OSAKA UNIVERSITY

DOCTORAL THESIS

Development of a versatile domain mapping system for intermediate-resolution electron microscopy using a portable peptide tag

電子顕微鏡観察における新規蛋白質ラベリングシステムの開発

Author:
Zuben Patrick BROWN

Supervisor:
Professor Junichi TAKAGI

Examining committee:
Professor Keichi NAMBA
Professor Atsushi NAKAGAWA
Professor Masahiro UEDA

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Laboratory of Protein Synthesis and Expression
Graduate School of Frontier Biosciences

March 5, 2018

"49. Be like the promontory against which the waves continually break, but it stands firm and tames the fury of the water around it.
. . . Remember, too, an every occasion which leads thee to vexation to apply this principle: not that this is a misfortune, but that to bear it nobly is good fortune."

Marcus Aurelius Antoninus, 121–180, *The Meditations of Marcus Aurelius.*

"51.
The Moving Finger writes; and,
having writ,
Moves on: nor all thy Piety nor Wit
Shall lure it back to cancel half a
Line,
Nor all your Tears wash out a Word
of it.

52.
And that inverted Bowl we call the
Sky,
Whereunder crawling coop't we live
and die,
Lift not your hands to It for help–
for It
Rolls impotently on as Thou or I."

Edward FitzGerald, 1859 *The Rubaiyat of Omar Khayyam*

OSAKA UNIVERSITY

Abstract

Graduate School of Frontier Biosciences

Doctor of Philosophy

Development of a versatile domain mapping system for intermediate-resolution electron microscopy using a portable peptide tag

by Zuben Patrick BROWN

Electron microscopy (EM) can be used to obtain the three dimensional structural information of macromolecular complexes. When data are obtained at a sufficiently high (approx. 3 Å) resolution, one can build atomic models of the complexes with relatively high accuracy. In contrast, when the EM images are collected at lower resolutions, even the identity of domains and subunits may be unclear. Gene fusion technologies are available to label subunits or domains to enable positional mapping by comparisons between the wild type and the mutated construct, but artificial fusion of protein domains within a target polypeptide may cause unwanted structural alterations. Exogenous labeling by Fab fragments from antibodies can bypass this complication because it does not require genetic modifications on the target protein. However, it requires a panel of high-affinity antibodies that recognize a wide variety of epitopes, making this method highly costly. Alternatively, insertion of small linear peptide tags into the target protein and then binding the Fab from the cognate antibody can be used. This, too, has limitation because standard peptide tag systems usually require their placement in terminal regions of a protein, reducing the versatility of the mapping. The recently developed PA dodecapeptide epitope tag (GVAMPGAEDDVV), forms a tight β-turn in the antigen binding pocket of its antibody (NZ-1), which allows for the epitope to remain reactive even when inserted into surface-exposed loops within central domains of various proteins. As a test case for EM labeling with PA tag and NZ-1 Fab, a multi-domain cell adhesion receptor $\alpha_{IIb}\beta_3$ integrin was chosen and I generated several different constructs with PA tag inserted centrally with the α_{IIb}-subunit. I confirmed that the purified PA-tagged integrin ectodomain fragments form a stable complex with NZ-1 Fab, a property that is necessary for a successful EM label. Furthermore, the negatively stained EM images showed that a majority of the particles exhibited a clear density corresponding to the NZ-1 Fab. The positions of the bound Fab were also in good agreement with the predicted location of the inserted PA tag. The high affinity and insertion compatibility of the PA tag system allows for its use as a new EM labeling methodology applicable to proteins for which good antibodies are not available.

Acknowledgements

First and foremost, I need to thank professor Junichi Takagi, the boss of the lab and the driving force behind this project and all the interesting science that is being done in his lab.The head of the electron microscopy group, associate professor Kenji Iwasaki was indispensable for all the imaging and analysis. Dr. Takao Arimori has been very helpful with this project and in general as a very knowledgeable source for all kinds of scientific questions and discussions.

Various members of the Takagi lab that have since moved to other laboratories or countries have also been very helpful and I need to thank. Including Dr. Yu Kitago (location) who helped me when I started using Linux, and for many science questions that he always had a careful and accurate answer for. Kyoko Matoba helped me with both my first ever gel filtration and use of the Hitachi 80 kV EM. Her explanation and teaching was exemplary and any errors that I encountered where my own. Operation of the JEM-2200FS was helped greatly by Dr. Shintaro Maeda (who has since moved to Scrips) and Hirose. Emiko Mihara has been vital for all my general cell culturing questions and advice, as well as lending me (apparently) the only USB I could find when I needed to practice my presentations. Dr. Hidenori Hirai and Dr. Masataka Umitsu are two excellent researchers who I relied on excessively during my master's and can't emphasize how important their advice, and example as exemplary scientists was for both my research and development as a scientist.

I need to think my parents for supporting me during the entirety of my PhD. Finally, I need to thank my wife who has been a vital support for the last four years and without I would not have been able to do half of what I did.

Zuben P. Brown
March 5, 2018

Contents

List of Figures

List of Tables

List of Abbreviations

ASA	**A**ccessible **S**urface **A**rea
CCF	**C**ross **C**orrelation **F**unction
CCR	**C**ross **C**o**r**relaton
CDR	**C**omplimentary **D**etermining **R**egion
CTF	**C**ontrast **T**ransfer **F**unction
DID	**D**ynein Light chain-**I**nteracting **D**omain
DOLORS	**D**o**L**ocalisation by **R**CT **S**ampling
DMEM	**D**ulbecco's **M**odified **E**agle **M**edium
DNA	**D**eoxy**R**ibonucleic **A**cid
DQE	**D**ose **Q**uantum **E**fficiency
DTT	**D**i**t**hio**t**hreitol
Fab	**F**ragment **a**ntigen-**b**inding
FSC	**F**ourier **S**hell **C**orrelation
GFP	**G**reen **F**luorescent **P**rotein
GST	**G**lutathione **S**-**T**ransferase
IgG	**I**mmuno**g**lobulin **G**
LIBS	**L**igand **I**nduced **B**inding **S**ite
MBP	**M**altose-**B**inding **P**rotein
MTF	**M**odulation **T**ransfer **F**requency
MWCO	**M**olecular **W**eight **C**ut-**O**ff
PAGE	**P**oly **A** **G**el **E**lectrophoresis
PDB	**P**rotein **D**ata **B**ank
RMSD	**R**oot **M**ean **S**quare **D**eviation
SDS	**S**odium **D** **S**
SEC	**S**ize **E**xclusion **C**hromotography
SNR	**S**ignal to **N**oise **R**atio
TBS	**T**ris-**B**uffered **S**aline
Tris	**T**ris(hydroxymethyl)aminomethane

Of course, to my wife.

Chapter 1

Introduction

1.1 Introduction

While pursuing my education in structural biology I have often been struck with the observation that "life is structure". While there is surely more to life than the mechanical motion of its constituent parts, there is some close link between the structure and function of many biological components. This is perhaps much less profound than I am implying, as on a larger scale the function of objects is based very strongly on their macro-structure and chemical (or atomic) composition. The utility of an object as mundane as my desk is strongly linked to its structure (flat) and relatively inert chemical composition (consider how different writing a paper would be if we used desks made of metallic sodium). Objects at the molecular and atomic scale also seem[1] to follow this principle, that their function is linked to their structural and chemical properties. Although making a different point, the observation made by Thomas Hobbes could also be applied to the molecular world:

> For seeing life is but a motion of Limbs, ... For what is the Heart, but a Spring; and the Nerves, but so many Strings; and the Joynts, but so many Wheeles, giving motion to the whole Body, such as was intended by the Artificer?[2]
>
> Thomas Hobbes, 1651, *The Leviathan*

Each body part has a function that even in the 1650 was linked in some way, although incorrectly, to its structure. This underlies an important point: if we know

[1] This is from my perspective as a biologist. Perhaps physicists would dispute this, and it's possible that atoms don't even have shapes since their constituent electrons are just some statistical property.

[2] The full quote in Japanese: 生命とは四肢の運動に他ならず、その運動は内部のある中心部分からはじまる、ということを知るならば、すべての自動機械 Automata(時計のように発條と車で自らうごく機關)は人工的生命をもつといってならぬわけがあろうか。心臓は何かといえば、それは発條にほかならず、神経はそれだけの數の紐にほかならず、關節はそれだけの數の車にほかならないのであって、創造者の意図するような運動を全身にあたえるものではないだろうか。; 水田洋譯 1949年

the structure of something, then we are on the path to understanding its properties and then using this knowledge to remodel the world in a way that is more in line with our desires. My PhD has been, among other things, the development of some skills to investigate the structure and function of some of the fundamental components of life, the protein. I have studied how a beam of highly ionising electrons can be used to determine the structure of very small proteins, and how we can design tools to improve our understanding of their architecture and ultimately gain a better understanding of their function. When I started in this field I made a number of assumptions about imaging with electrons, but the more I have read the less probable the entire field has seemed become. It appears to me now that it truly is *magic* that we can know anything at all about such tiny parts of the universe using only a beam of 'electricity'. Luckily, the evidence that structural biology provides real information about the world is so overwhelming that we can safely enjoy all the amazing structures that have been published and continue to be discovered. The following chapter will present some of that evidence and build a story of how we can use electron microscopy to determine the structure of proteins. Following this I will detail the methods used, with results and discussion in later chapters. Finally, I have tried to confine my whimsical and philosophical musings to this section alone, and so hopefully it doesn't leak into any other part of this thesis.

This thesis is organised into four chapters: The introduction (**Chapter 1**), detailed methods (**Chapter 2**), results (**Chapter 3**), and finally a discussion on the quality, possible uses, future experiments and weaknesses of these presented data (**Chapter 4**). Various additional experiments were completed and are presented in the supplementary data (**Chapter 5**). Additional appendices are included for data that helps complete this work but couldn't find a home in the results chapter. There are some general appendices dealing with some of the more technical aspects of EM research and image analysis (**Appendix A**: the contrast transfer function, **Appendix B**: envelope functions, **Appendix C**: EM guns, and **Appendix D**: image alignment). Finally there is a list of published works relating to this project (**Appendix E**).

