the undiluted concentration of ~ 0.5 mg/ml using DLS, and at a thirty-fold dilution using mass photometry. The DLS showed a tendency for aggregation in both forms of HSET at a temperature of 20°C, which was slightly higher for Apo-HSET. The level of aggregation was found to reduce in both samples at lower temperatures (fig 3.11, *Page 69*). Thus, DLS confirmed that dispersity of the HSET samples, similar to their structural stabilities, could be improved by the reduction of temperature.

In line with these results, mass photometry was performed on HSET samples that were around 4°C. These experiments provided a measure of the dispersity as a function of the concentration of buffer components such as salt

(fig 3.12, *Page 72*) and glycerol (fig 3.13, *Page 74*). As mentioned in section 3.2.5 (*Page 70*), glycerol is infamous for its high density which greatly impacts the contrast and limits the visibility of the target molecules (Drulyte, Johnson et al. 2018). Salts on the other hand, reduce the contrast of an image by the excessive scattering of electrons by metal ions like Na+ (Sodium) and K+ (Potassium) (Bhella 2019). However, the undesired glycerol also ensures that the protein is well-folded and keeps aggregation in check by providing an amphophilic solvation layer around protein molecules (Vagenende, Yap et al. 2009). The salt content on the other hand, is the irreplaceable source of the ionic strength required to solubilise proteins (Timasheff 1993). In other words, both components that are detrimental to EM play significant roles in the stabilisation of proteins. It is therefore imperative to reach a compromise by arriving at concentrations that are sufficient to maintain the integrity of the protein sample, while also being as low as possible to limit their negative impact on the contrast in cryo-electron micrographs. The titration of the salt and glycerol showed that the samples of Apo and ADP-HSET contained a homogeneous population of more than 90% dimeric HSET, provided the buffer contained at least 140 mM of salt and 0.4% (v/v) glycerol.

4.1.3 Summary

In this manner, I was able to employ different types of biochemical and biophysical characterisation approaches to help me identify the buffer composition and temperature that would maximise my chances of successfully imaging Apo and ADP-FL HSET by cryo EM.

4.2 Describing the Size and Shape of HSET

Some of the most important attributes of macromolecules are size and shape, and these parameters can be studied by different biophysical methods. Most proteins and protein-complexes are typically of two kinds of shapes- globular or extended (Erickson 2009). While the word globular implies compact and does not indicate spherical shapes in most cases, the size of a globular protein can be defined by approximating the shape of the protein to a sphere (Pusey, 1972). In contrast, proteins which contain extended structures would require more than one dimension to define them accurately. However, most software that estimate molecular mass or size operate under the assumption of globularity, and molecular standards designed to calibrate instruments typically comprise globular proteins (Erickson 2009). As a result, many biophysical techniques characterise elongated proteins based on the radius of a sphere, which is inaccurate and uninformative. Being a member of the kinesin family, HSET is known to possess an elongated coiled coil stalk that spans 45% of its sequence (Vale and Fletterick 1997) and is predicted to be at least 200 Å in length by Alphafold. This makes HSET a prime candidate for erroneous characterisation that can only be solved using multiple orthogonal techniques, particularly those that make anisotropic measurements to acquire information about the shape of HSET.

4.2.1 Oligomerisation state

The first hint at the oligomeric state of FL HSET came from Size Exclusion Chromatography, which can be used to estimate molecular weight as a function of elution volume (Erickson 2009), when calibrated with globular protein standards (Supplementary fig. S2, *Page 201*). Based on sections 3.1.2.2 (*Page 52*) and 3.1.3.1 (*Page 54*), Apo-HSET was found to elute around the 10 ml mark in a Superdex 200 10/300GL column, while both forms of HSET eluted around 14 ml in a Superose 6 Increase 10/300 SEC column. Both volumes correspond approximately to the elution of calibrant protein whose molecular weight is 470 kDa (Supplementary fig S2). Based on the 77 kDa mass of a single chain of $Strep_2$-FL HSET that was calculated from its sequence, the prediction of 470 kDa would imply that Apo and ADP-HSET are hexamers.

Herein lies the shortcoming of molecular weight estimation by SEC- the inability to disentangle the shape, size, and hydration state of the proteins passing through the column (Irvine 2001). In order to determine the molecular weight of HSET independent of its shape or hydration, I decided to use a static light scattering (LS) approach such as SEC-MALS (see section 3.2.3, *Page 65*). As explained in section 3.2.3.1 (*Page* 65), the molecular weight of the protein exiting the column is calculated from first principles using Rayleigh's LS equation, rather than being extrapolated from calibration curves (Wen, Arakawa et al. 1996). In other words, SEC-MALS provides a means for the absolute determination of molecular weight in contrast to its relative estimation by SEC. The true molecular weight of both Apo and ADP-HSET were thus found to be close to 155 kDa, implying that both forms of HSET were homodimeric as opposed to the hexameric state estimated by SEC. This oligomeric state was also confirmed by mass photometry and SEC-SAXS and is consistent with reports of dimeric kinesin-14 family members from other organisms (Endow, Henikoff et al. 1990, Meluh and Rose 1990). Moreover, both Apo-HSET and ADP-bound HSET are shown to possess the same oligomeric state, which matches the intuitive image of a kinesin moving along microtubules, remaining a biped while cycling through different nucleotide-bound states.

4.2.2 Size and shape

While SEC-MALS successfully provided the absolute molecular weight of HSET, it was unable to provide information about the shape of the protein due to the large wavelength of the laser used compared to the size of HSET, giving rise to isotropic scattering, as explained in section 3.2.3.1 (*Page 65*). I was, however, able to obtain this information from alternate biophysical approaches such as SAXS, as the wavelength of X-rays is smaller than HSET. This makes the scattering process anisotropic i.e., it is sensitive to the shape of HSET. The indirect Fourier transform of the scattering intensities obtained from SAXS results in a P(r) function (fig 3.17 a, *Page 83*), which is a distribution of the interatomic distances within the HSET molecule, calculated in pairs. According to the

distribution, the shape of HSET geometrically resembles a cylinder to a certain degree, whose maximum dimension "D_{max}" is 368 or 374 Å (or 36.8 or 37.4 nm) in the Apo and ADP-bound forms, respectively. The elongated shapes of both forms of HSET could be visualised in the form of 3D dummy atom models that provided the first glimpse of HSET in solution (fig 3.17 b and c, *Page 83*). As the nominal resolution of these models is low (around 20-30 Å) and they are generated based on data limited to low q values like in my case, they cannot be used to discern the domain architecture and further details about the structures of the two forms of HSET.

4.2.3 Summary

Overall, I used two static light scattering techniques SEC-MALS and SEC-SAXS to determine the shape, size, and stoichiometry of the subunits of FL HSET in solution. This information was important in helping me visually identify HSET particles in electron micrographs.

4.3 EM-based Structure Determination of a Small and Elongated Target like FL HSET

A major consequence of working on a small target with an unconventional shape such as HSET is the opportunity to pioneer every step in the cryo EM structure determination pipeline. From the type of grids to the selection of regions in the grid that are worth imaging and most importantly, the choice of parameters for accurate image-processing, there were several opportunities to understand and appreciate the uniqueness of my target. The biophysical studies helped identify the conditions that stabilised HSET, which functioned as a starting point for the sample preparation in negative staining and cryo EM. Aside from the visual evaluation of the HSET sample, negative stain EM was instrumental in uncovering a plethora of information that helped every stage of the cryo EM pipeline, as described in this chapter. Perhaps the most crucial benefit was its impact on sample preparation for cryo EM. The negative stain images in fig 3.29 (*Page 103*) confirmed that Quantifoil grids with carbon support were the best choice for both forms of HSET. It was also invaluable in establishing the sample concentrations that gave an optimal distribution of HSET particles on the grid. Unlike negative stain, the correlation between the applied sample concentration and particle density visible on a cryo-grid is difficult to establish because the visibility of particles is affected by the ice thickness on the grid. For instance, aside from a low concentration of the applied sample, the absence of particles on a cryo-electron micrograph could be the result of the ice on the grid being too thick or too thin. The concentration that proved to yield optimal particle distribution in negative stain EM cannot be transposed to unsupported grids, due to the presence of support film in the former. However, in the case of supported grids, the conditions used for negative stain can be extrapolated to cryo-conditions, which significantly reduced the number of parameters to be optimised from scratch.

4.3.1 Particle-picking

The other benefit of negative stain EM was the availability of a small-scale high-contrast dataset to trial and optimise strategies for HSET structure determination,

starting with the particle-picking procedure. The two commonly used types of picking algorithms filter the micrographs and subsequently match areas of the filtered micrographs to a set of templates. The first, now known as blob pickers, use a soft-edged circular disk of a user-specified diameter as the template (Sigworth 2004). The second, known as semi-automated pickers, use the reference-free 2D class averages generated from a set of manually picked particles as the templates for auto-picking (Scheres 2015). The absence of a manual picking step combined with the use of a circular disk as a template to speed up calculations implies that blob pickers are fully automatic and can pick particles from an entire dataset in real-time. However, their accuracy is dependent on contrast, and is therefore significantly reduced for small particles that deviate from a globular shape (Grant, Rohou et al. 2018) such as HSET. The semi-automated pickers overcome this issue by using the 2D classes and are thus sensitive to the shape of the particles. The success of the algorithm is highly dependent on the ability of the software to accurately align and classify the manually picked particles, which is not trivial to achieve, especially for a *de novo* system. As section 3.4.1 (*Page 97*) showed for the His_6-FL HSET dataset, most classes that resulted from the use of default alignment and classification parameters did not adequately capture the defining features of the molecule and therefore served as poor templates for the semi-auto picker in RELION. Thus, non-globular particles such as HSET may require optimisation of 2D processing parameters prior to auto-picking.

Unlike its contemporaries, CRYOLO (Wagner, Merino et al. 2019) is one of a new generation of pickers like Deep picker (Wang, Gong et al. 2016) and Topaz (Bepler, Morin et al. 2019) that train convolutional neural networks (CNNs) for fast and accurate object recognition . CNN pickers work by predicting the coordinates of target particles in a training set of images and scoring the predictions against the manually picked coordinates of the target. The score influences the next round of predictions, resulting in an iterative loop that continues until the score shows no further improvement (Schmidhuber 2015). This exhaustive feedback loop imparts a high degree of accuracy to the final picking model with minimal user involvement (Wagner, Merino et al. 2019). Though the negative stain picking model could not be used on cryo EM data due

to the reversal of contrast, CRYOLO did generate a reliable set of picked coordinates in HSET cryo-datasets as well (see fig 3.46, *Page 136* and fig 3.54, *Page 150*).

One of the features in CRYOLO that proved to be especially useful was the choice of a denoiser in addition to a conventional low-pass filter, which is usually the only option available in most conventional pickers. The latter enhances the contrast of the particles by filtering out the high frequency components above a user-defined cut-off spatial frequency, leaving behind only the low frequency components that are crucial for particle identification (Guo and Jiang 2014). While this approach of boosting the signal in the entire micrograph is effective in aiding the manual selection of target particles (see fig 3.46, *Page 136*), it does not enable automated pickers to distinguish target from background. By contrast, JANNI is trained on cryo-datasets of different molecular systems to specifically boost the contrast of the target particles against the background (Wagner and Raunser 2020). It is thus not surprising that the picking model trained on low-pass filtered data picked more false positives than the JANNI-based picking model in the Glacios dataset in fig 3.46. However, it is important to note that the pretrained default JANNI denoising model is still prone to errors, because it is trained on a non-HSET training set. To obtain a denoising model specific for the shape of HSET and the background offered by the carbon support film, I decided to create my own JANNI model using a few movies of varying contrast levels from my Krios dataset (see section 3.6.2.1, *Page 148*). As shown in fig 3.54 (*Page 149*), the custom denoising model enabled CRYOLO to recognize HSET particles that are more representative of the training set, with fewer false positives than the default denoising approach. Although the picking was not exhaustive, this was subsequently improved by iterative retraining of the CRYOLO model to maximize the number of true positive coordinates being picked per micrograph (fig 3.55, *Page 150*).

4.3.2 2D Classification

Based on the success of the neural network-based particle picking approaches described above, it seems that "iterative and exhaustive" is the recipe for successfully picking HSET particles. Not surprisingly, this also proved to be the case for the reference-free 2D alignment and classification of HSET (fig 3.28, *Page 101*). Of all the classifiers that were tested, ISAC operates by a mechanism that thoroughly evaluates the assignment of the particles to their classes after each iteration. In this evaluation, the orientations of the aligned particles within each class are randomised and realigned to the class average repeatedly to be able to calculate the deviation in the alignments (Penczek, Radermacher et al. 1992). The deviation between the multiple alignments, calculated for a single image, is measured in pixels, and expressed in terms of the pixel-error-threshold (PET) (Yang, Fang et al. 2012). Typically, a PET of 0.7 is used in ISAC for globular targets. However, accounting for the elongated nature and flexibility of the stalk in HSET, I used a generous PET of 2.1. Any particles whose alignments showed a standard deviation higher than the threshold were removed and reclassified, resulting in a high degree of purity within the classes. Consequently, the class-averages of HSET produced by ISAC were of superior signal to noise ratio than those produced by the other algorithms that I tested (fig 3.30, *Page 105*).

Another distinction that sets ISAC apart from RELION and cryoSPARC is that the alignments in ISAC are based on cross-correlation (Yang, Fang et al. 2012), whereas RELION (Scheres 2012) and cryoSPARC (Punjani, Rubinstein et al. 2017) use maximum likelihood-based particle alignment. The strength of maximum likelihood-based alignment and classification lies in its treatment of the uncertainty that is inherent to the alignment of noisy images. This is done by assigning a given particle to more than one class with a weight that is defined by the likelihood of the class being a perfect fit for the particle (Scheres 2010). Despite this cautious approach, the lack of a subsequent evaluation of these alignments makes the maximum likelihood-based algorithms more prone to producing classes with missing details. This effect is striking for targets that display curvature and elongation like HSET as shown in fig 3.30 (*Page 105*). The

classes produced by RELION (fig 3.30 b) and cryoSPARC (fig 3.30 c) show poor recovery of the stalk compared to the ISAC classes of the same sample (fig 3.30 a) that manage to go as far as highlighting details around the most mobile part of HSET- the CBD. The low quality of the classes in RELION and cryoSPARC is compounded by the fact that the distribution of particles between classes is not balanced, which results in many empty classes and the majority of particles accumulating in a few classes with a high degree of heterogeneity (fig 3.27, *Page 98* and fig 3.30, *Page 105*). This phenomenon can be linked to the fact that both RELION and cryoSPARC employ an approach known as “k-means” clustering, which segregates the particles into a user-defined number (k) of clusters i.e., classes, with the objective of maximising inter-class variance (Scheres 2010, Punjani, Rubinstein et al. 2017). Having the user estimate the number of subsets in the population of particles may very well lead to under-sampling of distinct views in the classes. As a result, most of the classes suffer from heterogeneity and produce noisy averages that are dominated by low-frequency features. This in turn causes poor alignments leading to a cumulative amplification of heterogeneity. In contrast, ISAC uses a variation of this approach that adds an extra constraint to the inter-class variance optimisation- ensuring a nearly identical number of particles in every class. This “EqK-means” approach requires the user to define the average number of particles that are expected in every class, rather than the expected number of classes. Restricting the particle count per class minimises the occurrence of scenarios such as the accumulation of more than one population in some classes and the collapse of other classes. However, the computational inefficiency that results from the exhaustive nature of ISAC, along with the inflexibility of the data formats in the software package mentioned in section 3.4.2 (*Page 101*), made the SPHIRE suite unsuitable for a scale-up to larger datasets in the future of the project.

Thus, I optimised maximum likelihood-based algorithms in the negative stain dataset, to reproduce the ISAC output as closely as possible. A common approach to alleviate heterogeneity in Maximum Likelihood-based algorithms is by carrying out several rounds of reclassification which provide sufficient clusters to sample the diversity in the data, while also selecting the particles after each round to manually filter the stack (see sections 3.6.1.3, *Page 137* and 3.6.2.2,

Page 150). However, an approach that treats the root cause of the issue rather than the symptom is the optimisation of image alignment parameters. While this is limited to circular masking in cryoSPARC, RELION has additional parameters such as values for in-plane angular and translational sampling, which exerts a considerable amount of control over the alignment process. As shown in section 3.4.2.2 (*Page 104*) and section 3.6.2.2 (*Page 150*), varying these parameters has a significant impact on the accuracy of the alignment, as well as the density and homogeneity of the particles in each class, and therefore, the level of detail that is visible in the class averages. The availability of optimisable parameters shows that RELION can overcome some limitations inherent to the algorithm, providing the scope to analyse targets of different sizes and shapes. With the Krios data, I used the optimised parameters and a circular mask in RELION v3.1 to obtain classes of dimeric HSET at a resolution of 8.5 Å.

4.3.3 3D Analysis

As the optimised sampling parameters were significantly different from their default options, it is no surprise that I could not refine the 3D classes of the negative stain dataset in RELION (section 3.4.2.3, *Page 109*). The success of the 3D classification of the negative stain dataset can be attributed to the high contrast in the images that made them easier to align than the cryo-images, despite the unoptimized sampling parameters (fig 3.35, *Page 114*). Nevertheless, the stalled refinement is consistent with the limited resolution that arose because of inaccurate particle alignment. This was also the case for the cryo EM datasets obtained from the Glacios and the Krios. Arguably, the analysis of the Glacios dataset was also compromised by phase-plate overdevelopment (see section 3.6.1.1, *Page 129*). However, even in the case of the uncompromised Krios data that resolved up to 8.5 Å of detail in 2D classes, the direct transposition of the 2D alignment parameters to 3D did not yield well-resolved 3D classes (fig 3.59, *Page 156*). In fact, it worsened the level of detail in the cryo EM map, despite the addition of more particles known from the 2D classes to contain high resolution information. While the finer angular and translational sampling rates were

consistent with improved resolution in fig 3.60 (*Page 159*), high values of the translational search range were not beneficial. On the contrary, the higher resolution detail was favoured by lower values of the translational search range (fig 3.60 h). This inverse relation seemed reminiscent of the improvement in resolution observed in the 2D classes due to the use of circular masks (fig 3.57, *Page 153*). In other words, limiting the translational search range reduces the extent of information from the molecular images used for alignment and classification, similar to the effect of the circular masks. Thus, although the parameters used in the 2D and 3D particle alignments are not identical, they seem to have the same effect on the data.

The improved resolution of the Krios map allowed visualisation of the secondary structure information in the coiled coil and the motor domains of HSET. However, although other views are present, the 2D class averages are dominated by front facing views which lead to anisotropy in the resulting 3D density map (fig 3.58 a and c, *Page 154*). This map thus revealed a challenging issue in the sample preparation- the preferred orientation adopted by HSET on supported grids. However, for reasons currently unknown, this was not observed in the negative stain dataset. Preferred orientation on supported grids could be overcome by glow-discharging in the presence of polyamines like pentylamine. Pentylamine deposits a positive charge on the surface of the grid, as opposed to the negative charge obtained by glow-discharging in air (Dubochet, Ducommun et al. 1971). Glow-discharging with pentylamine has been shown to result in a re-orientation of nearly 90% of the particles, introducing a high degree of rotational heterogeneity in the population (Morris and da Fonseca 2017). However, I discovered the presence of preferential orientations late in my project and therefore, did not have the time to investigate glow-discharging with pentylamine. Instead, I focussed on maximising the available data from side views by collecting as much data as possible from the existing grids. However, this did not lead to a sufficient improvement due to the issues with the phase plate, that reduced the amount of usable data to 2,000-3,000 movies (compared to the expected value of at least 12,000 movies).

4.3.4 Summary

Overall, extracting structural information from a small and elongated protein like HSET required an extensive amount of customisation in each step of the processing pipeline. In addition to the dimensions of HSET, a major factor that made the image processing challenging was the poor contrast, limited further by the unavoidable use of carbon supported grids that contributed to a significant level of background. Nevertheless, the project has reached a stage where preferred orientation appears to be the only obstacle in the way of resolving the side chains of HSET. Therefore, collecting more data from existing grids using a Krios with a non-problematic VPP is the first approach to improve the resolution of the maps. Depending on the outcome, the pentylamine strategy could be an alternative to remove the preferential HSET orientations on the carbon supported grids.

4.4 Phase Plate- Advantageous or Disadvantageous?

In a technique like single particle cryo EM that involves identifying, aligning and back-projecting noisy molecular images to create 3D reconstructions, one crucial property of the images- contrast, is the cornerstone of success. As explained in chapter 3.3 (*Page 88*), proteins are weak scatterers of electrons and are largely rendered visible by the difference in phase between scattered and unscattered electrons (phase contrast) rather than a difference in numbers (amplitude contrast) (Orlova and Saibil 2011). The most common approach to enhancing the phase shifts is defocussing the objective lens, which results in a trade-off between contrast and resolution (see section 3.3.3.2, *Page 91*). Its alternative, a device known as a Volta Phase Plate, was designed to enhance the contrast of cryo-images without compromising their resolution, while operating the microscope in-focus (Danev and Baumeister 2016) (see section 3.3.3.3, *Page 95*). However, in-focus acquisition demanded an impractical level of accuracy and was therefore replaced by a defocussed VPP acquisition. This was a breakthrough in the field of phase plate development as it made the VPP more feasible in routine data collections. The defocussed VPP acquisition also proved beneficial in solving the structures of difficult targets (Khoshouei, Radjainia et al. 2016). However, this approach polarised the cryo EM community.

To a significant part of the single particle cryo EM community that worked on compact targets that were readily visible with routine defocus, the VPP represented extra set-up time and reduced the throughput of data acquisition compared to routine defocussed acquisitions. The developers' claim that the VPP resulted in a 20% signal loss added to the backlash (Danev, Tegunov et al. 2017). Regardless of the evidence of improved particle picking, alignments, and 3D variability analyses in VPP datasets and resolutions of final structures that were comparable to non-VPP datasets (von Loeffelholz, Papai et al. 2018), the VPP was largely seen as a lazy solution to thick ice films on grids. Its use was actively discouraged by data collection facilities and relegated to a last resort for single particle cryo EM. Data collection facilities such as eBIC (Diamond Light Source, Oxford, UK) discourage VPP data collections today. This was quite concerning for my project, because the successful HSET grids had consistently shown an

irreconcilable level of background during grid optimisation experiments. Nevertheless, I proceeded to explore the phase plate because it seemed to be the best way to enhance the contrast for an elongated thin particle such as the HSET dimer on carbon supported grids.

4.4.1 Initial challenges to VPP data acquisition

The issues that I raised with the VPP are specific to my target protein and the high background on the grids. The factor that affected the quality of my data the most was the activation of the phase plate, a consequence of the on-the-fly development of the beam-induced Volta potential, as demonstrated by the activation profile in fig 3.43 (*Page 130*). While a minimum phase shift of 0.2π rad already showed significant contrast in the case of the 800 kDa GroEL (Danev, Tegunov et al. 2017), a comparable level of contrast was only achieved between 0.3- 0.35π rad for both my Glacios and the Krios HSET data. The contrast began to be compromised by the onset of blurring (see fig 3.45, *Page 134*) at phase angles of 0.7π rad and above, which was less than the upper limit of 0.8π rad used by the developers. Therefore, I was confined to a narrower window of desirable phase angles, significantly reducing the number of exposures per position of the VPP. Additionally, excessive phase development reduced the quality of the data, and the persistence of this trend further lowered the yield of a usable data to 20-28% (table 3.2, *Page 133*). While this phenomenon was acknowledged by the developers as a symptom of higher phase shifts, they did not regard it as a hindrance to their processing pipeline, most likely due to the clear visibility of their 800 kDa target that yielded a 3 Å structure from a mere 400 images (Danev, Tegunov et al. 2017). However, despite the enhanced phase contrast, the low signal of the HSET dimer implied that particles from a dataset twenty times larger had to be averaged to accumulate enough information to reveal the secondary structure of the protein (fig 3.57, *Page 153*). Thus, the greatest limitation in my analysis was the number of usable micrographs in each session. Attempting to compensate for this with more data collection sessions required more microscope time than I could access.

4.4.2 The consequences of over-development

The Glacios datasets of HSET were different from the Krios dataset and the published GroEL dataset because the data could not be improved by simply filtering out the micrographs corresponding to undesirable phase angles. Therefore, I undertook a detailed investigation of the causes and consequences of phase shift over-development (see section 3.6.1.1, *Page 129*). A lot of manual inspection was required to pinpoint the different parameters that could lead to blurring in images, as listed in table 3.2 (*Page 133*). Additionally, a different kind of blurred appearance that can best be described as a blurry and shaky appearance of the image (fig 3.45, *Page 134*) was found to be the result of a phase development so extreme that it crossed π rad. These higher values began to reappear in the development profile of the same position, similar to a reactivation starting from zero, due to the inability of the CTFFind algorithm to distinguish between phase angles of 'x' rad and 'π + x' rad, for any positive value of x. This meant that several images associated with normal metadata values were still contaminating the dataset, because the seemingly normal phase shift was spurious. This compromised the efficiency of filtering out poor quality data and resulted in a lack of higher resolution features in the 2D classes, despite several rounds of cleaning (fig 3.48, *Page 139*), masking (fig 3.49, *Page 140*) and high pass filtering (fig 3.50, *Page 142*). The spurious nature of the data presented itself in the 3D reconstructions as apparently high-resolution detail, which proved inconsistent between the three maps obtained (fig 3.52, *Page 145*). Thus, I could only extract a limited amount of usable data from the Glacios datasets.

4.4.3 The Do's and Don'ts of using a VPP

Based on the stable behaviour of the VPP and the success of the data obtained from the Krios (fig 3.53, *Page 147*), I propose the following causes of the issues encountered above. The foremost is the lack of available expertise with the Glacios-VPP, which would have avoided some of my mistakes and undoubtedly reduced much of the time spent learning to operate it by trial and error. When

collecting data on the LonCEM Krios, Dr. Nora Cronin's guidance made a tremendous difference to my understanding of the VPP and the quality of my dataset, which is evident from the success of the 2D classes and 3D models presented in section 3.6.2.2 (*Page 150*) and 3.6.2.3 (*Page 154*). Another prominent factor in the success of the acquisition was the phase plate heater current. The heater maintains the VPP at a temperature of 200°C to prevent contamination by beam-induced formation of organic deposits on the film, which could cause erratic development of the potential and therefore, the phase shifts. Even in the case of the Krios datasets, over-development was a dominating phenomenon, but was promptly resolved by increasing the phase plate heater current. Extrapolating this to the Glacios, it is likely that the gradually worsening over-development in the phase shifts was due to a contaminated VPP caused by a defective heater. Another difference that could have contributed to the defects in the Glacios dataset was the use of image beam shift. While this was not intentional, EPU provided a default value for minimal beam shift which I was not aware of. In contrast, the Krios dataset was acquired solely using stage shift. It seems intuitive that the movement of the beam would affect the spot on the film that is exposed and disturb the development of the surface potential, thereby distorting the resultant phase shifts that are imposed on the electrons. It is thus highly likely that the accidental inclusion of image beam shift was responsible for limitations other than the erratic development, including the absence of high-resolution features and the appearance of spurious details in the seemingly normal images. In hindsight, a subtle quality control measure was the visibility of aqueous ice rings in the power spectra of the Krios images, a sign of a resolution of up to 3.7 Å (McMullan, Vinothkumar et al. 2015), which was consistently absent in all images collected on the Glacios.

4.4.4 Summary

In summary, the issues faced make the VPP less reproducible and reliable and will have to be addressed. Though the issue of over-development was mitigated by increasing the heater current, I was unable to find a way to avoid it entirely.