Since this introduction has the broadest range of topics I will also give a short outline about its structure. In the following section I will discuss various presently available purification tags (**Section 1.2**), give much more detail about one in particular (**Section 1.3**), offer a brief outline of how electron microscopy can be used to determine the structure of unknown proteins (**Section 1.4**), give some detail on what tools are available for structural study of difficult intermediate resolution structures (**Section 1.4.2**), and finally outline how the PA tag and NZ-1 can be used as a novel EM label (**Section 1.5**).

1.2 Purification tags

Biological science depends on being able to purify and concentrate a protein of interest so we can be confident that we are experimenting on, or observing the 'right' protein. Most target proteins are expressed as some recombinant construct rather than purified from a natural source, and the additional peptides that are added as a purification handle are collectively called affinity tags. They range in size from fewer than 10 residues (such as poly-His) to large inserted domains (such as Fc, GST, or MBP), but they all exploit some enhanced affinity for a purification matrix relative to other components of the target protein mixture (this is reviewed by Terpe, (2003)).

The interaction between the tag and affinity matrix needs to have a number of properties that include: 1) high specificity to reduced contamination, 2) high affinity to promote complete capture of the target protein from the sample, 3) slow dissociation kinetics to allow for extended washing to remove any non-specific binding to the affinity matrix, 4) availability of mild elution conditions to completely remove the target protein from the affinity matrix without damaging it. The fusion tag also must be attached in such a manner to not induce unwanted[3] structural or functional changes to the target protein. Often the chance of impairing a target protein when inserting a fusion tag can be lowered by placing the moiety either the C-, or N-terminals of the protein so as to spatially separate the main polypeptide chain from the introduced sequence by as many residues as possible. Given the complexities of biological systems and the vast array of potential questions there is obviously no perfect tagging system that can be applied to all experiments and so often the best system based on desired yield, required affinity, specificity, and cost, among other parameters, must be empirically determined.

To this end, many different purification systems have been developed (**Table 1.1**) including those based on ion exchange interactions between polyarginine (Sassenfeld and Brewer, 1984), metal-chelate interactions between polyhistidine (Hochuli et al., 1988), affinity for amyloid precursor proteins (Sangawa et al., 2013), interactions between glutathione S-transferase binding to glutathione-resin (Smith and Johnson, 1988), chitin-binding domain and chitin resin (Cantor and Chong, 2001), calmodulin binding peptide binding to calmodulin (Stofko-Hahn et al., 1992), and streptavidin-binding peptide and streptavidin (Keefe et al., 2001). Given the high affinity and specificity of some peptide-antibody interactions they also offer an attractive source of purification tags as their peptides can be relatively small. There are several popular tags including FLAG (Hopp et al., 1988), Myc (Evan et al., 1985), and HA (Field et

[3] In some cases any change at all is unwanted, whereas in other cases it might be less important. It all depends on the 'resolution' of the unwanted change, our data collection, and at what spatial range our scientific question is.

al., 1988). Since the affinity, epitope size, and elution conditions varies there is ongoing research to develop improved epitope based purification systems with TARGET (Nogi et al., 2008; Tabata et al., 2010), AGIA (Yano et al., 2016), CP5 (Takeda et al., 2017) and the MAP (Fujii et al., 2016a), RAP (Fujii et al., 2017), and PA series (Fujii et al., 2014) all being developed in recent years. Of these the last, the PA tag, has shown particular promise as a powerful structural biology tool that has utility above and beyond just use as a protein purification handle.

interactions between glutathione S-transferase binding to glutathione-resin

1.3 The PA tag

The Takagi lab recently reported the development of a novel epitope purification tag that uses the high affinity interaction between the NZ-1 antibody and a dodecapeptide (GVAMPGAEDDVV) called PA tag (Fujii et al., 2014). NZ-1 (IgG_{2a}, λ)was discovered during the search for anti-cancer antibodies as an inhibitor of platelet aggregation by its strong binding to the PLAG domain of podoplanin, a type I transmembrane protein that is over-expressed in cancer cells (Kato et al., 2006). NZ-1 recognised a synthetic peptide based on the PLAG domain and so it was reasoned that it could recognise artificially inserted epitope. The NZ-1/PA peptide interaction has a number of very useful properties that allow it to be used as a powerful protein purification tool. The initial characterization of NZ-1 and the PA peptide it showed a binding affinity that was an order of magnitude higher than similarly prepared FLAG, Myc, or HA peptides against their respective antibodies (M2: anti-FLAG, 9E10: anti-Myc, 4B2: anti-HA) when measured using Biolayer interferometry (**Figure 1.1**). Importantly, the high affinity reaction had a very slow dissociation as shown by the stable signal during the dissociation phase. Such a property is very useful for epitope tag systems as it allows for extensive washing to remove contaminating molecules that have bound non-specifically. The specificity of the PA tag and NZ-1 interaction is also very high, allowing purification of very low concentration proteins from whole cell lystate. Finally, the elution conditions are quite mild (0.1 mg/mL PA peptide) which reduces the chance of damaging the target protein if harsh conditions were required. The PA peptide used to elute a protein sample can be removed from the NZ-1 bound sepharose with 3 M $MgCl^{2+}$ allowing the reuse of the antibody up to 60 times (Fujii et al., 2014). Given these excellent properties further characterisation of the PA tag and NZ

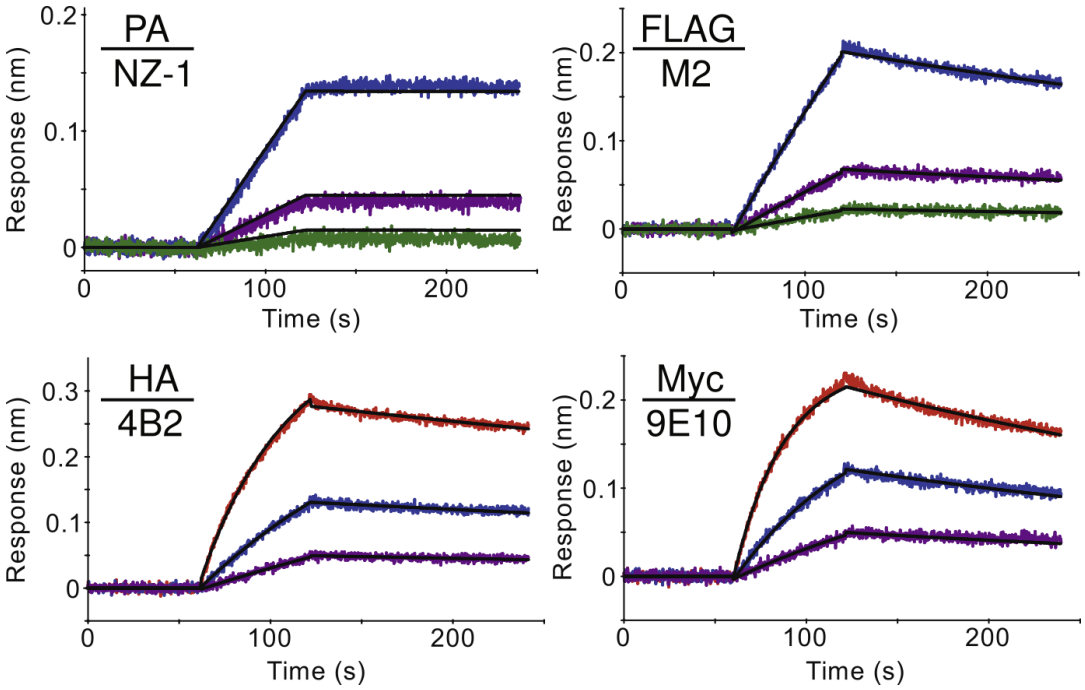


FIGURE 1.1: **Various epitope tags and their binding affinity to their cognate antibodies**. Anti-epitope antibodies (NZ-1, M2, 4B2, 9E10) were immobilised on a sensor tip and the C-terminally fused T4L protein (T4L-PA12, T4L-FLAG, T4L-HA, T4L-Myc) were tested. Buffer equilibration (0-60s), association (60-120s), and dissociation (120-240s) are shown for serial dilutions (red, 30 nM; blue, 10 nM; purple, 3 nM, green, 1 nM). The data was fitted with a global fit using a 1:1 binding model. Reprinted from Protein Expr. Purif., 95, Fujii, Y. et al. PA tag: A versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin, 240-247, Copyright (2014), with permission from Elsevier.

TABLE 1.1: Purifcation tags

Tag name	Seq. or size[A]	Binding mode	Affinity matrix	Location	kD[B]	Elution cond.	Ref.
polyarginine	5-6nR	ion X	SP-sephadex	C	—	NaCl gradient	[1]
polyhistidine	6-10nH	ion X	Ni^{2+}-NTA	N,C	—	imidazole	[2]
FLAG	DYKDDDDK	ab	mAb-matrix	C, N	2.80e−8	EDTA	[3]
Strep-tag	DYKDDDDK	Streptactin			—	Desthiobiotin	[4]
HA	YPYDVPDYA	AB		C?,N	1.60e−9	peptide	[5]
AGIA	EEAAGIARP	ab		C,N	4.90e−9	peptide	[6]
c-myc-tag	EQKLISEEDL	ab		C N	2.20e−9	Low pH, c-myc peptide	[7]
PA tag	GVAMPGAEDDVV	ab	mAb-matrix	various	4.00e−10	PA peptide, $MgCl^{2+}$	[8,9]
MAP	GDGMVPPGIEDK	ab	mAb-matrix	N	3.70e−9	peptide	[10]
RAP	DMVNPGLRDRIE	ab	mAb-matrix	N	9.70e−9	peptide	[11]
S-tag	103 a.a S-protein	domain	RNAaseA		1.00e−7	GITC[D], citrate, $MgCl^{2+}$	[12]
TARGET v1	3x(GYPGQV)	ab		C, N	1.90e−8	propylene glycol, peptide	[13]
TARGET v2	5x(YPGQ)V	ab		C,N	1.00e−8	propylene glycol, peptide	[14]
3x FLAG	3(DYKDDDDK)	ab		C, N	—	EDTA, FLAG peptide	[15]
Calmodulin-binding peptide	26 residues	calmodulin		C, N	—	EGTA	[16]
SBP-tag	size	streptavidin	streptavidin	C	2.50e−9		[17]
chitin-binding domain	seq	chitin	chitin matrix	C, N	—	self cleavage with thiol	[18]
Cellulose-binding domain	27-158 res.	cellulose		N,C	—	Ethylene glycol, low ionic	[19]
GST[E]	201 a.a 26 kDa	domain tag	GST	C	—	reduced glutathione	[20]
MBP[F]	396 a.a 40 kDa	domain tag	amylose	loc	—	10 mM Maltose	[21]
GFP[G]	28 kDa (238 res.)	domain tag		C, N, int	—	Elution	[22]
Starch-binding domain	seq	binding	starch	C	—	β-cyclodextrin	[23]