Similarly, according to the manual, it is possible to use minimal amounts of beam shift. However, this was found to negatively affect the contrast and it has the potential to reduce the detail in the images irreversibly. Both issues that affect the robustness and reproducibility of the VPP arise from the usage of a film, as do the other drawbacks noted by the developers (Danev, Tegunov et al. 2017). In light of this, the laser phase plate (Schwartz, Axelrod et al. 2019) seems like a promising film-free alternative to achieve enhanced phase contrast. The laser phase plate is based on the principle of retarding electrons using the panderomotive force generated by a continuous pulsed laser. This force can be moderated to increase or decrease the phase shift by changing the power of the laser. Overall, the use of enhanced phase contrast proved to be invaluable to the visualisation of HSET. The phase plate strategy made it possible to solve the structure of the small, elongated, and flexible HSET dimer, on a high-background support film to sub-nanometer resolutions, which is unprecedented in the field of cryo EM. In doing so, it also provided the first glimpse of the structure of a dimeric full-length kinesin-14. Despite the issues faced in the collection and the processing, it is important to note that the factors that ultimately limited the resolution of HSET were not due to the phase plate, but probably more due to the preferential orientations and low sample concentration.

4.5 Interpreting the FL HSET Structure

A factor that plays an important role in deciphering the mode of action of the members of the kinesin-14 family is the ability to visualise their molecular structures at high resolution. As discussed in section 1.4.4 (*Page 21*), the mechanical cycle of kinesin-14 family members such as Kar3 and Ncd was pieced together from the cryo EM maps of MT-bound intermediates (Hirose, Lockhart et al. 1995, Rank, Chen et al. 2012, Zhang, Dai et al. 2015). However, the low resolution of the maps meant that a crucial part of their interpretation required fitting an atomic model to observe the relative displacement between different binding modes (Endres, Yoshioka et al. 2006). The first structure of a kinesin-14 obtained at high resolution was the crystal structure of the monomeric motor domain of Ncd (Sablin, Kull et al. 1996). This structure was the first to show the appearance of the MT-binding motif at a resolution of 2.5 Å and provided insight into nucleotide-binding, through a bound ADP molecule. This was followed by the crystallisation of a larger construct that included residues at the N-terminal end of the motor domain, belonging to the neck linker and a small portion of the coiled coil which is important for dimerization (Sablin, Case et al. 1998). The dimeric structure was solved to the same resolution as the monomer and revealed an alpha helical strand extending from each motor domain and beginning to coil around each other, with the entire structure showing a two-fold symmetry around this coiled coil stalk. To date, this structure is the most complete depiction of a homodimeric kinesin-14 member, and any attempts to crystallize full-length kinesin-14 in a nucleotide-bound or apo state have been unsuccessful. Moreover, kinetic studies revealed an inherent difference in the affinity of each motor domain to ADP, which could not be discerned from the crystal structure (Foster, Mackey et al. 2001). However, there have been no reports of visualising this asymmetry in a kinesin-14 member to date. Therefore, although there is considerable knowledge about the mechanism of the motor-MT interactions that drive kinesin-14 members, there is no molecular understanding into the coupling of the energy from ATP hydrolysis with kinesin-14 movement.

4.5.1 The overall structure of FL HSET

In this work, I present the 3D structures of the human kinesin-14 HSET (KIFC1) which provide the first insight into the full length of a kinesin-14 molecule. The structure of ADP-FL HSET was initially determined at a resolution of 20 Å using negative stain EM, and subsequently resolved to an estimated resolution of 8.5 Å using cryo EM. The overall FL-HSET structure spans a length of 300 Å and encompasses the motor domains, a neck region, a coiled coil stalk, and a cargo-binding tail domain. The cryo EM structure of ADP-HSET shows a total length of 190 Å and lacks the cargo-binding domain (CBD) as well as a portion of the coiled coil stalk. Both the negative stain and cryo EM maps of ADP-HSET show an off-centre stalk i.e., the proximity of the stalk to one of the motor domains, as opposed to being centred and equidistant from both motor domains (fig 4.1 a and b). Full-length Apo-HSET was also solved by negative stain EM and revealed two conformational states that demonstrated two different off-centre motor-stalk arrangements, which I refer to as type K (fig 4.1 a) and type S (fig 4.1 b). The off-centred HSET conformations are different from the recently obtained negative stain reconstruction of kinesin-1 (Tan, Yue et al. 2023) which shows the stalk centred with respect to the two motor domains. It is possible that both families of kinesins attain more states than reported in these studies, sampling other orientations of the stalk with respect to the motor domains. Another possibility is that this observed difference in the position of the stalk relative to the motor domains between a kinesin-14 like HSET and kinesin-1 is responsible for the two key points of divergence between the two families- directionality and processivity of motion. Kinesin-1 members are known to display a processive i.e., "walking" motion towards the growing + end of MTs (Svoboda, Schmidt et al. 1993), while kinesin-14 members move towards the - ends displaying a "hopping" or non-processive motion (McDonald, Stewart et al. 1990).

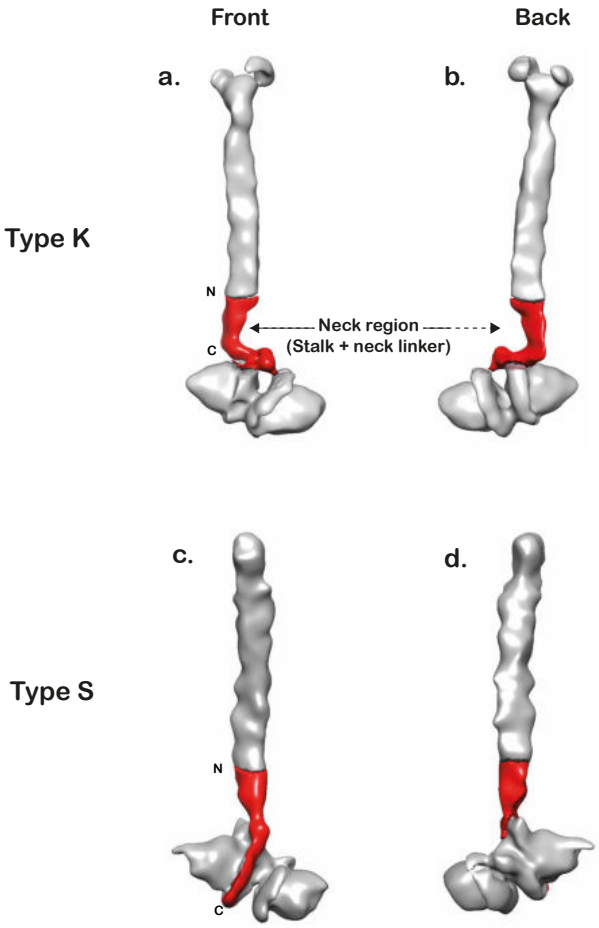


Figure 4.1 Negative stain 3D Reconstructions of Apo-HSET

Conformational states of Apo-HSET reveal the (a) front and (b) back views of the type K (Kinked) state which shows the two motor domains arising from a central point where the kinked neck terminates; (c,d) type S (Straight) state which possesses a straight neck terminating at one of the motor domains; the neck (red) comprises the neck linker and a bit of the coiled coil stalk; N and C denote the corresponding terminals of the neck region.

4.5.2 FL HSET: The neck and stalk

The different conformations in HSET versus kinesin-1 are most likely due to different conformations in the smallest and one of the most important parts of a kinesin's polypeptide chain- the neck linker. The kinesin neck linker is composed of a short stretch of 10-15 amino acids separating the motor from the stalk. This region is crucial in communicating the binding events that occur at the motor domain to the stalk, thereby generating the mechanical force necessary to

drive the kinesin's motion (Sablin, Case et al. 1998, Kozielski, De Bonis et al. 1999). Crystal structures of monomeric kinesin-1 and kinesin-14 Ncd that both consist of the motor domain and the neck region (which comprises the neck linker and a portion of the coiled coil) showed that the main differences between them are in the neck region (fig 4.2), and that the neck linker sequence was conserved in a family-specific manner (Kozielski, De Bonis et al. 1999).

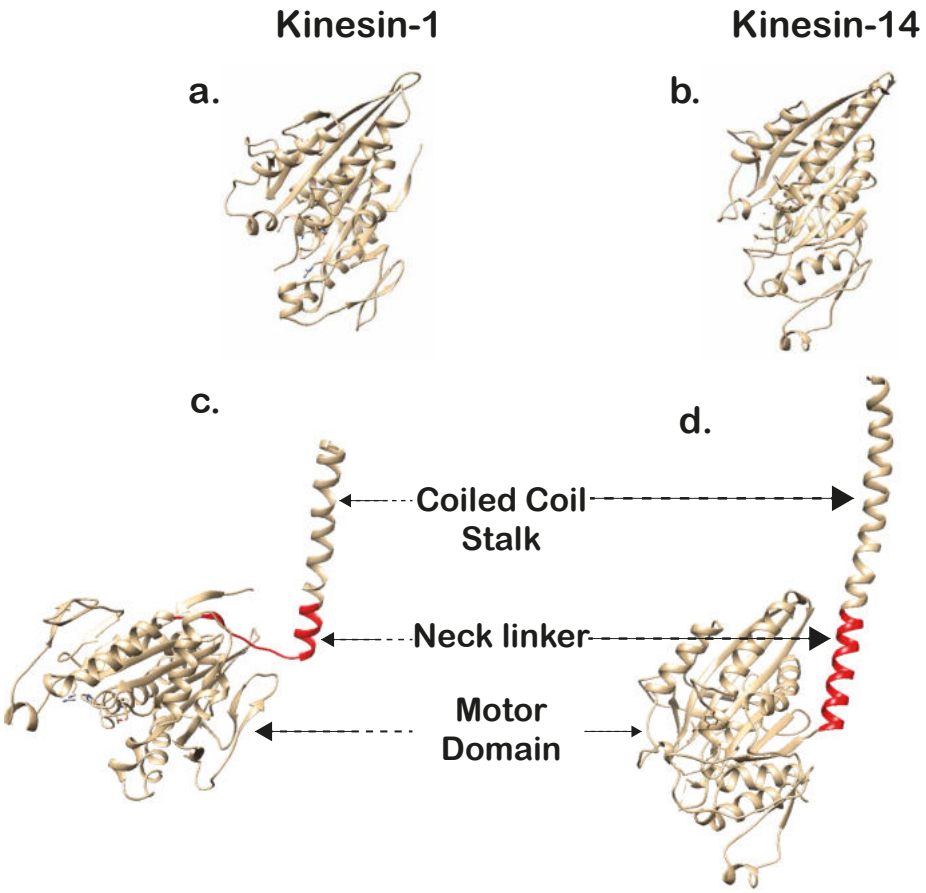


Figure 4.2 Kinesin-1 v/s. Kinesin-14

Atomic models obtained from crystal structures of monomeric kinesin-1 (PDB ID: 3KIN) and kinesin-14 (PDB ID: 2NCD) bound to ADP demonstrate the (a,b) similarity between their motor domains and the (c,d) differences arising due to the incorporation of the neck linker (red) in the construct.

Thus, although kinesin families are defined based on the position and homology of their motor domains, the neck linker also reflects the defining characteristics of

each family. This is further reinforced by the fact that any attempts made to decipher the basis for the difference in directionality using the motor domain-MT interactions alone were unsuccessful as the binding modes of the motor domains from both families appeared identical (Lockhart and Cross 1994). Single molecule motility studies of chimeric motors that fused the kinesin-1 motor to the kinesin-14 neck and vice versa provided irrefutable evidence that the neck linker controlled the directionality of a kinesin's motion (Yamagishi, Shigematsu et al. 2016). A subsequent structural analysis of the MT-bound chimeric motors illustrated that the neck orients itself to face the direction of motion of the kinesin, thereby positioning the unattached motor domain to face the opposite end (Yamagishi, Shigematsu et al. 2016).

With the recent negative stain reconstruction of kinesin-1 and the HSET structures that I determined, full-length structures of both the kinesin-14 and kinesin-1 families are now available, which provide a more complete picture of the conformation of the neck and stalk regions with respect to the motor domains. While the neck is not completely visible at the resolution of the negative stain kinesin-1 map (Tan, Yue et al. 2023), the maps of HSET obtained in this study are of a higher resolution and highlight features in the neck and stalk that have not been previously observed. For instance, the conformational state of Apo-HSET referred to as the "type K" shows a kink of nearly 90° in the neck (fig 4.1 a and b) leading into a straight stalk, while the "type S" states of both Apo and ADP-HSET shows no kink in the neck region and a visibly curved stalk (fig 4.1 c and d). Moreover, the kinked neck in the type K state appears to terminate at a single point to which both motor domains connect, while the neck in the type S state terminates at one motor domain (fig 4.1 c and d). Despite the differences in the C-terminal end of the neck region, the N-terminal end leads into a stalk that is offset towards one of the motor domains in the same direction in both conformers. The coordination between the curvature of the stalk, the kink in the neck, and the motor-neck connection suggests that these two conformers of Apo-HSET might represent intermediates corresponding to different stages in the movement of HSET on MTs. Further exploration of the significance of these states will ideally require structures of HSET bound to MTs at a comparable or higher resolution.

4.5.3 FL HSET: The motor domain

The asymmetry in kinesin-14 member Ncd was first observed as a kinetic discrepancy between the rate of ADP-release by the motor domains, with one showing a weaker affinity to ADP and hence, a 700-fold quicker rate of release than the other (Foster, Mackey et al. 2001). The motor with the weaker affinity was shown to be crucial to establishing contact with the MTs (Chen, Porche et al. 2012). While the kinetics of the asymmetry has been investigated in detail, how this would manifest in the structure of a homodimeric kinesin-14 still remains unclear.