[A]kDa, [B]mM units of kD, [C]Peptide is concentrated free tag peptide, [D]Guanidinium thiocyanate ,[E]Glutathione S-transferase, [F]Maltose binding protein, [G]Green fluorescent protein.
ref: [1](Sassenfeld and Brewer, 1984), [2](Hochuli et al., 1988), [3](Hopp et al., 1988), [4](Schmidt and Skerra, 2007), [5](Field et al., 1988), [6](Yano et al., 2016), [7](Evan et al., 1985), [8,9](Fujii et al., 2014; Fujii et al., 2016b), [10](Fujii et al., 2016b), [11](Fujii et al., 2017), [12](Connelly et al., 1990), [13](Nogi et al., 2008), [14](Tabata et al., 2010), [15](Hernan et al., 2000), [16](Stofko-Hahn et al., 1992),[17](Keefe et al., 2001), [18](Cantor and Chong, 2001), [19](Tomme et al., 1998), [20](Smith and Johnson, 1988), [21](Maina et al., 1988), [22](Wang and Hazelrigg, 1994), [23](Guillén et al., 2013)

1.3.1 The crystal structure of the PA tag

The association of antigen and antibody depends on the complementarity of the antigen and the antigen active site. This complementarity is regulated amino acids inside or adjacent to the hypervariable region, and is often referred to as the complementarity determining region (CDR) of an antibody (Absolom and Oss, 1986). Determining the structure of this region in the presence and absence (*apo*) of antigen can allow rational discussion of the causes for high or low affinity. To this end, the crystal structure of soluble PA peptide and the NZ-1 fragment antigen-binding (Fab) fragment was determined to better understand the binding behaviour and affinity of the PA peptide (Fujii et al., 2016b). By comparison between the *apo* and bound state it seems apparent that the cause for the high affinity is due to the minimal changes in the CDR upon binding and the unique β-turn conformation of the PA peptide.

Atomic resolution crystals were obtained for NZ-1 Fab fragment in both *apo* form at 1.65 Å and bound with PA14 peptide (EGGVAMPGAEDDVV) at 1.70 Å. Fortunately, each crystal had two Fab per asymmetric unit allowing for both within crystal and between crystal structural comparisons to be made which could allow some statements to be made about the stability of the bound and unbound states. Fujii et al., (2016b) calculated the C_α root-mean-square deviation (RMSD)[4] which allows for some analysis and generalisations to be made about the innate flexibility of the NZ-1 CDR structure. The RMSD of C_α atoms within the CDR was only 0.123 Åfor the *apo* form, increasing to 0.237 Å upon peptide binding which shows that the NZ-1 binding pocket is relatively structurally stable in either the bound or unbound states. Given the low RMSD this could help explain the high specificity of the NZ-1/PA peptide interaction (i.e., a high RMSD would indicate that the antibody can be present in many different conformations and therefore might recognise many different antigens).

Importantly, the RMSD between the *apo* and bound states was only 0.466 Å which is relatively small and shows that the binding pockets between these two states are essentially the same. This suggests that NZ-1 does not use an "induced-fit" binding mode that is common when the CDR undergoes a large conformational change. Furthermore, several water molecules that take part in the hydrogen bonding network that stabilises the bound PA peptide where already present in the *apo* form. The small RMSD between *apo* and bound structures, along with the presence of water prior to peptide binding suggests that the NZ-1 binding pocket it 'primed'

[4]The root-mean-square deviation is the measure of the average distance between eqiuvalent atoms (such as C_α, but can be any atom) of superimposed molecules. It can be calculated with the following equation:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \tag{1.1}$$

δi is the distance between atom i and the average of the equivalent atom in the reference structure.

for antigen recognition and would only pay a small entropic cost.

A second factor that could help explain the high binding affinity is the overall structure of the binding pocket. The heavy and light chains of NZ-1 form a deep cleft that buries over 1200 Å of the total solvent-accessible surface area (ASA) of the PA peptide (**Figure 1.2**). Relative to other known antibody-peptide interactions sites this is not particularly large (Chen et al., 2013), but there are many hydrogen bonds and salt bridges that can form across the interface and the shape complimentarity of the van der Waals forces could account for the large enthalpic gain upon antigen binding.

Finally, the secondary structure of the PA peptide appears to have a contribution to the high binding affinity with NZ-1. Alanine scanning showed that the central 7 residues of the PA peptide were necessary for recognition by NZ-1 (Fujii et al., 2014). The cyrstal structure supported these data by showing that these central residues (shown in bold GVA**MPGAEDD**VV) made direct contact with the NZ-1 antibody. The central MPGA motif formed a type II β-turn which is a commonly observed structural motif for proteins in solution that have a Pro-GLy motifs (Guruprasad and Rajkumar, 2000). If the PA peptide forms this in solution then it could also be 'primed' for recognition by NZ-1 and would further explain the high binding affinity as it is entropically favourable.

The crystal structure of the NZ-1 Fab and PA peptide revealed several structural features that likely contribute to the very high binding affinity. The CDR is primed in a conformation that is ready to bind to PA peptide with the binding pocket undergoing minimal conformational change between *apo* and bound states. Furthermore, the pocket forms a deep cleft to capture the PA peptide and has a large surface area for contacts which could contribute to a lower chance of dissociation. Finally, the PA peptide also forms a β-turn that can may pre-form in solution further allowing entropically favourable binding. From a structural perspective, the β-turn conformation suggested that there are other applications for the PA peptide, as the N-, and C-terminal of the peptide are oriented in the same direction and only 10.2 Å apart suggesting that this peptide could be inserted into loops within protein central domains (such as a loop) and still remain reactive with its antibody, something that is currently lacking with current generation epitope tags.

1.3.2 The PA tag as a mobile epitope

The conformation β-turn conformation of PA peptide suggested that it can be inserted into surface-exposed unstructured loops in central domains of proteins and still remain reactive with NZ-1. This was demonstrated with the platelet adhesion

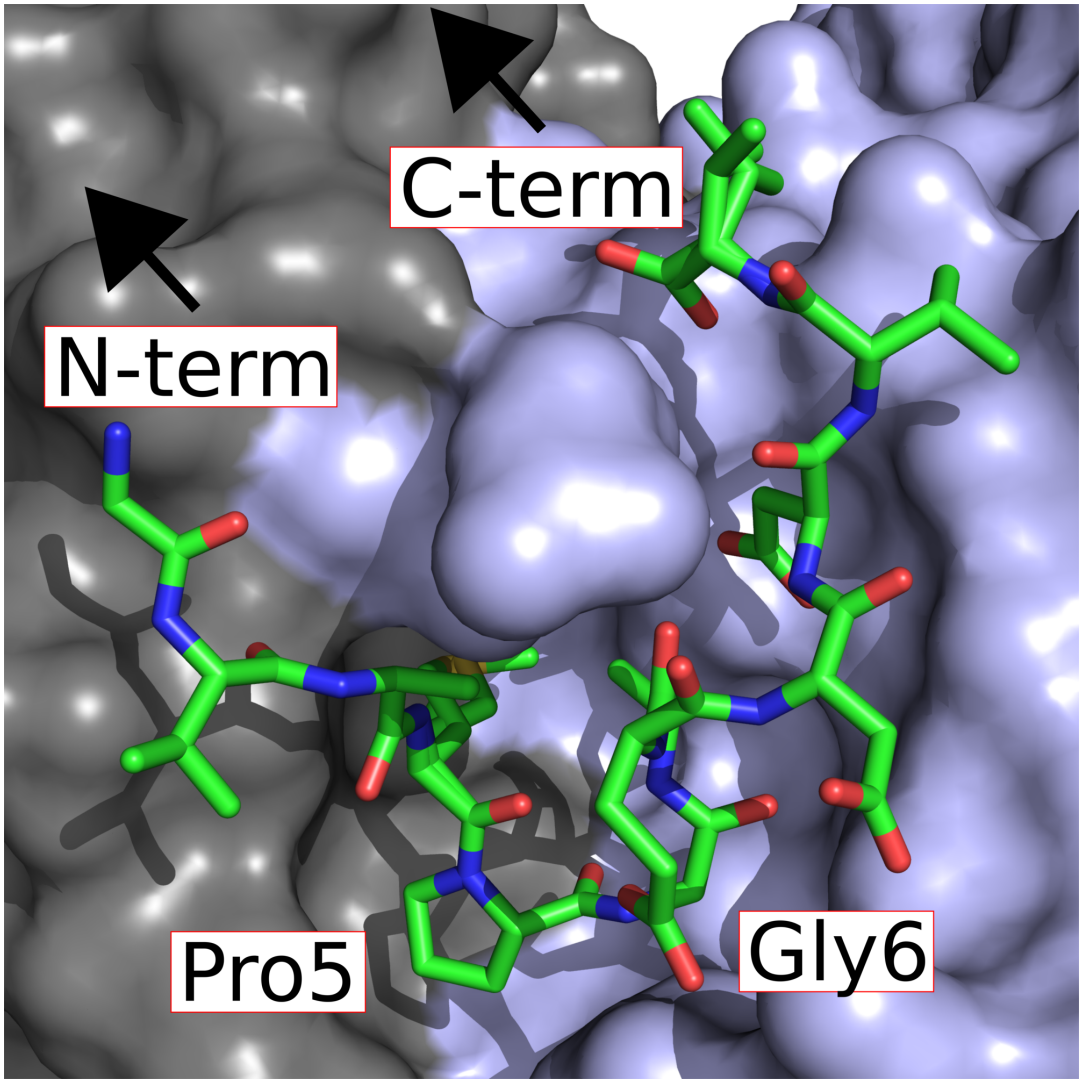


FIGURE 1.2: **PA14 peptide in NZ-1 binding pocket**. Surface representation of the NZ-1 Fab binding pocket with heavy chain (light blue) and light chain (grey). PA14 peptide as stick structure with nitrogen (blue), carbon (green), sulphur (gold), and oxygen (red) shown. Prepared in Pymol from PDB file 4yo0.

receptor $\alpha_{IIb}\beta_3$ integrin that had either the FLAG, Myc, or PA tag inserted into identical unstructured loops with between zero and six additional linker residues (Fujii et al., 2016b). Only the PA tag insertions remained reactive with its cognate antibody, whereas both FLAG and Myc lost this ability, presumably due to the distortion of their epitopes by the neighbouring polypeptide sequence. Several other unstructured loops were chosen with integrin and the PA tag insertion remained reactive with the NZ-1 antibody.