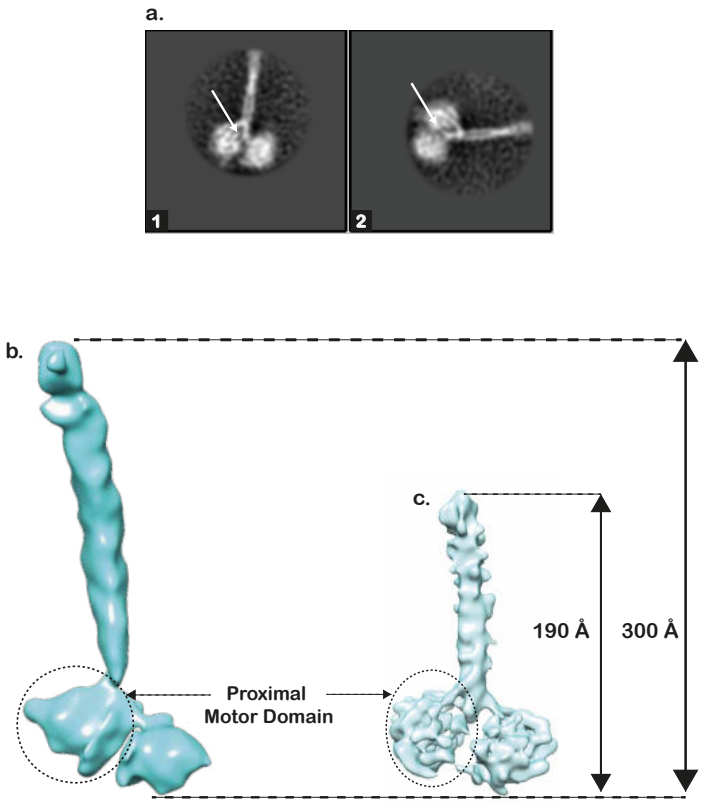


Figure 4.3 Asymmetry in the Motor Domains of ADP-HSET

(a) Reference-free 2D classes of the cryo-dataset of ADP-HSET obtained at a resolution of 8.5 Å show the second connection between the stalk and one of the motor domains (arrows); 3D maps of ADP-HSET obtained from (b) negative stain EM and (c) Cryo EM show the asymmetry that results from one of the motor domains (circled) being more proximal to the stalk than the other.

The first glimpse of structural asymmetry between the motor domains of homodimeric HSET can be seen in the dummy atom 3D models obtained by SAXS (fig 3.17 b and c, *Page 83*). At a maximal resolution of 30 Å, this property manifested as an apparent one-sided protrusion at the bottom of the stalk of Apo-HSET. However, as discussed in section 4.2.2 (*Page 166*), the exact nature and implications of this protrusion could not be deciphered at the resolution of the models. While the higher resolution information from negative stain and cryo EM confirmed the asymmetry, they also showed that it was not limited to Apo-HSET. In the negative stain reconstructions of ADP-FL HSET in fig 4.3 b, the greater proximity of one of the motor domains to the stalk than the other is a clear sign of a structural asymmetry within the homodimer, which may be related to the functional asymmetry observed in the kinetic studies.

This feature is also observed in the cryo EM structure of ADP-HSET (fig 4.3 c). At the higher resolution of the cryo EM map, the asymmetry further appears in the form of a consistently visible feature in FL HSET from the 2D classes (fig 4.3 a) to the final refined map in fig 4.3 c- the apparently clearer recovery of density in the proximal motor domain (circled, fig 4.3 a, b, and c) compared to the distal motor. As pointed out in section 3.6.2.2 (*Page 150*), this could be a sign of conformational variability in the distal motor domain and a comparative rigidity in the proximal one. The constraints that would be imposed by the apparent second attachment between the densities of the proximal motor and the stalk in the 2D classes (fig 4.3 a) and the 3D map (fig 4.3 c) further supports this motor hypothesis. The asymmetry is further reinforced by the difference in the fit of both subunits of dimeric HSET at once (fig 4.4 a) versus individually (fig 4.4 b) to the density corresponding to the motor domains. The latter case displaces the subunits compared to the former to fit 78% of the atoms in the proximal motor and 81% of the distal motor within the contour, compared to the former that incorporates 62% of the atoms of both motors into the density. The vertical displacement of the individually fitted subunits positions the nucleotide-binding site of the proximal motor domain in fig 4.4 d closer to the stalk than the distal motor domain.

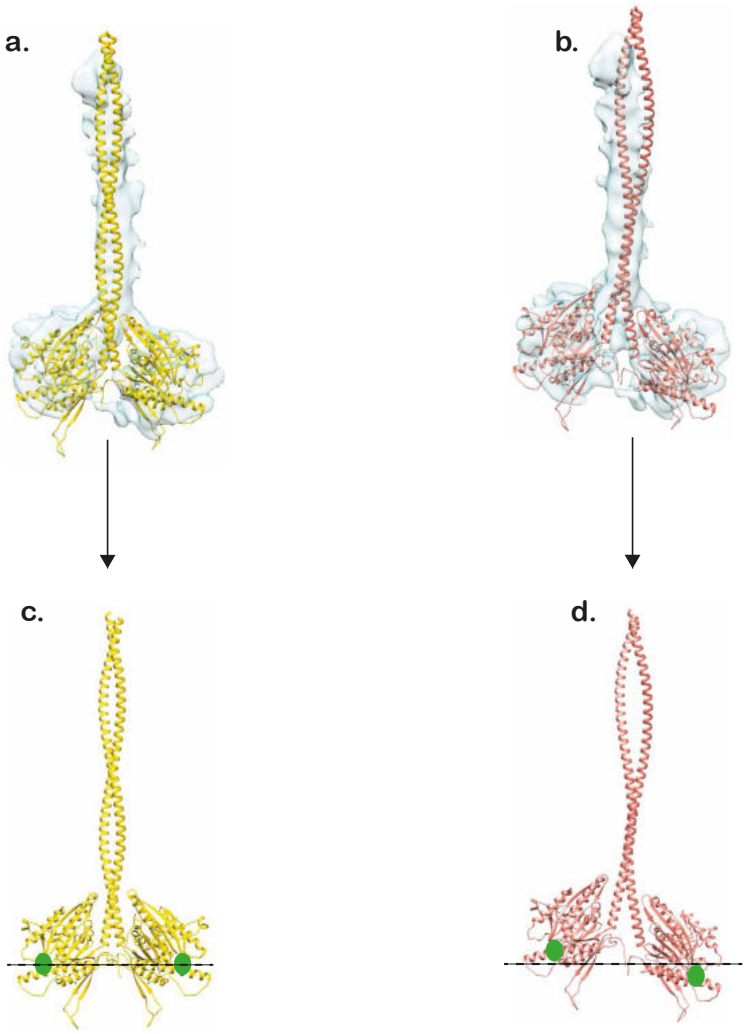


Figure 4.4 Nucleotide-binding Sites of the Motor Domains in ADP-HSET

Asymmetry in the motor domains of ADP-HSET is illustrated by the change in positions of the motor domains between the (a) dimeric fit and (b) individual monomeric fit into the 8.5 Å cryo EM map; the nucleotide-binding sites (green) in the (c) symmetric model are at the same horizontal level, while those in the (d) asymmetric model show a vertical displacement relative to each other.

While it is not possible to obtain a more detailed view at this resolution and with rigid body fits of the atomic models, the above results suggest that the motor domain distal to the stalk is most likely to establish first contact with a MT. As MT-binding is correlated with ADP release, I propose that in this HSET conformation, this motor is probably the one with the weaker ADP affinity. In addition to the differences in the position and the connections between the stalk and each motor domain, the dissimilarities in the features visible in the cryo EM map for each

motor domain highlighted in section 3.6.2.3 (*Page 154*) hint at conformational differences between the two, which might be related to different ADP affinities. Overall, the asymmetry observed in a kinesin-14 like HSET appears to be a conformational difference between the motor domains. While structural asymmetry has been observed before in the kinesin superfamily as early as 1996 in the crystal structure of kinesin-1 (Kull, Sablin et al. 1996), the type of asymmetry described above has not been reported earlier.

4.5.4 FL HSET: The CBD

The N-terminal CBD of the kinesin-14 family, which is known to interact with MTs, is poorly characterised. Its understanding is limited to the fact that it contains a basic motif suggested to be involved in electrostatic interactions with the MT (Furuta and Toyoshima 2008). The HSET CBD is predicted to be unstructured (Supplementary fig S3, *Page 204*). In this study, I established through the shape of the Kratky analysis in SAXS that while Apo and ADP-bound FL HSET do possess a similar degree of inter-domain mobility, there was no evidence in the plot of any part of the dimer being unstructured (as discussed in chapter 4.1, *Page 162*). In my negative stain EM analysis of the two forms of HSET, the 2D classes obtained by the reference-based classification revealed density beyond the length of 250 Å predicted by Alphafold for the structured part of HSET i.e., from the bottom of the motor domain to the top of the coiled coil, indicating the recovery of the N-terminal CBD (fig 4.5). The recovery of the CBD resulted in a total length of 300 Å for HSET, which is 65-75 Å shorter than the estimate obtained by SAXS (see section 3.2.6.4, *Page 81*). This discrepancy could be a combined consequence of a possible overestimation of the SAXS measurement or the flexibility of the terminal residues that renders the full extent of the CBD invisible in EM maps. However, the recovery of a portion of the CBD in the reference-based 2D classes such as the one shown in fig 4.5 suggests that it may have some structured regions. This was further reinforced by a similar CBD density in the refined negative stain EM maps for ADP-HSET (fig 4.3) and both conformers of Apo-HSET (fig 4.1).

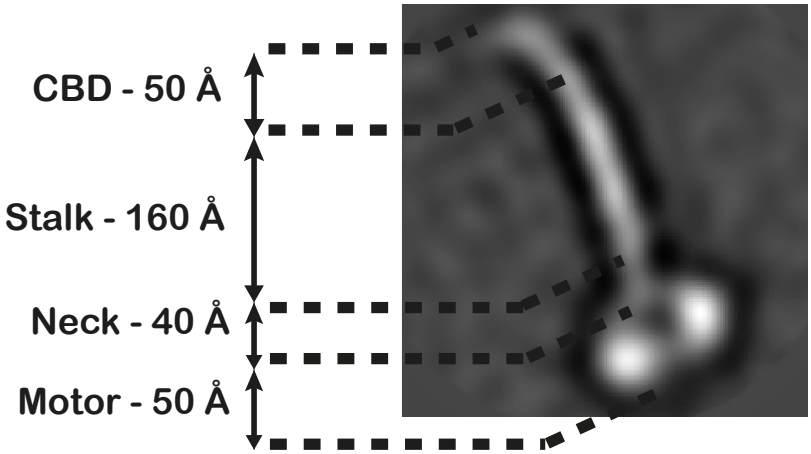


Figure 4.5 Lengths of Domains in FL HSET

A representative 2D class average of ADP-HSET obtained from a negative stain EM dataset by reference-based approach shows the recovery of information pertaining to the entire domain architecture of HSET, including the N-terminal CBD.

Additionally, as observed in the negative stain EM maps of the Apo-HSET conformers in fig 4.1, the presence of this feature may hint at inter-domain crosstalk. The previous section described a coordinated presence of three features in the type K and type S conformers- the point of origin of the motor domains from the neck, a kink in the neck, and the curvature of the stalk. In other words, the kinesin appears to display two independent sets of motion- the hopping of the motor domains at the C-terminal end, and a bending motion brought about by the alternation between the curved and straight stalk-CBD regions at the N-terminal end. The correlation between the conformation of the stalk and CBD domains with the different positions of the motor domains shows that the two recovered forms of the stalk-CBD region are not random instances but may represent discrete conformations that play a role in the function of HSET.

4.5.5 Summary

Thus, the full-length structures provided a first glimpse of inter-domain communication in the Apo and ADP-bound forms of HSET. As mentioned earlier, investigation of HSET in the presence of MTs would be required to test these hypotheses and further build the complete picture of its mechanism of action.

The 3D structures of Apo and ADP-FL HSET obtained by negative stain EM at a resolution of 20 Å provided the first visual evidence of a full-length kinesin-14, encompassing the motor, neck linker, stalk, and cargo-binding tail domains. The 3D analysis uncovered two conformations for Apo-HSET, while a single conformation of ADP-HSET was observed. The type K (Kinked) conformer consists of a distinct right-angled kink in the neck and a relatively straight stalk-CBD region, while the type S (Straight) conformer comprises a straight neck and a visibly curved stalk-CBD region. In addition, the two conformers differ in the orientation of the motor domains relative to the neck. The ADP-bound form resembled the type S state of Apo-HSET. Overall, the presence of distinct features in the domains of Apo and ADP-FL HSET provided a glimpse into the communication across the three domains that may be linked to the mechanism of kinesin movement along MTs.

Motivated by the success of the negative stain analysis, I prepared a sample of ADP-HSET for cryo EM data collection with the aid of the VPP to enhance the contrast and render the HSET particles visible in the cryo-electron micrographs. The data yielded 2D classes with a high level of detail and a final 3D refined map at an estimated resolution of 8.5 Å, showing the motor domains, the neck linker a little more than one cross-over repeat of the coiled coil stalk, encompassing a total length of 190 Å. Globally, the structure appeared similar to the negative stain map of ADP-HSET, with the addition of higher resolution details. Most obvious were the resolved strands of the coiled coil in the 2D classes, showing a helical pitch of 100 Å and a maximal inter-strand spacing of 10 Å, and the detail in the neck separating the motor domains from the stalk. While this was also visible in the ab-initio 3D map at a resolution of 9 Å, it was obscured by noise in the refined map and by anisotropy from a lack of side

views. Another detail was the asymmetry of the motor domains with respect to the stalk, as initially observed in the negative stain maps, and at a higher resolution in the cryo EM maps. The 3D map further revealed that the proximal motor contained more than one attachment to the coiled coil as opposed to the single connection through the neck linker in the distal motor, which could possibly help to stabilise the conformation of the proximal motor domain. Overall, the observed asymmetry is consistent with a difference in affinity of each motor domain for ADP and initial MT binding through one of the motor domains.

5. THE NEXT STEPS IN THE STRUCTURAL INVESTIGATION OF HSET

5.1 Short-Term Aims

The 3D maps of Apo and ADP-FL HSET obtained from negative stain EM as well as the recovery of secondary structure detail in the cryo EM map at a resolution of around 8.5 Å showed that FL HSET is indeed a viable target for single particle structure elucidation. While the long-term aim of the structural studies on HSET is to understand its interaction with MTs, the immediate next steps of the project would be set to maximize the resolution of the isolated HSET structures by overcoming the limitations faced in the EM data collection and processing during my PhD.