If the PA tag to be inserted into a central surface exposed loop without additional linkers and still remain reactive with the NZ-1 antibody this means that it could function as a mobile epitope[5]. Using rational placement of the PA tag the binding of NZ-1 (or non-binding) could be used as a conformational reporter in proteins that undergo some large structural change during their function. This has already been shown for a number of proteins that fortuitously had high affinity antibodies available against some domain that alternated between exposed and hidden during conformational change (Dennison et al., 2014; Humphries et al., 2003; Irannejad et al., 2013; Walker et al., 2004). However, native antibodies that can function as conformational reporters are not available for many proteins and may only be discovered by chance or determined investigation. Other conformational reporter strategies have used small proteins such as GFP (Calleja et al., 2003) or cutinase (Bonasio et al., 2007) but they have not seen wide spread use possibly due to adverse structural changes that may occur from insertions of additional whole domains. The PA tag, however, is only 12 amino acids and has minimal effect on the target protein structure upon insertion, and so this kind of conformational reporter could be quite useful. The demonstration of PA tag as a functional mobile epitope (Fujii et al., 2016b) suggests that it can have a wide range of other uses include in electron microscopy.

1.4 EM and structural determination

Electron microscopy (EM): the use of electrons to determine the structure of some unknown protein is a very powerful technique with a long history that is covered in much detail in many very good textbooks (Frank, 2006; Reimer and Kohl, 2008). For our purposes that level of detail is unnecessary, although it is worth being at least partially familiar with the details. Here I will not present the physics behind image formation but rather will give the view that I built over my PhD and from the perspective of a biologist[6]. I will present things dogmatically and will leave it to my reader to investigate the more technical textbooks that outline and explain

[5] A mobile epitope is a sequence that can be moved with relative freedom and remain reactive with a single paratope. I would not consider generating different monoclonal antibodies against various regions of a protein to be a 'mobile epitope', for example.

[6] That is, wrong about the physics but right enough to be able to obtain biochemically verifiable and useful data.

the precise physics behind transmission electron microscopy. The simplest approximation for why we can use electron microscopy to determine the structure of an object is that by placing something in the path of an electron beam we can find out something about its structure by observing the change in the electrons path. Things become complicated very quickly since there are different sources of contrast in our images (amplitude and phase) and by considering the different orientations and conformations that real objects can take, but initially we should consider our object to have only one orientation and be comprised of a single shape. It should be clear that there will be a different image if the object is transparent to electrons compared to one that is opaque. The first point to consider is that there is some linear[7] relationship between the object we imaging and our obtained projection (Hawkes, 2006). A very simplified[8] representation of this can be that for some real object U, there is a relationship between it and a projection so that $U \propto p$. For biological samples electrons are very damaging and quickly destroy first the high resolution information and soon all desired information in the sample (Glaeser, 2016). This means that there is a trade off between increasing the beam intensity (which allows us to see the projection) and not destroying the sample we are observing. Practically that means that we operate the microscopes with a very low[9] beam intensity which increases the noise that is associated with each image. Here a we can complicate out model by introducing a Gaussian[10] noise term to our projection so that the projection we obtain is degraded by noise caused by the low beam intensity. In reality the noise associated with the EM actually follows both Gaussian and Poisson distributions depending on the source (the electronics or the scope) but they can both be approximated by an additive Gaussian distribution (Hegerl and Hoppe, 1976).

$$U \propto p = S + N_{Gaussian} \tag{1.3}$$

In simple terms, there is a relationship between the real object its projection *modified* by a noise term. The 'true' signal S has a noise term N that we assume is Gaussian with a mean of zero. We can then sum many identical projections to eliminate the noise while improving the signal component.

[7] This turns out to quite important and perhaps not intuitive as pointed out by Hawkes, (2006). Thankfully, there is a linear relationship between the electron density of a real object and the projection that we can obtained (of course there are many caveats and corrections we need to apply, in particular CTF correction which I discuss in **Appendix A**), but at least there is some *linear* link between increasing electron density in the object and the signal we obtain in the projection.

[8] Simplified and hence open to much criticism from not only physicists but anyone even slightly more intelligent than I (which is basically everyone).

[9] usually around 8 e-/Å^2/s

[10] The equation for a generalised Gaussian fuction is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-(x-\mu)^2/(2\sigma)^2} \tag{1.2}$$

σ is the standard deviation, and μ the mean or expected value.

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{n} p \tag{1.4}$$

With this we can use the averaged projection to overcome the inherent sensitivity of biological samples to ionising radiation and increase the signal component of our projection. Real samples are usually a mixture of different projections (difference views of the same protein) as well as different conformations. This means that there needs to be some image classification, to separate the different conformations, and image alignment to sum identical projections[11] (Frank, 2006). If we can correctly classify the different projections, and align them accurately, it can be possible to build the 3D electron density of our target protein.

1.4.1 Reconstructing 3D dimensional structures from EM projections

The principle behind building a 3-dimensional model from a series of projections was first demonstrated mathematically by Johann Radon in 1917 (Radon, 1986), but it was a number of years before it was used to build a model of a real structure. Projections of a 3D object will produce 2D images with varying image intensity. These 2D projections can be converted into Fourier space and provided they come from the same 3D object then they will share a line in 3D Fourier space. If this central line is aligned and there are sufficient projections to 'fill' the 3D Fourier space, then it can be converted to real space and it should produce a model of the original object (**Figure 1.3**). Problems arise if there are not enough projections to completely cover the real object in Fourier space causing information to be missing when the conversion occurs. The first demonstration of this using biological samples and electron microscopy was the 3D reconstruction of the bacteriophage T4 tail. Since the tail is periodic all views (i.e., sufficient projections to 'cover' all the views in Fourier space) were present allowing the 3D model of the tail to be built. Since 1968 there have been many changes to the EM field, but the fundamental principle that allows the reconstruction of high resolution structures has remained the same (Heel et al., 2016).

1.4.2 Making sense of the electron density

After obtaining a 3D electron density from 2D projections of a biological object, the next step is interpreting these data. If the amino acid sequence of the target is known, then it can be possible to fit this into the electron density and build a model of the 'real' structure. The resolution of the electron density is one factor that dictates how easy or difficult it will be to built the atomic model from the sequence of amino acids.

[11] I go into more detail in **Appendix D**

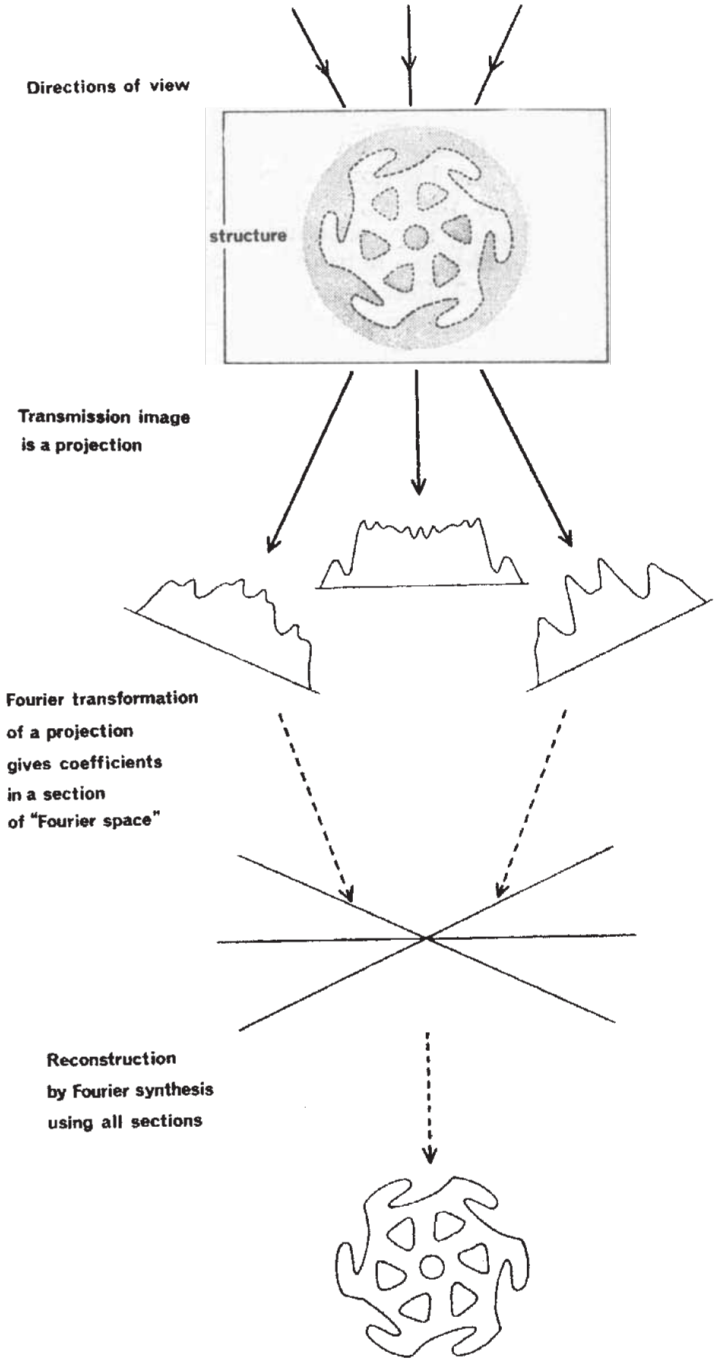


FIGURE 1.3: **3D reconstruction from projections**. Various projections of some real structure are taken, giving different views with varying intensity. When these 2D images are converted to Fourier space they share a central line that can be aligned and used to build a 3D Fourier representation of the object. Converting this 3D Fourier object back to real space gives a representation of the real object built from the 2D projections. Reprinted by permission from Macmillan Publishers Ltd: Nature De Rosier, D. J. & Klug, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature 217, 130–134, copyright (1968)

Since 1968 there has been a large number of technical and computational improvements which greatly improved the resolution we can obtain for some proteins with electron microscopy. Some of the highest resolution structures have even passed below 3 , which is a high enough resolution so that the identify of side chains of a polypeptide can be determined from the electron density alone(Merk et al., 2016). In this case identifying the underlying molecular architecture is much simplified[12] and can be done in a shorter time frame and with more confidence. However, there are a number of potential limitations that can seriously abrogate the obtained resolution including sample heterogeneity, alignment errors, instrument limitations, and numerical errors (Frank, 2006). The fact that the EM field has not overcome these limitations should be evident from the fact that the theoretical minimum protein size that can be determined at atomic resolution is 17-30 kDa (Henderson, 1995) but the current world record is only 93 kDa[13] (Merk et al., 2016).

1) *Sample heterogeneity*. Proteins that are flexible and occupy a wide range of conformational space are difficult targets as each micrograph can contain multiple conformations (which in essence can be considered to be different proteins). In some cases classification can be used to separate the different structures and achieve high resolution (Scheres, 2016) there is a cost in the size of dataset that must be collected. If the protein is small, or the conformational change is not easily identifiable then the maximum resolution obtained can be limited.

2) *Alignment errors*. Alignment errors are a limiting factor in the same manner as proteins with high heterogeneity. Projections that are classified as identical but in reality are from different views[14] will cause a loss in resolution. Alignment errors increase for smaller proteins (Henderson et al., 2011) as the available information to use for alignment decreases. If classification is difficult, either due to sample heterogeneity or small protein size, then alignment errors will be a limiting factor in resolution. Many drug targets are small membrane bound proteins and may be particularly vulnerable to alignment errors limiting the obtained resolution.

3) *Instrument limitations*. There are many technical issues with the available microscopes that might limit the maximum resolution that can be obtained. The higher the accelerating voltage the lower level of inelastic scattering (which correlates with sample damage) that occurs and so higher voltage microscopes can 'use' more electrons per sample area to collect images before unacceptable levels of damage occurs

[12] This is not to say that it is *fast*, building an atomic model even with high resolution data can still take a number of weeks or months.

[13] I used the word 'only' quite liberally here. This is still quite amazing given the technical issues that have had to be overcome to reach this threshold.

[14] of either the same protein or as in the case for proteins with high heterogeneity two different conformations

(Glaeser, 2016). This also has an influence on the CTF which I discuss in more detail in **Appendix A**. As electrons are ionising radiation they also cause heating and movement of the sample (Zheng et al., 2017) which blurs the collected image unless some sophisticated direct electron detectors with a high read speed can be used (McMullan et al., 2014). Finally, to image smaller proteins new phase plates are needed (Danev et al., 2014; Danev et al., 2017) to increase the contrast and allow for correct particle alignment. Clearly there are many aspects of a microscope that need to be optimised for imaging, and without the high end machines the maximum resolution obtained can be limited.

4) *Numerical errors*. Frank, (2006) listed numerical errors as a limiting factor related to the digitisation of photographic plates with a pixel value larger than the the structural information present. This is a problem related to the minimum sample distance necessary to capture a signal (Shannon, 1949). Since direct electron detectors are in common use issues related to scanning photographic film are unlikely to be a limiting factor, however, incorrect binning of data in a similar manner could limit the resolution. Binning is done to reduce the computational load by averaging together some number of pixels (for example, 4x4) to reduce the number of calculations that need to be made. In my experience I have always had access to ample computational resources that I have not had a need to bin any data. Furthermore, if the maximum resolution had a characteristic that was linked with the Nyquist or binned Nyquist I trust that most scientists would realise this is the limiting factor and recalculate the structure with differently binned, or unbinned, data. Given the explosion in computing power I assume that this kind of error is not responsible for many resolution limitations, and if it can be identified it should be trivial, although time consuming, to resolve (rerun the analysis without that level of binning).

There are number of potential sources of limitations and so not all proteins can be determined to atomic resolution. However, even intermediate resolutions provide structural information that can be useful in understanding the function of macromolecular complex (Matoba et al., 2017). If the electron density is not high resolution then identifying the molecular architecture can be quite complicated. One strategy is to use some kind of EM labelling technique that provides additional information about the underlying structure. Although there are many available methods (see **Table 1.3**) the fundamental principle involves the comparison between a labelled and unlabelled (wild-type) structure, and provided something is known about the localisation of the tag then then the domain identity can be inferred (**Figure 1.4**).

1.4.3 Available tools

Conceptually I have grouped each labelling strategy as either 1) domain deletion, 2) domain insertion, 3) native ligand or antibody based, and 4) peptide tags. Each of these strategies have some advantages, and limitations, and so it comes down to the question and target system as to which is the more advantageous to use. Generally, the fewer modifications that are required to the genetic sequence of the target molecule the lower the likelihood that there has been some unwanted structural change, however, this limits the available tags to naturally occurring ligands, or artificially produced antibodies which are not always available or may be low quality (**Section 1.4.6**). **Table 1.2** outlines the general properties of the different tagging strategies as I have defined them.

1.4.4 Domain deletions to identify subunits

The first type of localisation tool is the deletion of a subunit of domain of a complex and comparing this mutant structure to the wild-type to identify the location of the missing feature. It is thereby possible to infer the identity of this feature by its absence from the wild-type complex. This has been used extensively to determine the structure of the components of the cilia and flagella, two eukaryotic organelles that provide motive force or sensory function. They share a common superstructure, the 9+2 axoneme[15], and are present in single cell Eukaryotes to more complex metazoa (Porter and Sale, 2000). There are over 400 axonemal components (Pazour et al., 2005), and mutations in single components can render the cilia or flagella unfunctional (see references in Bui et al., (2008)). The structure and organisation of the 9 + 2 axoneme has been determined in part by comparative studies of various mutants that lack one protein, and can be compared to the wild-type to show the location of the missing subunit (Bui et al., 2008; Heuser et al., 2009; Heuser et al., 2012; Lin et al., 2012; Pigino et al., 2011). Although the use of domain deletions to identify the architecture of the axonome has been used very successfully, take not that the axoneme is a vary large macromolecular complex over 2500 Å in diameter. Such a large complex can afford to have a single component removed and still remain structurally stable, allowing the identification of the missing component. If however, a much smaller complex is the target then removing a subunit may interfere with the correct structural formation and prevent such a comparison being made.

[15]Each microtubule is paired, and nine of these pairs ring a central pair, at which point the naming convention becomes obivous

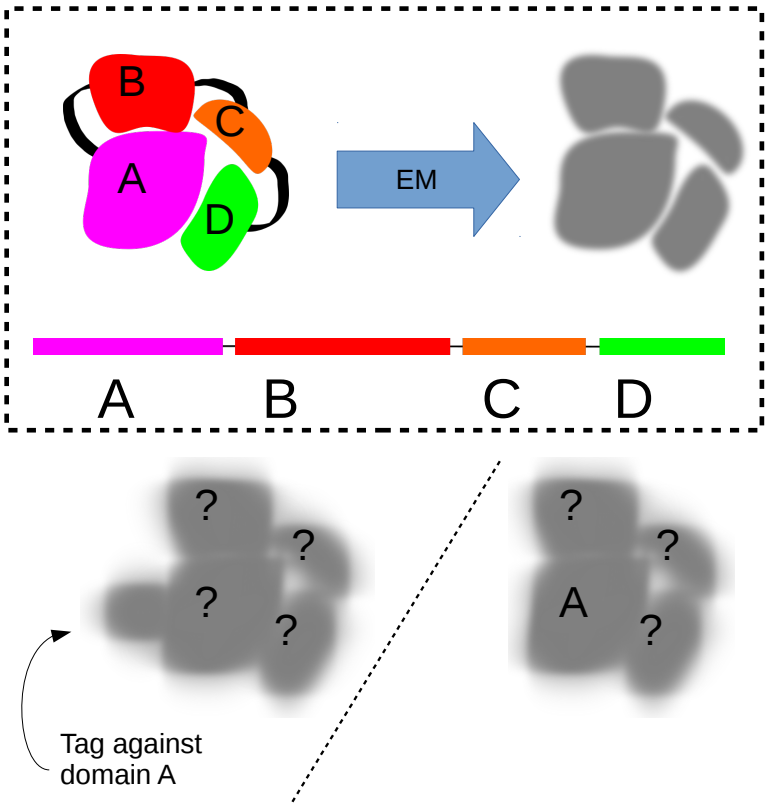


FIGURE 1.4: **EM labelling to identify domains or subunits**. A gene that encodes for a protein with four domains (A, B, C, D) which forms some 'real' structure. If EM can only obtain a low or intermediate representation of this structure then determining the identity of the domains can be difficult. By comparing labelled and unlabelled structures we can infer the identify of the domains if we have some prior knowledge about the localisation behaviour of the tag.