5.1.1 Preferred Orientation

The domination of the 2D classes of the cryo EM analysis by the front facing views of ADP-HSET led to the observed anisotropy of the final refined map. This is most likely caused by the tendency of HSET to adopt a preferred orientation of front views on the grid. As shown, the loss of information from the side views significantly compromises the resolution of the map in the corresponding directions and results in a non-uniform distribution of density in the map. As discussed in chapter 4.3 (*Page 173*), I would overcome these preferred orientations by collecting more data from more grids of ADP-HSET to accumulate a sufficient number of side views that would yield an isotropic structure. However, the issue was not observed in the negative stain data, which was also collected on carbon supported grids. Thus, the elimination of biased orientation in the cryo grids could also be achieved from grid-to-grid variability. As an alternative strategy, I would attempt to prevent preferred orientations by glow discharging supported grids in the presence of pentylamine prior to sample application, as mentioned in chapter 4.3 (*Page 173*).

5.1.2 Apo-HSET

The pursuit of a cryo EM structure Apo-HSET was abandoned midway through the Glacios sessions, on account of the problems with the VPP and the limited availability of microscope time, which led to the prioritization of ADP-HSET. However, as shown in negative stain EM, Apo-HSET is an equally feasible cryo EM target that warrants a set of VPP data collections on the Titan Krios at LONCEM. Considering Apo-HSET showed two conformers, it is important to bear in mind that resolving each of these states to sub-10 Å resolutions would warrant sufficiently greater quantities of data than ADP-HSET. Obtaining high resolution structures of Apo-HSET would provide interesting insight into the conformational difference in the motor domains and the neck linkers between the two conformers as well as between Apo and ADP-HSET overall. As kinesin-14 members have been shown to bind to the MT through a nucleotide-free motor domain (Zhang, Dai et al. 2015), a comparison between the motor domains of the ADP and Apo forms of HSET should provide clues into the conformational changes that are undergone by HSET to transition from the isolated to MT-bound forms. Moreover, due to the significant difference between the motor domains in terms of their affinity to ADP in a homodimeric kinesin-14 (Foster, Mackey et al. 2001), the isolation of a kinesin-14 with both motor domains in the Apo state i.e., Apo-Apo form has not been reported previously. While the Apo form of HSET reported in this study does bind ADP as indicated by DSF, there is no way to ascertain the stoichiometry of ADP-binding to the motor domains of HSET. In light of this, a high-resolution structure of Apo-HSET would be required to determine whether I have purified an Apo-Apo or Apo-ADP form of FL HSET.

5.1.3 Data Processing Strategies

While an optimised sample preparation strategy is a pre-requisite, it does not guarantee a high-resolution structure (Orlova and Saibil 2011). As shown in sections 3.4.2 (*Page 101*) and 3.6.2 (*Page 146*), even the best datasets required extensive customisation of the processing pipeline. Optimised computational

approaches also play a significant role in breaking the resolution barrier for a dataset. For instance, a computational approach to tackle the lack of HSET side views could be to isolate the particles belonging to the few 2D classes of the side views and using them to train a CRYOLO model that is specific to side views. I would also attempt segregating the CRYOLO models for picking front and side views, to prevent loss of information due to non-specificity of the picking model. Any overlaps between the picked coordinates can be effectively detected and deleted in RELION's program to append datasets.

For the 3D analysis, both Apo and ADP-HSET datasets suffer from the lack of resolution that arises during the accommodation of greater lengths of the stalk, as described in fig 3.57. While limiting the length of the molecule by the application of a circular mask in 2D and a short translational search range in 3D proved to be a successful approach to recover high resolution details, this was achieved by compromising the rest of the stalk and N-terminal CBD. In order to visualize the N-terminus of HSET, I propose two strategies to be applied to a homogeneous stack of particles that can yield a high-resolution structure through the approaches employed in section 3.6.2 (*Page 146*). The first involves repeating the optimised classification outlined in section 3.6.2.2 (*Page 150*) with steadily increasing diameters of the circular mask, to capture the information corresponding to the stalk-CBD region. The second involves recentering the aligned particles to a point that is visible well within the stalk i.e., closer to the N-terminal end, and classifying them without 3D alignment i.e., a reference-based 3D classification. This would average the stalk and CBD regions without including them in the calculations of an optimal alignment of the entire particle, thereby preventing the loss of resolution resulting from any inaccuracy in aligning this region. Depending on the strength of the signal and the flexibility at this distal end of the stalk, the class-averaging would either recover discrete conformations of this region or sample its continuous movement.

5.2 Long-Term Aims

While the ideas described above describe further analysis of current data or further use of currently available constructs of FL HSET, this section focusses on the creation of alternate systems for structural exploration of HSET.

5.2.1 Shorter constructs

One of the issues raised in the earlier section was the challenge that arose from the inclusion of the N-terminal CBD. In addition to the problems faced in processing the data, an important consequence of an elongated system was the increased frequency of overlap or cross-over between neighbouring particles, thereby imposing a limit on the length of the molecule that could be enclosed in a box during particle extraction. In other words, the full length of HSET hindered the single particle analysis of full length HSET. A typical solution to this would be to reduce the concentration of HSET on the grid, effectively spreading the particles further and minimising the occurrence of overlaps. However, the consequent reduction in the number of particles obtained per micrograph would limit the resolution. An alternative could be to produce shorter constructs of HSET which include the coiled coil up to the length of 100 Å that is resolved in the current analysis. The shorter length might achieve a higher optimal particle density on the grids before overlapping becomes an issue. As a result, the dataset could potentially attain a higher resolution for the length of the construct being used rather than compromise a part of it.

5.2.2 Microtubule binding

Studying the interaction of HSET with MTs is a crucial last piece of the puzzle to understand the mechanism of action of HSET. While my study so far dealt with HSET as the sole target, the incorporation of MTs would result in a two-target system, with the added predicament of a 1000-fold size difference between HSET and MT, and multiple HSET molecules bound to a single MT (McDonald, Stewart et al. 1990). Thus, the visualisation of HSET-MT interactions using single particle

EM would require the consideration of a different optimum at every step of the study, from sample preparation to data processing. Fortunately, kinesin-MT complexes have been explored by cryo EM since the 1980s (Amos 1987) and have also been analysed by single particle EM methods since the 1990s (Lockhart and Cross 1994, Hirose, Lockhart et al. 1995, Endres, Yoshioka et al. 2006, Chen, Porche et al. 2012). Furthermore, the resolution revolution in cryo EM has yielded advances in processing strategies, leading to routinely solved high resolution structures of MTs that are decorated with proteins (Moores 2008, Cook, Roberts et al. 2021). Thus, there is significant precedence to the cryo EM of kinesin-MT complexes.

However, these studies were conducted using constructs restricted to the motor domain and the neck linker with a small portion of the stalk to allow dimerization. Hence, the first strategy I propose would involve the generation of such truncated constructs of HSET motor domains to decorate MTs in the presence of different nucleotides such as ADP and non-hydrolysable ATP analogs. This would enable a uniform decoration of a sufficient number of HSET particles on the MTs required to obtain high resolution structural snapshots of the motor domain at different stages in its mechanical cycle. The second strategy concerns the fact that HSET binds MTs using the N-terminal end as well. However, this region is too short to be visualised on its own. On the other hand, the use of the full-length structure poses the problem of non-selectivity of binding modes, in which MTs would be non-uniformly decorated with the motor domains of some HSET molecules, the tails of some others, and both. To overcome this, I propose the generation of a full-length HSET construct with a mutation of the highly conserved residue R616 in the microtubule-binding L12-α5 loop-helix region of the motor domain. In kinesin-1, mutating this residue to alanine has been shown to cause a fifteen-fold reduction in the affinity of kinesin to MT (Woehlke, Ruby et al. 1997). Therefore, an HSET construct with this impaired motor-MT interaction would be more likely to participate in MT binding at the N-terminal CBD, thereby enriching the HSET CBD-MT contacts.

5.2.3 Compound binding

The last aim of this project arises from the potential of HSET as an anti-cancer drug target, which was the motivation for this entire study. Recent drug discovery efforts have unveiled allosteric inhibitors (Wu, Mikule et al. 2013) as well as ATP-competitors that arrest HSET on MTs (Saint-Dizier, Matthews et al. 2023). However, the lack of structural data to visualise the binding of the inhibitors limits the drug design pipeline. Thus, the final step of the HSET project would involve the preparation of HSET-MT complexes in the presence of inhibitors. Obtaining a structure of HSET bound to an inhibitor and MT would provide valuable insight into the molecular requirements for HSET inhibition. A comparison of this intermediate to the nucleotide-bound states of HSET would further aid the understanding of the mode of action of such an HSET inhibitor. Furthermore, a sufficiently high resolution could lead to the visualisation of the inhibitor molecule in the density of the map, allowing the analysis of interactions at an atomic level to guide structure-based drug design. Thus, in this manner, cryo EM could be used to drive drug discovery efforts on a protein like HSET for which no protein-inhibitor crystal structures could be obtained.

SUPPLEMENTARY INFORMATION

S1. Constructs for FL-HSET

S1A. His_6-TEV-FLHSET

MGSAHHHHHHASNLYFQGMDPQRSPLLEVKGNIELKRPLIKAPSQLPLSGSRL
KRRPDQMEDGLEPEKKRTRGLGATTKITTSHPRVPSLTTVPQTQGQTTAQKV
SKKTGPRCSTAIATGLKNQKPVPAVPVQKSGTSGVPPMAGGKKPSKRPAWD
LKGQLCDLNAELKRCRERTQTLDQENQQLQDQLRDAQQQVKALGTERTTLE
GHLAKVQAQAEQGQQELKNLRACVLELEERLSTQEGLVQELQKKQVELQEER
RGLMSQLEEKERRLQTSEAALSSSQAEVASLRQETVAQAALLTEREERLHGL
EMERRRLHNQLQELKGNIRVFCRVRPVLPGEPTPPPGLLLFPSGPGGPSDPP
TRLSLSRSDERRGTLSGAPAPPTRHDFSFDRVFPPGSGQDEVFEEIAMLVQS
ALDGYPVCIFAYGQTGSGKTFTMEGGPGGDPQLEGLIPRALRHLFSVAQELS
GQGWTYSFVASYVEIYNETVRDLLATGTRKGQGGECEIRRAGPGSEELTVTN
ARYVPVSCEKEVDALLHLARQNRAVARTAQNERSSRSHSVFQLQISGEHSSR
GLQCGAPLSLVDLAGSERLDPGLALGPGERERLRETQAINSSLSTLGLVIMALS
NKESHVPYRNSKLTYLLQNSLGGSAKMLMFVNISPLEENVSESLNSLRFASKV
NQCVIGTAQANRK

S1B. $Strep_2$-HRV3C-FLHSET

MGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKASLEVLFQGPHMDPQRS
PLLEVKGNIELKRPLIKAPSQLPLSGSRLKRRPDQMEDGLEPEKKRTRGLGAT
TKITTSHPRVPSLTTVPQTQGQTTAQKVSKKTGPRCSTAIATGLKNQKPVPAV
PVQKSGTSGVPPMAGGKKPSKRPAWDLKGQLCDLNAELKRCRERTQTLDQE
NQQLQDQLRDAQQQVKALGTERTTLEGHLAKVQAQAEQGQQELKNLRACVL
ELEERLSTQEGLVQELQKKQVELQEERRGLMSQLEEKERRLQTSEAALSSSQ
AEVASLRQETVAQAALLTEREERLHGLEMERRRLHNQLQELKGNIRVFCRVR
PVLPGEPTPPPGLLLFPSGPGGPSDPPTRLSLSRSDERRGTLSGAPAPPTRH
DFSFDRVFPPGSGQDEVFEEIAMLVQSALDGYPVCIFAYGQTGSGKTFTMEG
GPGGDPQLEGLIPRALRHLFSVAQELSGQGWTYSFVASYVEIYNETVRDLLAT

GTRKGQGGECEIRRAGPGSEELTVTNARYVPVSCEKEVDALLHLARQNRAVA
RTAQNERSSRSHSVFQLQISGEHSSRGLQCGAPLSLVDLAGSERLDPGLALG
PGERERLRETQAINSSLSTLGLVIMALSNKESHVPYRNSKLTYLLQNSLGGSAK
MLMFVNISPLEENVSESLNSLRFASKVNQCVIGTAQANRK

S1C. Primers for PCR of Strep$_2$-FLHSET

Forward- 5’ TTCCAGGGTCCCCATATGGATCCGCAGAGGTCCCCCC 3’

Reverse- 5’ TAGGCCTTTGAATTCTCACTTCCTGTTGGCCTGAGCAGTAC 3’

S2. Purification of FLHSET

S2A. Molecular Weight Determination from SEC

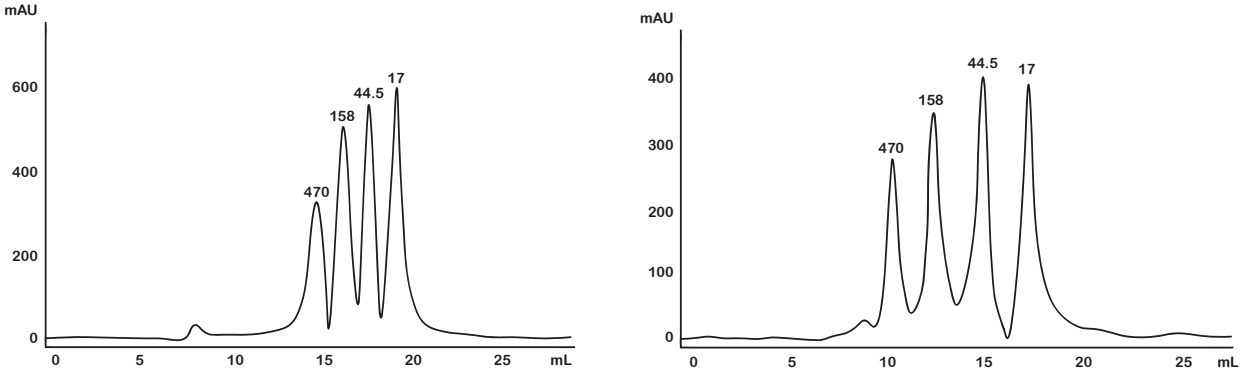


Figure S2: Calibration of Size Exclusion Columns

Superdex 200 10/30 GL and Superose 6 Increase 10/30 GL SEC columns were calibrated using Bio-rad's gel filtration standard mixture of proteins with known MW; MW of each eluted protein is indicated above the elution peak corresponding to the protein.