TABLE 1.2: Broad classes of different tagging strategies

Strategy	Potential limitation
Domain deletions	Availability of mutants. Usually limited to stable large complexes
Domain insertions	Insertion into a non-structurally important site
Ligands and Ab	Availability
Peptide tags and Ab	Limited to terminal regions, or long linkers

1.4.5 Domain insertions

In cases where mutants are unavailable, or the removal of component of a target complex will alter the overall structure to an unacceptable level then another strategy is the insertion of an additional domain to use as an EM labelling handle. There are several options that use whole proteins such as maltose-binding protein (Ciferri et al., 2012), green fluorescent protein (Ciferri et al., 2015), or various metallothionein

proteins that conjugate metal to increase EM contrast (Mercogliano and DeRosier, 2007; Nishino et al., 2007). While these methods have been used to identify the structure of various proteins the constitutive expression of the domain tag could interfere with the correct formation of the target macromolecular complex if the inserted domain is in a region that undergoes some necessary conformational change prior to reaching the final correct conformation. Techniques that do not have this limitation involve the timed construction, or addition of the label. The N-terminal insertion of the 87 residue Dynein light chain-interacting domain (DID) tag can be used to construct a visible EM label upon addition of the appropriate protein components (Flemming et al., 2010) thereby minimising the potential structural interference from a large permanently attached domain. Other techniques that also have temporal control over their tagging reactions include various methods based on strategic biotinylation and labelling with streptavidin (Lau et al., 2012; Oda and Kikkawa, 2013), or production of proteins with unnatural amino acids as labelling sites for domain tags (Dambacher and Lander, 2015). All these methods involve the genetic modification of the target protein and may have an experimental 'cost' associated with the insertion of a permanently attached domain such as GFP and MBP, or require various reactions to initiate tagging.

1.4.6 Antibodies and ligands

For some experiments the target protein may be particularly vulnerable to structural perturbation, in which case the insertion of large domains is not a viable strategy for labelling. Instead, native ligands or antibodies for a protein could be used as labels without the requirement for genetic modification. For example, toxins with a known specificity to a protein domain, such as snake (Zingsheim et al., 1982) or scorpion venom (Samsó et al., 1999), have been used successfully to identify domains from intermediate resolution EM maps. These kinds of ligands are effective as labels as they bind semi-irreversibly to their target domain and have sufficient mass that they can be distinguished from the unbound state. Antibodies that have been found to block some functional site can also be used as their binding location should provide localisation about the site that they interfere with (Prasad et al., 1990). Another source are antibodies raised against some portion of target protein, such as a single domain (Kelly et al., 2010), or against a whole subunit of a complex (Boisset et al., 1993; Boisset et al., 1995). In all these cases the tag that is used needs to satisfy two minimum conditions, 1) the affinity and specificity is sufficiently high so a majority of particles have a tag attached, and 2) there is some localisation information known about the binding site (e.g., it blocks a functional site, or targets a specific domain etc). Obviously, it could be quite difficult to identify toxins or find naturally occurring antibodies against all potential targets of interest, and so antibodies can be generated to provide a good source of EM labels.

1.4.7 Limitations to Antibody generation

The development of phage-display libraries has meant that a wide range of antibodies can be generated against a target protein (Fellouse et al., 2007). Bacteriophages can be engineered to express an inserted gene as a protein on their surface so that large libraries can be screened for an interaction between this surface protein and some target. Since the bacteriophage contains the genes that encode for the surface protein, once an interaction has been identified then the phage is sequenced and the causative gene is found. The complementary determining region of an antibody can be expressed in this way, and with the introduction of random mutations large libraries of 'antibodies' can be screened against target peptides. This method can identify antibodies for many peptides but it does not guarantee that the native protein will also show high affinity to identified antibody as the peptide linear sequence in solution may not match the *in situ* structure (Dyson et al., 1988). Conventional antibody production also suffers from this limitation and in some cases can prevent antibody generation[16] (Hancock and O'Reilly, 2005). Due to the potential structural difference between *in situ* and in solution polypeptides when antibodies are found the interaction may be low affinity as the interaction is now based only on the chemical properties of the target sequence as the structure has changed (Van Regenmortel, 1992). Examples from the literature that use generated antibodies as labelling tools have produced nice structures (Wu et al., 2012) but have a large commitment for searching the epitope-paratope space to identify quality antibodies, obviously something that can be costly to do.

1.4.8 Epitope tags

If developing antibodies against a particular target epitope is not possible, then another strategy is to insert an epitope with a proven high affinity antibody. One source of these epitope are the protein purification tags and their conjugate antibodies. Epitope tags are a good source of EM labels as they generally are small in size and show high affinity for their conjugate antibodies. However, as the recognition mode for many antibodies against epitope tags is linear (for example see Roosild et al., (2006)), their placement is often limited to the terminal regions of the target polypeptide. Examples from the literature that used purification tags and their antibodies for EM labelling placed the epitope at the terminal ends of the protein (Kelly et al., 2010; Yip et al., 2010). During protein purificaiton if an epitope tag is inserted into some central domain of a protein they often include additional linker residues(Facey and Kuhn, 2003; Kendall and Senogles, 2006) or are strategically placed inside a long pre-existing loop (Dinculescu et al., 2002; Morlacchi et al., 2012; Smith et al., 2004).

[16] Antibodies may be generated, but their affinity could be predicated on the structure of peptides free in solution that have limited resemblance to the *in situ* protein structure.

This is likely due to the linear recognition mode of the purification antibody, and so if it is inserted into a central domain then the neighboring residues may deform the linear conformation and reduce or abrogate the antibody binding. On the other hand, if these epitopes were inserted with additional linker residues then the proximity of the antibody to the insertion site is reduced, which lowers the accuracy of the localisation information.

1.4.9 Summary of EM tags

For both large and small proteins electron microscopy is becoming a highly popular structural method, however, there are many intrinsic and extrinsic factors that can limit the obtained resolution (Frank, 2006). At intermediate resolutions it can be difficult to determine domain and subunit identity, and so other strategies are required to aid in the mapping of subunits and domains. While there are many methods available each suffers from some potential deficiency. Domain deletion and insertion strategies may not be feasible if the target subunits(s) are structurally important and the recombinant protein loses resemblance to the wild-type structure. Alternatively, ligands that can function as labelling tools are not always available, or can be associated with significant production costs. On the other hand, small epitope tags do have high quality antibodies available, but their insertion site in the target protein has a number of limitations. Overall, a generally applicable EM labelling strategy is not currently available, but could have widespread application given the valuable mechanistic and structural information that intermediate resolution EM structures provides (Matoba et al., 2017; Suzuki et al., 2016).

1.5 Overview and implementation

The unique β-turn conformation of the PA tag allowed it to be used as a mobile epitope against the NZ-1 antibody for various experiments (Fujii et al., 2016b). Given the high level of freedom when selecting an insertion site, and the high affinity NZ-1/PA interaction, it stands to reason that this system could also be a high quality EM label. The PA tag is only 12 residues and so should have minimal impact on the structure of any recombinant protein, and NZ-1 can be added after protein expression and folding so that there is unlikely to interfere with the native-like conformation of the protein. **Figure 1.5** shows the general principle, where a known gene can be modified to include the 12 amino acid PA peptide sequence and then the binding site of the NZ-1 Fab is known. This allows the domain or subunit structure to be determined. Using this strategy depends on having some prior structural knowledge about the target protein so that the PA tag can be inserted into some surface exposed unstructured loop region. Obviously, the more structural information that

TABLE 1.3: A table of various EM localisation tags

Tag name[A]	Size	Binding mode	Temporal control	Location	Ref.
Native ligands	various	various	∇	various	[1]
Domain deletions	various	various	no	various	[2]
MBP	40 KDa (396 res.)	domain	constitutive	N-terminal	[3]
GFP	28 kDa (238 res.)	Domain tag	constitutive	Internal	[3,4]
Metallothionein	6 kDa	Domain tag	constitutive	terminal	[5]
DID	87 residues	assembly tag	yes	N	[6]
FLAG tag	8 residues DYKDDDDK	Ab	yes	terminal	[7]
His-tag	size	Ni-NTA gold cluster	yes	N-terminal	[8]
Gold-conjugated Ab	approx. 50 kDa	Ab	yes	various	[9]
DOLORS	15 residues	monovalent streptavidin	yes	internal	[10]
PA tag	12 residues	Ab	Yes	surface exposed loops	[11]
CBP	26 residues	Ab	yes	C-terminal	[12]

[A]Abbreviations: (MBP) Maltose binding protein, (GFP) Green fluorescent protein, (DID) Dynein light chain-interaction domain, (DOLORS) Domain localization by RCT sampling

ref: [1](Samsó et al., 1999; Zingsheim et al., 1982), [2](Bui et al., 2008), [3](Ciferri et al., 2012), [4](Ciferri et al., 2015), [5](Mercogliano and DeRosier, 2007; Nishino et al., 2007), [6](Flemming et al., 2010), [7](Kelly et al., 2010), [8](Büchel et al., 2001), [9](Mercogliano and DeRosier, 2007), [10](Lau et al., 2012), [11](Brown et al., 2017), [12](Yip et al., 2010)

is available the better informed the placement of the PA tag can be, however, even with limited knowledge, such only having access to the predicted secondary structure based on amino acid sequence, it should still be possible to use the PA tag, but it may require more constructs to find one that has the PA tag placed in an accessible surface region. Although the PA tag has been shown to function as a mobile epitope the evidence that was presented relied on various biochemical characterisations (Fujii et al., 2016b), and so for this thesis I hope to show with convincing EM structural data that the PA tag and NZ-1 Fab can also be used as a mobile EM label, even of very structural heterogeneous and difficult to characterise proteins. What follows is a summary of the structural and biochemical data for the test protein that was selected, $\alpha_{IIb}\beta_3$ integrin.

1.5.1 Integrin

Integrins are a family of adhesion receptors comprising 18 α- and 8 β-subunits that make up at least 24 non-covalently associated heterodimers (Takagi and Springer, 2002). The family of proteins is quite structurally and functionally diverse and they can be associated with a wide range of cell types. Integrins have a very fast ligand binding speed that allows them to modulate adhesion on time scales as short as 1 second. The typical domain organisation of integrin embedded in the cell membrane is shown in **Figure 1.6** along with the striking structural changes that integrin undergoes during activation. Integrin normally forms a bent conformation, but can be extended by the addition of manganese ions in the buffer (Mould et al., 1995) which induces the lower legs of the $\alpha\beta$ subunits to extent and the heterodimer takes on the extended form (Takagi et al., 2002).

Why integrin?

$\alpha_{IIb}\beta_3$ integrin was selected as a test case to demonstrate the use of the PA tag and NZ-1 Fab as an EM labelling handle based on the following criteria: 1) both a high resolution crystal (Zhu et al., 2008) and intermediate resolution EM structure is known for the protein (Eng et al., 2011; Takagi et al., 2002), 2) the PA insertion compatibility has been comprehensively demonstrated with biochemical data (Fujii et al., 2016b), 3) the binding affinity for various PA tag insertion sites is known (Fujii et al., 2016b), and 4) integrin is a difficult protein to image using EM and so a demonstration on a challenging case such as this would provide strong evidence that 'easier' proteins could also use this technology.

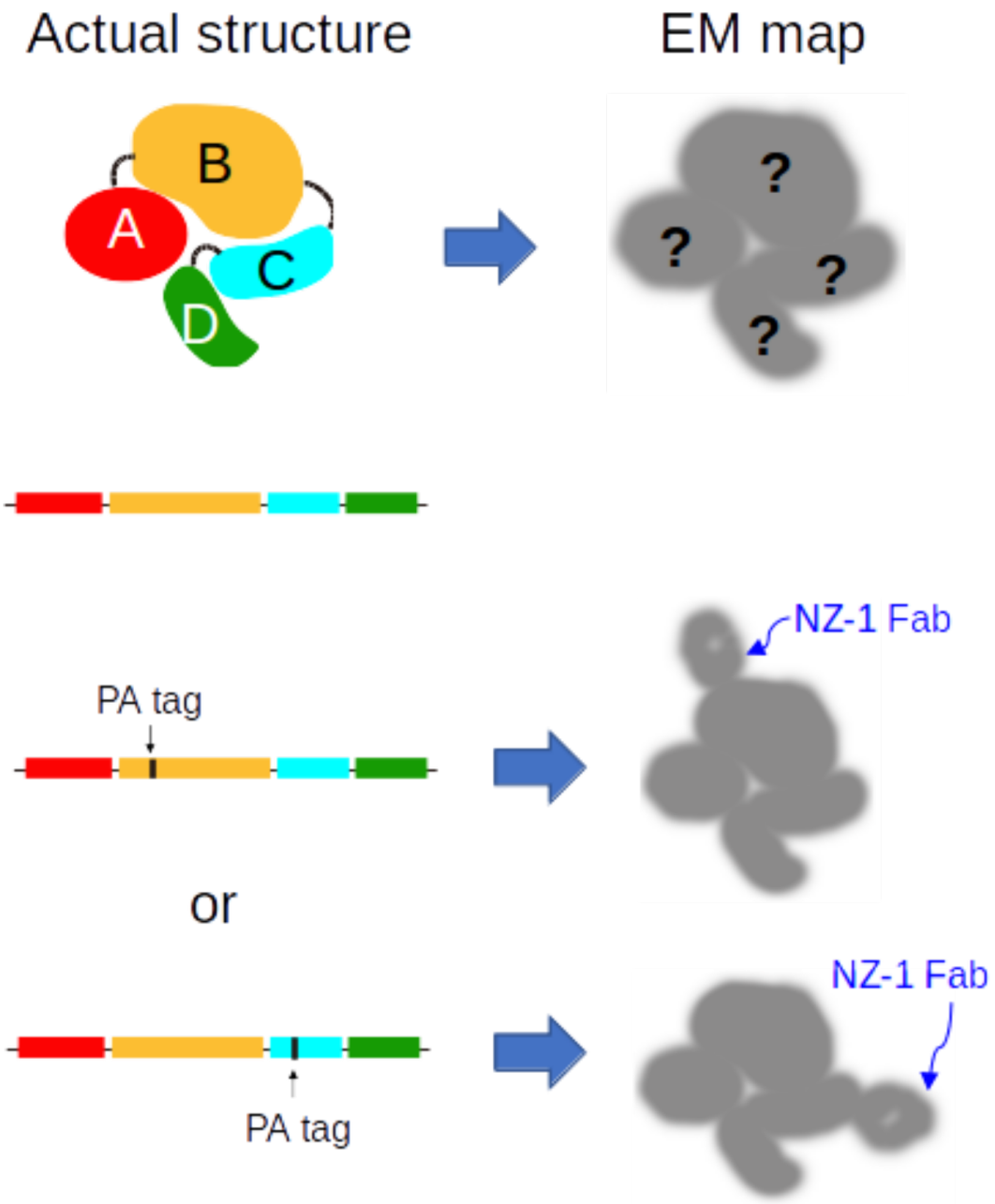
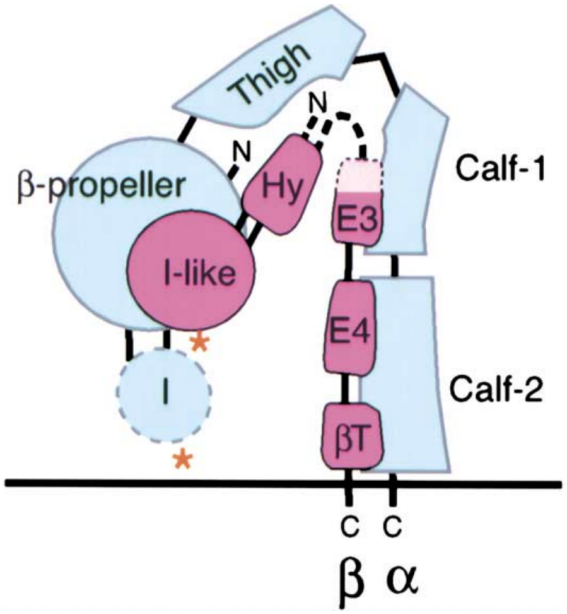
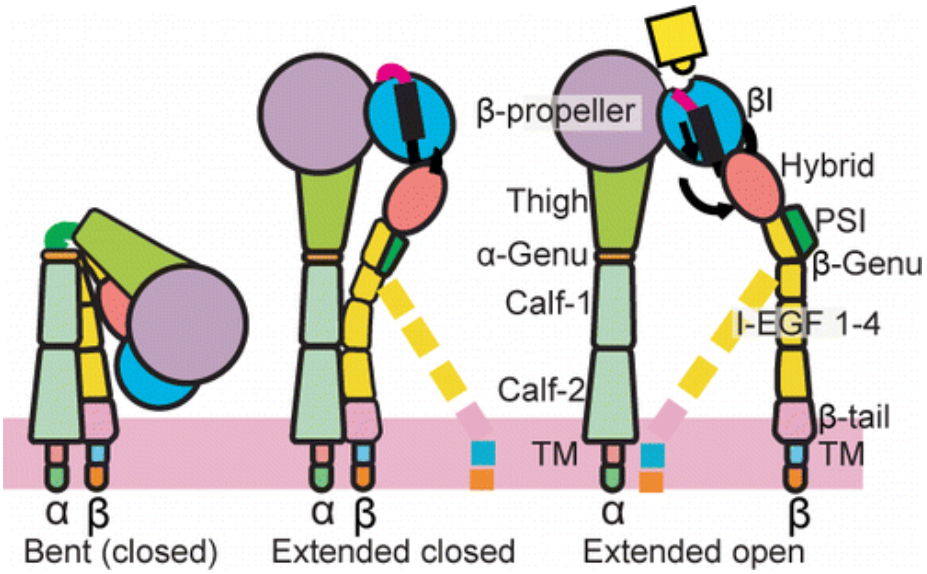


FIGURE 1.5: **PA tag/NZ-1 labelling strategy**. If the genetic sequence of a protein is known but during EM imaging only a low to intermediate resolution representation can be obtained identifying the domain architecture can be difficult. Creating different PA tag containing constructs that can be reacted with NZ-1 Fab to show the binding location could be used to map the various domains or subunits.



(A) Domain organisation of $\alpha V\beta 3$ integrin



(B) Conformational states of integrin

FIGURE 1.6: **Domain organisation of $\alpha V\beta 3$ integrin and conformational changes**. **(1.6a)** This shows the domain organisation of $\alpha V\beta 3$ integrin in the bent conformation. In wild type native integrin the C-terminal regions are embedded in the cell membrane which is shown by a solid horizontal line here. The various domains are as follows: βT, β tail domain, E4 and E3 are EGF domains, Hy is hybrid domain. Asterisk mark the ligand binding sites (the I-domain is absent in $\alpha V\beta 3$ integrin and so is marked with a dashed line here). Design based on the crystal structure from Xiong, (2001). Reprinted from Cell, 95, Takagi, J. et al. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling, 599-611, Copyright (2002), with permission from Elsevier. **(1.6b)** Shows the three main conformational states of integrin. Reprinted with permission from Dong, X. et al. Avβ3 Integrin Crystal Structures and Their Functional Implications. Biochemistry 51, 8814–8828. Copyright (2012) American Chemical Society.

1.6 Summary

Here I will present data that shows the use of the PA tag and NZ-1 Fab as a novel EM labelling handle. Using data from various sources, including gel filtration chromatography, SDS-PAGE, and electron microscopy I demonstrate that the PA tag and NZ-1 can function as a powerful EM labelling system, even on difficult to analyse proteins.

Chapter 2

Methods

2.1 Introduction: General Biochemical Methods

This section will outline the general so called 'biochemical' methods that I used. The design of the protein constructs used. Transfection of Expi293F cells. Purification. Gel filtration. Protein concentration. SDS-PAGE. Preparation of NZ-1.