S2B. Biochemical Properties of FL HSET

Biochemical properties of FL HSET were estimated from its amino acid sequence using the ProtParam tool in the ExPasy suite (Gasteiger, E. et al., 2011).

Number of amino acids: 716

Molecular weight: 78142.37

Theoretical pI: 9.1

Atomic composition:

Carbon	C	3376
Hydrogen	H	5510
Nitrogen	N	1024
Oxygen	O	1064
Sulfur	S	21

Formula: $C_{3376}H_{5510}N_{1024}O_{1064}S_{21}$
Total number of atoms: 10995

Amino acid composition:

Ala (A)	50	7.0%
Arg (R)	56	7.8%
Asn (N)	20	2.8%
Asp (D)	19	2.7%
Cys (C)	10	1.4%
Gln (Q)	55	7.7%
Glu (E)	61	8.5%
Gly (G)	67	9.4%
His (H)	13	1.8%
Ile (I)	15	2.1%
Leu (L)	84	11.7%
Lys (K)	35	4.9%
Met (M)	11	1.5%
Phe (F)	16	2.2%
Pro (P)	48	6.7%
Ser (S)	63	8.8%
Thr (T)	41	5.7%
Trp (W)	4	0.6%
Tyr (Y)	8	1.1%
Val (V)	40	5.6%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 80
Total number of positively charged residues (Arg + Lys): 91

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 34545
Abs 0.1% (=1 g/l) 0.442, assuming all pairs of Cys residues form cystines

Ext. coefficient 33920
Abs 0.1% (=1 g/l) 0.434, assuming all Cys residues are reduced

S2C. Concentration and Purity of Eluted HSET Fraction

HSET Variant	Eluted Fraction of Apo-HSET	Concentration (mg/mL)
His_6-FLHSET	Peak - 1	0.343
	Peak	0.441
	Peak + 1	0.286
$Strep_2$-FLHSET-I	Peak - 1	1.522
	Peak	1.708
	Peak + 1	1.439
$Strep_2$-FLHSET-II	Peak - 1	0.462
	Peak	0.497
	Peak + 1	0.487

Table S2: Concentrations of peak Apo-HSET fractions

The concentrations of the HSET fractions eluted from the final step of each purification strategy were measured using the NanoDrop UV-Vis Spectrophotometer; the data is shown for the peak fraction from each purification strategy as well as two fractions adjacent to the peak, upstream (peak -1) and downstream (peak + 1) of it.

S3. Secondary Structure Prediction of FLHSET

The secondary structure elements in FL HSET were predicted from its amino acid sequence using the PSIPRED online tool (McGuffin, L.J. et al., 2000)

Figure S3 (Page 1): Predicting the Secondary Structure of FL HSET

PSIPRED chart predict the regions in the amino acid sequence of FL HSET that are α-helices (H, pink), ß-strands (E, yellow) or random coils (C, uncoloured).

Figure S3 (Page 2): Predicting the Secondary Structure of FL HSET

PSIPRED chart predict the regions in the amino acid sequence of FL HSET that are α-helices (H, pink), ß-strands (E, yellow) or random coils (C, uncoloured).

6. BIBLIOGRAPHY

1. Adhikari, S., P. V. Manthena, K. Sajwan, K. K. Kota and R. Roy (2010). "A unified method for purification of basic proteins." Anal Biochem **400**(2): 203-206.
2. Alexander, C. G., R. Wanner, C. M. Johnson, D. Breitsprecher, G. Winter, S. Duhr, P. Baaske and N. Ferguson (2014). "Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings." Biochim Biophys Acta **1844**(12): 2241-2250.
3. Allen, R. D., J. Metuzals, I. Tasaki, S. T. Brady and S. P. Gilbert (1982). "Fast axonal transport in squid giant axon." Science **218**(4577): 1127-1129.
4. Amos, L. A. (1987). "Kinesin from pig brain studied by electron microscopy." J Cell Sci **87 (Pt 1)**: 105-111.
5. Ando, A., Y. Y. Kikuti, H. Kawata, N. Okamoto, T. Imai, T. Eki, K. Yokoyama, E. Soeda, T. Ikemura, K. Abe and et al. (1994). "Cloning of a new kinesin-related gene located at the centromeric end of the human MHC region." Immunogenetics **39**(3): 194-200.
6. Basto, R., K. Brunk, T. Vinadogrova, N. Peel, A. Franz, A. Khodjakov and J. W. Raff (2008). "Centrosome amplification can initiate tumorigenesis in flies." Cell **133**(6): 1032-1042.
7. Benoit, M., A. B. Asenjo and H. Sosa (2018). "Cryo-EM reveals the structural basis of microtubule depolymerization by kinesin-13s." Nat Commun **9**(1): 1662.
8. Bepler, T., A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble and B. Berger (2019). "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs." Nat Methods **16**(11): 1153-1160.
9. Bhella, D. (2019). "Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility." Biophys Rev **11**(4): 515-519.
10. Bisby, M. A. (1976). "Axonal transport." Gen Pharmacol **7**(6): 387-393.
11. Bornhorst, J. A. and J. J. Falke (2000). "Purification of proteins using polyhistidine affinity tags." Methods Enzymol **326**: 245-254.
12. Brady, S. T., R. J. Lasek and R. D. Allen (1982). "Fast axonal transport in extruded axoplasm from squid giant axon." Science **218**(4577): 1129-1131.
13. Brinkley, B. R. (2001). "Managing the centrosome numbers game: from chaos to stability in cancer cell division." Trends Cell Biol **11**(1): 18-21.
14. Brunet, S. and I. Vernos (2001). "Chromosome motors on the move. From motion to spindle checkpoint activity." EMBO Rep **2**(8): 669-673.
15. Chen, C. J., K. Porche, I. Rayment and S. P. Gilbert (2012). "The ATPase pathway that drives the kinesin-14 Kar3Vik1 powerstroke." J Biol Chem **287**(44): 36673-36682.
16. Chen, J. Z., C. Sachse, C. Xu, T. Mielke, C. M. Spahn and N. Grigorieff (2008). "A dose-rate effect in single-particle electron microscopy." J Struct Biol **161**(1): 92-100.

17. Cole, D. G., S. W. Chinn, K. P. Wedaman, K. Hall, T. Vuong and J. M. Scholey (1993). "Novel heterotrimeric kinesin-related protein purified from sea urchin eggs." Nature **366**(6452): 268-270.
18. Collins, C. A. and R. B. Vallee (1986). "A microtubule-activated ATPase from sea urchin eggs, distinct from cytoplasmic dynein and kinesin." Proc Natl Acad Sci U S A **83**(13): 4799-4803.
19. Cook, A. D., A. J. Roberts, J. Atherton, R. Tewari, M. Topf and C. A. Moores (2021). "Cryo-EM structure of a microtubule-bound parasite kinesin motor and implications for its mechanism and inhibition." J Biol Chem **297**(5): 101063.
20. D'Assoro, A. B., W. L. Lingle and J. L. Salisbury (2002). "Centrosome amplification and the development of cancer." Oncogene **21**(40): 6146-6153.
21. D'Imprima, E., D. Floris, M. Joppe, R. Sanchez, M. Grininger and W. Kuhlbrandt (2019). "Protein denaturation at the air-water interface and how to prevent it." Elife **8**.
22. Danev, R. and W. Baumeister (2016). "Cryo-EM single particle analysis with the Volta phase plate." Elife **5**.
23. Danev, R., B. Buijsse, M. Khoshouei, J. M. Plitzko and W. Baumeister (2014). "Volta potential phase plate for in-focus phase contrast transmission electron microscopy." Proc Natl Acad Sci U S A **111**(44): 15635-15640.
24. Danev, R., D. Tegunov and W. Baumeister (2017). "Using the Volta phase plate with defocus for cryo-EM single particle analysis." Elife **6**.
25. deCastro, M. J., R. M. Fondecave, L. A. Clarke, C. F. Schmidt and R. J. Stewart (2000). "Working strokes by single molecules of the kinesin-related microtubule motor ncd." Nat Cell Biol **2**(10): 724-729.
26. Dinenberg, A. S., J. R. McIntosh and J. M. Scholey (1986). "Studies on sea urchin egg cytoplasmic ATPases of possible significance for microtubule functions." Ann N Y Acad Sci **466**: 431-435.
27. Droz, B., H. L. Koenig, L. D. Biamberardino and L. Di Giamberardino (1973). "Axonal migration of protein and glycoprotein to nerve endings. I. Radioautographic analysis of the renewal of protein in nerve endings of chicken ciliary ganglion after intracerebral injection of (3H)lysine." Brain Res **60**(1): 93-127.
28. Drulyte, I., R. M. Johnson, E. L. Hesketh, D. L. Hurdiss, C. A. Scarff, S. A. Porav, N. A. Ranson, S. P. Muench and R. F. Thompson (2018). "Approaches to altering particle distributions in cryo-electron microscopy sample preparation." Acta Crystallogr D Struct Biol **74**(Pt 6): 560-571.
29. Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall and P. Schultz (1988). "Cryo-electron microscopy of vitrified specimens." Q Rev Biophys **21**(2): 129-228.
30. Dubochet, J., M. Ducommun, M. Zollinger and E. Kellenberger (1971). "A new preparation method for dark-field electron microscopy of biomacromolecules." J Ultrastruct Res **35**(1): 147-167.
31. Durand, D., C. Vives, D. Cannella, J. Perez, E. Pebay-Peyroula, P. Vachette and F. Fieschi (2010). "NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers." J Struct Biol **169**(1): 45-53.

32. Endow, S. A., S. Henikoff and L. Soler-Niedziela (1990). "Mediation of meiotic and early mitotic chromosome segregation in Drosophila by a protein related to kinesin." Nature **345**(6270): 81-83.
33. Endow, S. A. and H. Higuchi (2000). "A mutant of the motor protein kinesin that moves in both directions on microtubules." Nature **406**(6798): 913-916.
34. Endow, S. A., S. J. Kang, L. L. Satterwhite, M. D. Rose, V. P. Skeen and E. D. Salmon (1994). "Yeast Kar3 is a minus-end microtubule motor protein that destabilizes microtubules preferentially at the minus ends." EMBO J **13**(11): 2708-2713.
35. Endres, N. F., C. Yoshioka, R. A. Milligan and R. D. Vale (2006). "A lever-arm rotation drives motility of the minus-end-directed kinesin Ncd." Nature **439**(7078): 875-878.
36. Enos, A. P. and N. R. Morris (1990). "Mutation of a gene that encodes a kinesin-like protein blocks nuclear division in A. nidulans." Cell **60**(6): 1019-1027.
37. Erickson, H. P. (2009). "Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy." Biol Proced Online **11**: 32-51.
38. Erickson H. P.; Klug A. *Philos. Trans. R. Soc., B* 1970, 261, 105
39. Fan, H. and F. Sun (2022). "Developing Graphene Grids for Cryoelectron Microscopy." Front Mol Biosci **9**: 937253.
40. Fan, X., L. Zhao, C. Liu, J. C. Zhang, K. Fan, X. Yan, H. L. Peng, J. Lei and H. W. Wang (2017). "Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM." Structure **25**(10): 1623-1630 e1623.
41. Fink, G., L. Hajdo, K. J. Skowronek, C. Reuther, A. A. Kasprzak and S. Diez (2009). "The mitotic kinesin-14 Ncd drives directional microtubule-microtubule sliding." Nat Cell Biol **11**(6): 717-723.
42. Forman, D. S., K. J. Brown and D. R. Livengood (1983). "Fast axonal transport in permeabilized lobster giant axons is inhibited by vanadate." J Neurosci **3**(6): 1279-1288.
43. Foster, K. A., J. J. Correia and S. P. Gilbert (1998). "Equilibrium binding studies of non-claret disjunctional protein (Ncd) reveal cooperative interactions between the motor domains." J Biol Chem **273**(52): 35307-35318.
44. Foster, K. A. and S. P. Gilbert (2000). "Kinetic studies of dimeric Ncd: evidence that Ncd is not processive." Biochemistry **39**(7): 1784-1791.
45. Foster, K. A., A. T. Mackey and S. P. Gilbert (2001). "A mechanistic model for Ncd directionality." J Biol Chem **276**(22): 19259-19266.
46. Frank J. Oxford University Press; USA: 2006. Three-dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State.
47. Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." J Appl Crystallogr **42**(Pt 2): 342-346.
48. Fukasawa, K. (2005). "Centrosome amplification, chromosome instability and cancer development." Cancer Lett **230**(1): 6-19.

49. Furuta, K. and Y. Y. Toyoshima (2008). "Minus-end-directed motor Ncd exhibits processive movement that is enhanced by microtubule bundling in vitro." Curr Biol **18**(2): 152-157.
50. Gao, K., R. Oerlemans and M. R. Groves (2020). "Theory and applications of differential scanning fluorimetry in early-stage drug discovery." Biophys Rev **12**(1): 85-104.
51. Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res **31**(13): 3784-3788.
52. Ghadimi, B. M., D. L. Sackett, M. J. Difilippantonio, E. Schrock, T. Neumann, A. Jauho, G. Auer and T. Ried (2000). "Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations." Genes Chromosomes Cancer **27**(2): 183-190.
53. Ghisaidoobe, A. B. and S. J. Chung (2014). "Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Forster resonance energy transfer techniques." Int J Mol Sci **15**(12): 22518-22538.
54. Glaeser, R. M. and K. A. Taylor (1978). "Radiation damage relative to transmission electron microscopy of biological specimens at low temperature: a review." J Microsc **112**(1): 127-138.
55. Glatter O, Kratky O. *Small angle x-ray scattering.* Academic Press; London ; New York: 1982.
56. Goodson, H. V., S. J. Kang and S. A. Endow (1994). "Molecular phylogeny of the kinesin family of microtubule motor proteins." J Cell Sci **107 (Pt 7)**: 1875-1884.
57. Grafstein, B. and D. S. Forman (1980). "Intracellular transport in neurons." Physiol Rev **60**(4): 1167-1283.
58. Grant, T., A. Rohou and N. Grigorieff (2018). "cisTEM, user-friendly software for single-particle image processing." Elife **7**.
59. Guo, F. and W. Jiang (2014). "Single particle cryo-electron microscopy and 3-D reconstruction of viruses." Methods Mol Biol **1117**: 401-443.
60. Hackney, D. D. (1988). "Kinesin ATPase: rate-limiting ADP release." Proc Natl Acad Sci U S A **85**(17): 6314-6318.
61. Hammel, M. (2012). "Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS)." Eur Biophys J **41**(10): 789-799.
62. Hartwell, L. H. and T. A. Weinert (1989). "Checkpoints: controls that ensure the order of cell cycle events." Science **246**(4930): 629-634.
63. Hayat, M.A. Glutaraldehyde: Role in electron microscopy, Micron and Microscopica Acta, Volume 17, Issue 2, 1986, Pages 115-135, ISSN 0739-6260, https://doi.org/10.1016/0739-6260(86)90042-0.
64. Hentrich, C. and T. Surrey (2010). "Microtubule organization by the antagonistic mitotic motors kinesin-5 and kinesin-14." J Cell Biol **189**(3): 465-480.
65. Higgins, D. G., A. J. Bleasby and R. Fuchs (1992). "CLUSTAL V: improved software for multiple sequence alignment." Comput Appl Biosci **8**(2): 189-191.
66. Hirokawa, N. (1982). "Cross-linker system between neurofilaments, microtubules, and membranous organelles in frog axons revealed by the quick-freeze, deep-etching method." J Cell Biol **94**(1): 129-142.

67. Hirokawa, N. (1998). "Kinesin and dynein superfamily proteins and the mechanism of organelle transport." Science **279**(5350): 519-526.
68. Hirokawa, N. and Y. Noda (2008). "Intracellular transport and kinesin superfamily proteins, KIFs: structure, function, and dynamics." Physiol Rev **88**(3): 1089-1118.
69. Hirose, K., A. Lockhart, R. A. Cross and L. A. Amos (1995). "Nucleotide-dependent angular change in kinesin motor domain bound to tubulin." Nature **376**(6537): 277-279.
70. Hong, P., S. Koza and E. S. Bouvier (2012). "Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates." J Liq Chromatogr Relat Technol **35**(20): 2923-2950.
71. Irvine, G. B. (2001). "Determination of molecular size by size-exclusion chromatography (gel filtration)." Curr Protoc Cell Biol **Chapter 5**: Unit 5 5.
72. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis (2021). "Highly accurate protein structure prediction with AlphaFold." Nature **596**(7873): 583-589.
73. Kashina, A. S., R. J. Baskin, D. G. Cole, K. P. Wedaman, W. M. Saxton and J. M. Scholey (1996). "A bipolar kinesin." Nature **379**(6562): 270-272.
74. Khoshouei, M., M. Radjainia, A. J. Phillips, J. A. Gerrard, A. K. Mitra, J. M. Plitzko, W. Baumeister and R. Danev (2016). "Volta phase plate cryo-EM of the small protein complex Prx3." Nat Commun **7**: 10534.
75. Kim, A. J. and S. A. Endow (2000). "A kinesin family tree." J Cell Sci **113 Pt 21**: 3681-3682.
76. Kirchner, J., G. Woehlke and M. Schliwa (1999). "Universal and unique features of kinesin motors: insights from a comparison of fungal and animal conventional kinesins." Biol Chem **380**(7-8): 915-921.
77. Kleylein-Sohn, J., B. Pollinger, M. Ohmer, F. Hofmann, E. A. Nigg, B. A. Hemmings and M. Wartmann (2012). "Acentrosomal spindle organization renders cancer cells dependent on the kinesin HSET." J Cell Sci **125**(Pt 22): 5391-5402.
78. Konig, S., D. Svergun, M. H. Koch, G. Hubner and A. Schellenberger (1992). "Synchrotron radiation solution X-ray scattering study of the pH dependence of the quaternary structure of yeast pyruvate decarboxylase." Biochemistry **31**(37): 8726-8731.
79. Kostanski, L. K., D. M. Keller and A. E. Hamielec (2004). "Size-exclusion chromatography-a review of calibration methodologies." J Biochem Biophys Methods **58**(2): 159-186.
80. Kozielski, F., S. De Bonis, W. P. Burmeister, C. Cohen-Addad and R. H. Wade (1999). "The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations." Structure **7**(11): 1407-1416.
81. Kuhlbrandt, W. (2014). "Biochemistry. The resolution revolution." Science **343**(6178): 1443-1444.

82. Kull, F. J., E. P. Sablin, R. Lau, R. J. Fletterick and R. D. Vale (1996). "Crystal structure of the kinesin motor domain reveals a structural similarity to myosin." Nature **380**(6574): 550-555.
83. Kuznetsov, S. A. and V. I. Gelfand (1986). "Bovine brain kinesin is a microtubule-activated ATPase." Proc Natl Acad Sci U S A **83**(22): 8530-8534.
84. Kuznetsov, S. A., E. A. Vaisberg, N. A. Shanina, N. N. Magretova, V. Y. Chernyak and V. I. Gelfand (1988). "The quaternary structure of bovine brain kinesin." EMBO J **7**(2): 353-356.
85. Kwon, M., S. A. Godinho, N. S. Chandhok, N. J. Ganem, A. Azioune, M. Thery and D. Pellman (2008). "Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes." Genes Dev **22**(16): 2189-2203.
86. Landen, C. N., Jr., Y. G. Lin, A. Immaneni, M. T. Deavers, W. M. Merritt, W. A. Spannuth, D. C. Bodurka, D. M. Gershenson, W. R. Brinkley and A. K. Sood (2007). "Overexpression of the centrosomal protein Aurora-A kinase is associated with poor prognosis in epithelial ovarian cancer patients." Clin Cancer Res **13**(14): 4098-4104.
87. Lawrence, C. J., R. K. Dawe, K. R. Christie, D. W. Cleveland, S. C. Dawson, S. A. Endow, L. S. Goldstein, H. V. Goodson, N. Hirokawa, J. Howard, R. L. Malmberg, J. R. McIntosh, H. Miki, T. J. Mitchison, Y. Okada, A. S. Reddy, W. M. Saxton, M. Schliwa, J. M. Scholey, R. D. Vale, C. E. Walczak and L. Wordeman (2004). "A standardized kinesin nomenclature." J Cell Biol **167**(1): 19-22.
88. Lee, B. G., F. Merkel, M. Allegretti, M. Hassler, C. Cawood, L. Lecomte, F. J. O'Reilly, L. R. Sinn, P. Gutierrez-Escribano, M. Kschonsak, S. Bravo, T. Nakane, J. Rappsilber, L. Aragon, M. Beck, J. Lowe and C. H. Haering (2020). "Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism." Nat Struct Mol Biol **27**(8): 743-751.
89. LINDQVIST, B., STORGÅRDS, T. Molecular-sieving Properties of Starch. *Nature* **175**, 511–512 (1955). https://doi.org/10.1038/175511a0
90. Lo, M. C., A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas and G. Ellestad (2004). "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery." Anal Biochem **332**(1): 153-159.
91. Lockhart, A. and R. A. Cross (1994). "Origins of reversed directionality in the ncd molecular motor." EMBO J **13**(4): 751-757.
92. Manalastas-Cantos, K., P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S. Molodenskiy, A. Panjkovich, H. D. T. Mertens, A. Gruzinov, C. Borges, C. M. Jeffries, D. I. Svergun and D. Franke (2021). "ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis." J Appl Crystallogr **54**(Pt 1): 343-355.
93. Maney, T., A. W. Hunter, M. Wagenbach and L. Wordeman (1998). "Mitotic centromere-associated kinesin is important for anaphase chromosome segregation." J Cell Biol **142**(3): 787-801.
94. Manning, B. D., J. G. Barrett, J. A. Wallace, H. Granok and M. Snyder (1999). "Differential regulation of the Kar3p kinesin-related protein by two associated proteins, Cik1p and Vik1p." J Cell Biol **144**(6): 1219-1233.
95. Matuliene, J., R. Essner, J. Ryu, Y. Hamaguchi, P. W. Baas, T. Haraguchi, Y. Hiraoka and R. Kuriyama (1999). "Function of a minus-end-directed

kinesin-like motor protein in mammalian cells." J Cell Sci **112 (Pt 22)**: 4041-4050.
96. McDonald, H. B., R. J. Stewart and L. S. Goldstein (1990). "The kinesin-like ncd protein of Drosophila is a minus end-directed microtubule motor." Cell **63**(6): 1159-1165.
97. McMullan, G., K. R. Vinothkumar and R. Henderson (2015). "Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy." Ultramicroscopy **158**: 26-32.
98. Meluh, P. B. and M. D. Rose (1990). "KAR3, a kinesin-related gene required for yeast nuclear fusion." Cell **60**(6): 1029-1041.
99. Moores, C. (2008). "Studying microtubules by electron microscopy." Methods Cell Biol **88**: 299-317.
100. Morris, E. P. and P. C. A. da Fonseca (2017). "High-resolution cryo-EM proteasome structures in drug development." Acta Crystallogr D Struct Biol **73**(Pt 6): 522-533.
101. Myers, S. M. and I. Collins (2016). "Recent findings and future directions for interpolar mitotic kinesin inhibitors in cancer therapy." Future Med Chem **8**(4): 463-489.
102. Nagayama, K. and R. Danev (2008). "Phase contrast electron microscopy: development of thin-film phase plates and biological applications." Philos Trans R Soc Lond B Biol Sci **363**(1500): 2153-2162.
103. Nakane, T., A. Kotecha, A. Sente, G. McMullan, S. Masiulis, P. Brown, I. T. Grigoras, L. Malinauskaite, T. Malinauskas, J. Miehling, T. Uchanski, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. Murshudov, A. R. Aricescu and S. H. W. Scheres (2020). "Single-particle cryo-EM at atomic resolution." Nature **587**(7832): 152-156.
104. Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nat Protoc **2**(9): 2212-2221.
105. O'Connell, M. J., P. B. Meluh, M. D. Rose and N. R. Morris (1993). "Suppression of the bimC4 mitotic spindle defect by deletion of klpA, a gene encoding a KAR3-related kinesin-like protein in Aspergillus nidulans." J Cell Biol **120**(1): 153-162.
106. O'Farrell, P. H., B. A. Edgar, D. Lakich and C. F. Lehner (1989). "Directing cell division during development." Science **246**(4930): 635-640.
107. Ohi, M., Y. Li, Y. Cheng and T. Walz (2004). "Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy." Biol Proced Online **6**: 23-34.
108. Orlova, E. V. and H. R. Saibil (2011). "Structural analysis of macromolecular assemblies by electron microscopy." Chem Rev **111**(12): 7710-7748.
109. Orthaber, D. and O. Glatter (2000). "Synthetic phospholipid analogs: a structural investigation with scattering methods." Chem Phys Lipids **107**(2): 179-189.
110. Orun, O., M. H. Koch, B. Kan, D. I. Svergun, M. V. Petoukhov and Z. Sayers (2003). "Structural characterization of T-protein of the Escherichia coli glycine cleavage system by X-ray small angle scattering." Cell Mol Biol (Noisy-le-grand) **49 Online Pub**: OL453-459.

111. Pannu, V., K. Mittal, G. Cantuaria, M. D. Reid, X. Li, S. Donthamsetty, M. McBride, S. Klimov, R. Osan, M. V. Gupta, P. C. Rida and R. Aneja (2015). "Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers." Oncotarget **6**(12): 10487-10497.
112. Park, H. W., Z. Ma, H. Zhu, S. Jiang, R. C. Robinson and S. A. Endow (2017). "Structural basis of small molecule ATPase inhibition of a human mitotic kinesin motor protein." Sci Rep **7**(1): 15121.
113. Passmore, L. A. and C. J. Russo (2016). "Specimen Preparation for High-Resolution Cryo-EM." Methods Enzymol **579**: 51-86.
114. Penczek, P., M. Radermacher and J. Frank (1992). "Three-dimensional reconstruction of single particles embedded in ice." Ultramicroscopy **40**(1): 33-53.
115. Penczek, P. A. (2010). "Resolution measures in molecular electron microscopy." Methods Enzymol **482**: 73-100.
116. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.
117. Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nat Methods **14**(3): 290-296.
118. Pusey PN. Correlation and light beating spectroscopy. In: Cummings HZ, Pike ER, editors. *Photon correlation and light beating spectroscopy.* New York: Plenum; 1972. pp. 387–428.
119. Quintyne, N. J., J. E. Reing, D. R. Hoffelder, S. M. Gollin and W. S. Saunders (2005). "Spindle multipolarity is prevented by centrosomal clustering." Science **307**(5706): 127-129.
120. Rambo, R. P. and J. A. Tainer (2010). "Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering." Curr Opin Struct Biol **20**(1): 128-137.
121. Rank, K. C., C. J. Chen, J. Cope, K. Porche, A. Hoenger, S. P. Gilbert and I. Rayment (2012). "Kar3Vik1, a member of the kinesin-14 superfamily, shows a novel kinesin microtubule binding pattern." J Cell Biol **197**(7): 957-970.
122. Receveur-Brechot, V. and D. Durand (2012). "How random are intrinsically disordered proteins? A small angle scattering perspective." Curr Protein Pept Sci **13**(1): 55-75.
123. Reinemann, D. N., S. R. Norris, R. Ohi and M. J. Lang (2018). "Processive Kinesin-14 HSET Exhibits Directional Flexibility Depending on Motor Traffic." Curr Biol **28**(14): 2356-2362 e2355.
124. Rieder, C. L. and S. P. Alexander (1990). "Kinetochores are transported poleward along a single astral microtubule during chromosome attachment to the spindle in newt lung cells." J Cell Biol **110**(1): 81-95.
125. Ring, D., R. Hubble and M. Kirschner (1982). "Mitosis in a cell with multiple centrioles." J Cell Biol **94**(3): 549-556.

126. Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." J Struct Biol **192**(2): 216-221.
127. Sablin, E. P., R. B. Case, S. C. Dai, C. L. Hart, A. Ruby, R. D. Vale and R. J. Fletterick (1998). "Direction determination in the minus-end-directed kinesin motor ncd." Nature **395**(6704): 813-816.
128. Sablin, E. P., F. J. Kull, R. Cooke, R. D. Vale and R. J. Fletterick (1996). "Crystal structure of the motor domain of the kinesin-related motor ncd." Nature **380**(6574): 555-559.
129. Sader, K., R. Matadeen, P. Castro Hartmann, T. Halsan and C. Schlichten (2020). "Industrial cryo-EM facility setup and management." Acta Crystallogr D Struct Biol **76**(Pt 4): 313-325.
130. Sahin, E. and C. J. Roberts (2012). "Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms." Methods Mol Biol **899**: 403-423.
131. Saint-Dizier, F., T. P. Matthews, A. M. Gregson, H. Prevet, T. McHardy, G. Colombano, H. Saville, M. Rowlands, C. Ewens, P. C. McAndrew, K. Tomlin, D. Guillotin, G. W. Mak, K. Drosopoulos, I. Poursaitidis, R. Burke, R. van Montfort, S. Linardopoulos and I. Collins (2023). "Discovery of 2-(3-Benzamidopropanamido)thiazole-5-carboxylate Inhibitors of the Kinesin HSET (KIFC1) and the Development of Cellular Target Engagement Probes." J Med Chem **66**(4): 2622-2645.
132. Sawin, K. E., T. J. Mitchison and L. G. Wordeman (1992). "Evidence for kinesin-related proteins in the mitotic apparatus using peptide antibodies." J Cell Sci **101 (Pt 2)**: 303-313.
133. Scheres, S. H. (2010). "Classification of structural heterogeneity by maximum-likelihood methods." Methods Enzymol **482**: 295-320.
134. Scheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." J Struct Biol **180**(3): 519-530.
135. Scheres, S. H. (2015). "Semi-automated selection of cryo-EM particles in RELION-1.3." J Struct Biol **189**(2): 114-122.
136. Scherzer, O (1936) Über einige Fehler von Elektronenlinsen.*Zeitschrift für Physik* 101(9-10): 593-603.
137. Schliwa, M. (1984). "Mechanisms of intracellular organelle transport." Cell Muscle Motil **5**: 1-82,403-406.
138. Schmidhuber, J. (2015). "Deep learning in neural networks: an overview." Neural Netw **61**: 85-117.
139. Schmidt, T. G., J. Koepke, R. Frank and A. Skerra (1996). "Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin." J Mol Biol **255**(5): 753-766.
140. Schmitt, J., H. Hess and H. G. Stunnenberg (1993). "Affinity purification of histidine-tagged proteins." Mol Biol Rep **18**(3): 223-230.
141. Schnapp, B. J., R. D. Vale, M. P. Sheetz and T. S. Reese (1985). "Single microtubules from squid axoplasm support bidirectional movement of organelles." Cell **40**(2): 455-462.
142. Schwartz, O., J. J. Axelrod, S. L. Campbell, C. Turnbaugh, R. M. Glaeser and H. Muller (2019). "Laser phase plate for transmission electron microscopy." Nat Methods **16**(10): 1016-1020.

143. Semisotnov, G. V., N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas and R. I. Gilmanshin (1991). "Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe." Biopolymers **31**(1): 119-128.
144. Serdyuk, I. N., T. N. Tsalkova, D. I. Svergun and T. D. Izotova (1987). "Determination of radii of gyration of particles by small-angle neutron scattering: calculation of the effect of aggregates." J Mol Biol **194**(1): 126-128.
145. Sgro, G. G. and T. R. D. Costa (2018). "Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis." Front Mol Biosci **5**: 74.
146. She, Z. Y. and W. X. Yang (2017). "Molecular mechanisms of kinesin-14 motors in spindle assembly and chromosome segregation." J Cell Sci **130**(13): 2097-2110.
147. Sigworth, F. J. (2004). "Classical detection theory and the cryo-EM particle selection problem." J Struct Biol **145**(1-2): 111-122.
148. Sjostrand, J. and M. Frizell (1975). "Retrograde axonal transport of rapidly migrating proteins in peripheral nerves." Brain Res **85**(2): 325-330.
149. Skold, H. N., D. J. Komma and S. A. Endow (2005). "Assembly pathway of the anastral Drosophila oocyte meiosis I spindle." J Cell Sci **118**(Pt 8): 1745-1755.
150. Some, D., H. Amartely, A. Tsadok and M. Lebendiker (2019). "Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS)." J Vis Exp(148).
151. Stearns, T. (2001). "Centrosome duplication. a centriolar pas de deux." Cell **105**(4): 417-420.
152. Stetefeld, J., S. A. McKenna and T. R. Patel (2016). "Dynamic light scattering: a practical guide and applications in biomedical sciences." Biophys Rev **8**(4): 409-427.
153. Stewart, R. J., J. Semerjian and C. F. Schmidt (1998). "Highly processive motility is not a general feature of the kinesins." Eur Biophys J **27**(4): 353-360.
154. Striegel, A. M. (2017). "Specific refractive index increment (partial differentialn/ partial differentialc) of polymers at 660 nm and 690 nm." Chromatographia **80**(6): 989-996.
155. Spence J. C. H.*High Resolution Microscopy*, 3rd ed.; Oxford University Press: Cary, NC, 2003.
156. Svoboda, K., C. F. Schmidt, B. J. Schnapp and S. M. Block (1993). "Direct observation of kinesin stepping by optical trapping interferometry." Nature **365**(6448): 721-727.
157. Tan, Z., Y. Yue, F. da Veiga Leprevost, S. E. Haynes, V. Basrur, A. I. Nesvizhskii, K. J. Verhey and M. A. Cianfrocco (2023). "Autoinhibited kinesin-1 adopts a hierarchical folding pattern." bioRxiv.
158. Tao, L., A. Mogilner, G. Civelekoglu-Scholey, R. Wollman, J. Evans, H. Stahlberg and J. M. Scholey (2006). "A homotetrameric kinesin-5, KLP61F, bundles microtubules and antagonizes Ncd in motility assays." Curr Biol **16**(23): 2293-2302.
159. Taylor, K. A. and R. M. Glaeser (1974). "Electron diffraction of frozen, hydrated protein crystals." Science **186**(4168): 1036-1037.

160. Theurkauf, W. E. and R. S. Hawley (1992). "Meiotic spindle assembly in Drosophila females: behavior of nonexchange chromosomes and the effects of mutations in the nod kinesin-like protein." J Cell Biol **116**(5): 1167-1180.
161. Thompson, R. F., M. Walker, C. A. Siebert, S. P. Muench and N. A. Ranson (2016). "An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology." Methods **100**: 3-15.
162. Timasheff, S. N. (1993). "The control of protein stability and association by weak interactions with water: how do solvents affect these processes?" Annu Rev Biophys Biomol Struct **22**: 67-97.
163. Unwin, P. N. (1971). "Phase contrast and interference microscopy with the electron microscope." Philos Trans R Soc Lond B Biol Sci **261**(837): 95-104.
164. Vagenende, V., M. G. Yap and B. L. Trout (2009). "Mechanisms of protein stabilization and prevention of protein aggregation by glycerol." Biochemistry **48**(46): 11084-11096.
165. Vale, R. D. and R. J. Fletterick (1997). "The design plan of kinesin motors." Annu Rev Cell Dev Biol **13**: 745-777.
166. Vale, R. D., T. S. Reese and M. P. Sheetz (1985). "Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility." Cell **42**(1): 39-50.
167. van Heel, M., G. Harauz, E. V. Orlova, R. Schmidt and M. Schatz (1996). "A new generation of the IMAGIC image processing system." J Struct Biol **116**(1): 17-24.
168. Vernos, I., J. Heasman and C. Wylie (1993). "Multiple kinesin-like transcripts in Xenopus oocytes." Dev Biol **157**(1): 232-239.
169. von Loeffelholz, O. and B. P. Klaholz (2021). "Setup and Troubleshooting of Volta Phase Plate Cryo-EM Data Collection." Methods Mol Biol **2305**: 291-299.
170. von Loeffelholz, O., G. Papai, R. Danev, A. G. Myasnikov, S. K. Natchiar, I. Hazemann, J. F. Menetret and B. P. Klaholz (2018). "Volta phase plate data collection facilitates image processing and cryo-EM structure determination." J Struct Biol **202**(3): 191-199.
171. Wade, R.H. A brief look at imaging and contrast transfer, Ultramicroscopy, Volume 46, Issues 1–4, 1992, Pages 145-156, ISSN 0304-3991, https://doi.org/10.1016/0304-3991(92)90011-8.
172. Wagner, T., F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum, D. Quentin, D. Roderer, S. Tacke, B. Siebolds, E. Schubert, T. R. Shaikh, P. Lill, C. Gatsogiannis and S. Raunser (2019). "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM." Commun Biol **2**: 218.
173. Wagner, T. and S. Raunser (2020). "The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows." Commun Biol **3**(1): 61.
174. Walczak, C. E., S. Verma and T. J. Mitchison (1997). "XCTK2: a kinesin-related protein that promotes mitotic spindle assembly in Xenopus laevis egg extracts." J Cell Biol **136**(4): 859-870.
175. Wang, F., H. Gong, G. Liu, M. Li, C. Yan, T. Xia, X. Li and J. Zeng (2016). "DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM." J Struct Biol **195**(3): 325-336.

176. Wen, J., T. Arakawa and J. S. Philo (1996). "Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions." Anal Biochem **240**(2): 155-166.
177. Wendt, T. G., N. Volkmann, G. Skiniotis, K. N. Goldie, J. Muller, E. Mandelkow and A. Hoenger (2002). "Microscopic evidence for a minus-end-directed power stroke in the kinesin motor ncd." EMBO J **21**(22): 5969-5978.
178. Wilgenbusch, J. C. and D. Swofford (2003). "Inferring evolutionary trees with PAUP*." Curr Protoc Bioinformatics **Chapter 6**: Unit 6 4.
179. Wilson AJC. Small-angle scattering of X-rays by A. Guinier and G. Fournet. *Acta Cryst*. 1956;**9**:326.
180. Woehlke, G., A. K. Ruby, C. L. Hart, B. Ly, N. Hom-Booher and R. D. Vale (1997). "Microtubule interaction site of the kinesin motor." Cell **90**(2): 207-216.
181. Wu, J., K. Mikule, W. Wang, N. Su, P. Petteruti, F. Gharahdaghi, E. Code, X. Zhu, K. Jacques, Z. Lai, B. Yang, M. L. Lamb, C. Chuaqui, N. Keen and H. Chen (2013). "Discovery and mechanistic study of a small molecule inhibitor for motor protein KIFC1." ACS Chem Biol **8**(10): 2201-2208.
182. Yamagishi, M., H. Shigematsu, T. Yokoyama, M. Kikkawa, M. Sugawa, M. Aoki, M. Shirouzu, J. Yajima and R. Nitta (2016). "Structural Basis of Backwards Motion in Kinesin-1-Kinesin-14 Chimera: Implication for Kinesin-14 Motility." Structure **24**(8): 1322-1334.
183. Yang, J. T., R. A. Laymon and L. S. Goldstein (1989). "A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses." Cell **56**(5): 879-889.
184. Yang, J. T., W. M. Saxton and L. S. Goldstein (1988). "Isolation and characterization of the gene encoding the heavy chain of Drosophila kinesin." Proc Natl Acad Sci U S A **85**(6): 1864-1868.
185. Yang, Z., J. Fang, J. Chittuluru, F. J. Asturias and P. A. Penczek (2012). "Iterative stable alignment and clustering of 2D transmission electron microscope images." Structure **20**(2): 237-247.
186. Yao, Y. and W. Dai (2014). "Genomic Instability and Cancer." J Carcinog Mutagen **5**.
187. Young, G., N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E. G. Marklund, M. P. Collier, S. A. Chandler, O. Tkachenko, J. Allen, M. Crispin, N. Billington, Y. Takagi, J. R. Sellers, C. Eichmann, P. Selenko, L. Frey, R. Riek, M. R. Galpin, W. B. Struwe, J. L. P. Benesch and P. Kukura (2018). "Quantitative mass imaging of single biological macromolecules." Science **360**(6387): 423-427.
188. Zegzouti, H., M. Zdanovskaia, K. Hsiao and S. A. Goueli (2009). "ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases." Assay Drug Dev Technol **7**(6): 560-572.
189. Zhang, P., W. Dai, J. Hahn and S. P. Gilbert (2015). "Drosophila Ncd reveals an evolutionarily conserved powerstroke mechanism for homodimeric and heterodimeric kinesin-14s." Proc Natl Acad Sci U S A **112**(20): 6359-6364.
190. Zhang, W., L. Zhai, Y. Wang, R. J. Boohaker, W. Lu, V. V. Gupta, I. Padmalayam, R. J. Bostwick, E. L. White, L. J. Ross, J. Maddry, S.

Ananthan, C. E. Augelli-Szafran, M. J. Suto, B. Xu, R. Li and Y. Li (2016). "Discovery of a novel inhibitor of kinesin-like protein KIFC1." Biochem J **473**(8): 1027-1035.

191. Zivanov, J., T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl and S. H. Scheres (2018). "New tools for automated high-resolution cryo-EM structure determination in RELION-3." Elife **7**.