2.2 Integrin related methods

2.2.1 Integrin constructs

Integrin is a membrane bound heterometic receptor comprised of an α- and β-subunit that are constrained by the cell membrane and so the production of soluble heterodimers requires a careful design strategy that was initially used for $\alpha_V\beta_3$ integrin (Takagi et al., 2002) and applied to $\alpha_{IIb}\beta_3$ integrin by Fujii et al., (2016b). Residues 1-963 of the α_{IIb}-subunit and residues 1-692 of the β_3-subunit were conjugated with ACID or BASE peptides at the C-terminal allowing the formation of a coiled-coil 'Velcro' linker (O'Shea et al., 1993) that constrained the subunits within approximately 40 Å to mimic their native state on the cell surface. Extension PCR was used to generate various PA tag (GVAMPGAEDDVV) containing constructs with the PA peptide inserted into that the W2(residues 136-137), Calf1$_{XZ}$(residues 680-681), or Calf1$_{EF}$(residues 711-712) domains.

2.2.2 Model building of the extended-open integrin

To build the extended-open conformation of integrin the crystal structure of full-length *bent conformation* $\alpha_{IIb}\beta_3$ integrin (PDB: 3FCS) (Zhu et al., 2008) was used. First, the headpiece (above the marked lines in **Figure 3.1**) was aligned with the

TABLE 2.1: Integrin constructs

Name	insertion location
W2	136-137
Calf1$_{XZ}$	680-681
Calf1$_{EF}$	711-712

Domain nomenclature defined by Xiao et al., (2004).

crystal structure of open form $\alpha_{IIb}\beta_3$ integrin (PDB: 1TYE) (Xiao et al., 2004), then the lower legs of both subunits were aligned as rigid bodies to match the overall extended structure seen in other EM investigations of integrin (**Eng2001**; Takagi et al., 2002). Structural manipulations were performed using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schr ödinger, LLC.).

2.2.3 Transformation of E. coli DH5α

Gently thaw *E. coli* DH5α (stored at -80°C) on ice making care not to shock the cells. After spinning down the cells take 50 uL and mix with 0.2 uL of plasmid DNA (should be approximately 50 ng of plasmid DNA). Incubate on ice for 30 minutes and then heat shock for at 42°C for one minute to induce uptake of plasmid DNA. Incubate for 30 minutes at 37°C and then spread on LB plates and incubate at 37°C for 12-16 hours until colony formation. Pick colonies and mix into 3 mL LB media with 1/1000 ampicilin and then incubate at 37°C in a shaker for 12-14 hours. Plasmid DNA was collected from bacterial cells by using the GenElute Plasmid Miniprep (Thermo Fisher Scientific) as per manufacturers instructions.

2.3 Expi293F Expression system

The Expi293F expression system is based on the human 293 cell line. They are a suspension based cell system that can grow to a very high density (5 x 10^6 cells/mL) and achieved very high expression yields of proteins with minimal culture volumes. In the following methods section I will elaborate on the preparation of cell cultures, transfection with plasmid DNA and preparation for protein purification. All culturing is conducted at 37°C in automatic shakers (125 rpm) and 8% CO_2.

2.3.1 Thawing and passaging

Cells were purchased from Thermo Fisher Scientific and provided as 1 mL aliquots stored in liquid nitrogen. Intial cultures were seeded from 1 mL aliquots thawed in a

water bath and then added to 29 mL of Expi293 Expression Medium (Thermo Fisher Scientific) to give a cell density of 0.3 x 10 ^{6} cells/mL. Cells were allowed to grow until density was over 1 x 10^6 cells/mL (usually within 2-4 days after thawing) and passaged only if viability was over 90% (visual assessment). For the first passage cultures were split into 25 mL and a cell density of 0.3 x 10 ^{6} cells/mL. After 3-4 days cells would reach a density of between 3-5 x 10 ^{6} cells/mL and could be transfected and passaged normally. Passage can be continued until over 50 times before cell viability is negatively effected and fresh aliquots of cells would be used to restart the cell culture line.

2.3.2 Transfection

Soluble ectodomain fragments of $\alpha_{IIb}\beta_3$ integrin were transfected to thawed and passaged Expi293F cells prepared as outlined above. Small volumes (usually less than 10 µL) were transferred to a cell counter plate and checked for a density of between 3-5 x 10 ^{6} cells/mL. From cultures with an appropriate density, 75 x 10 ^{6} passaged cells were added to Expi293 Expression Medium (Thermo Fisher Scientific) to make a final volume 25.5 mL (final cell density of 2.9 x 10 ^{6} cells/mL). Plasmid DNA of $\alpha_{IIb}\beta_3$ integrin was prepared by adding 20 µg of α_{IIb} and 20 µg β_3 DNA to 1.5 mL Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific) and incubating for five minutes. Preparation of the propriety transfection reagent is prepared by adding 81 uL ExpiFectamine to 1.5 mL of Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific) and also incubating for 5 minutes before gently mixing the plasmid DNA and transfection medium. The new volume is allowed to incubate for 20 minutes to allow formation of the DNA transfection complex before adding to the cell culture. Enhancers are added to improve the protein yield and was done between 16-22 hours after transfection. 75 hours after transfection the medium was centrifuged (5000 rpm, 5 minutes) to remove most of the cells and then filtered with a 0.22 um PVDF filter (Merck Millipore). To adjust the pH to reduce denaturation during purification 1:50 1 M Tris (pH 8.0) was added to the final volume of media. If purification was not conducted within 24 hours then the media was frozen by submersion in liquid nitrogen and stored at -80°C.

2.4 Preparation of NZ-1 bound Sepharose

This section describes the preparation of the coupled NZ-1 antibody to CNBr-activated Sepharose 4 Fast Flow or CNBr-activated Sepharose 4B (GE Healthcare, Chicago, IL). All reactions are completed at room temperature and the washing steps involved centrifugation at 1500 g for 5 minutes and then removal of the supernatant. Freeze dried sepharose gel (4.3 g) was mixed with 40 mL of 1 mM HCl and rotated for 2

hours until the beads swelled and total volume increased to approximately 20 mL. The supernatant was removed after centrifuging and the beads were washed with 20 mL 1mM HCl three times. The solution was then washed with equilibration buffer [0.1 M $NaHCO_3$, 0.5 M NaCl, pH 8.3] three times to ensure that the beads were at the optimal pH for the coupling reaction. 20 mL of NZ-1 IgG (40 mg) were dialysed against equilibration buffer and then added to the Sepharose beads for a total volume of 40 mL and then rotated for 2.5 hours. The supernatant was removed and checked at 280 nm for the presence of IgG. If the coupling reaction has occured successfully then there should be a very low concentration of IgG compared to the original dialysed sample. The beads were washed equilibration buffer three times and then 20 mL 0.5 M Tris-HCl (pH 8.0) was added to ensure the coupling reaction was stopped. This was rotated at room temperature for 1.5 hours, and the supernatant replaced with 20 mL of 0.5 M Tris-HCl (pH 8.0) rotated for 1 hour. The beads were then washed three times with 20 mL of 0.1 M NaAcO pH 5.2, 0.5 M NaCl and then three times with TBS [20 mM Tris-HCl, 150 mM NaCl, pH 7.5]. Finally the beads were stored at 4°C as 50% slurry in TBS with 0.05% sodium azide.

2.5 Protein Purification

Protein purification was completed either with Ni-NTA beads or NZ-1 bound Sepharose.
Ni-NTA purification: 40 µl Ni-NTA beads per mL (stored as 50% slurry giving final dry volume of 20 µl) of conditioned media were washed with TBS [50 mM Tris-HCl, 150 mM NaCl, pH 8.0] and added to filtered and purified media and rotated for 4 hours at 4°C. Media was poured into a polyprep column and washed with TBS [20 mM imidazole, 50 mM Tris-HCl, 150 mM NaCl, pH 8.0] five times (each wash was twice bead volume). Low concentration of imidazole is used to prevent non-specific binding to the His tag (Terpe, 2003). Elution was completed with 10 washing steps using the bead volume of TBS [300 mM imidazole, 50 mM Tris-HCl, 150 mM NaCl, pH 8.0]. Elutions were pooled and dialysed as described in section 2.6)
NZ-1 Sepharose: 40 µl NZ-1 bound Sepharose per mL (stored as 50% slurry giving final dry volume of 20 µl) of conditioned media were washed with TBS [50 mM Tris-HCl, 150 mM NaCl, pH 8.0] and added to filtered and purified media and rotated for 4 hours at 4°C. Media was poured into a polyprep column and washed with TBS [50 mM Tris-HCl, 150 mM NaCl, pH 8.0] five times (each wash was twice bead volume). 10 elution steps (bead volume) were prepared by incubating PA14 peptide (EGGVAMPGAEDDVV) in TBS [0.1 mg/mL PA14 peptide, 20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 8.0]for 5 minutes. Elutions were pooled and dialysed as described in section 2.6.

2.6 Dialysis of purified protein

Purified protein contained either soluble PA peptide or imidazole and so before gel filtration (section 2.8.2) the sample would need to be dialysed to remove these contaminants, and then concentrated and concurrent buffer exchange to the same buffer that will be used for the SEC. Pooled elutions were dialysed against TBS [20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 8.0] in a Spectra/Por dialysis tubing (Spectrum Labs) with a molecular weight cut off of 6-8 kDa. Dialysis tubing was clipped and placed into 100 times the sample volume TBS (i.e., for 5 mL of pooled elution that would be 500 mL) and gently stirred in the cool room (4°C). Buffer was changed three times every 10-12 hours for a final 1,000,000 fold dilution.

2.7 Concentration and buffer exchange

During dialysis the volume of the purified sample increases due to osmosis and so the sample must be concentrated and the buffer exchanged to the gel filtration buffer (section 2.8.2). Concentration and buffer exchange can be done in a single step using Amicon Ultra ultrafiltration devices (30 kDa MWCO, Merck Millipore). Initially the regenerated cellulose membrane is washed with 2 mL of dialysis buffer (section 2.6) by spinning at 7500 g for five minutes. The flow through is discarded and the dialysed sample added and spun at 7500 g for approximately 5 minutes until 250 µl remains. The gel filtration buffer contains 1 mM $MnCl^{2+}$ to induce the extension of integrin from bent to extended closed (**Figure 1.6b**). Since this is a toxic compound I performed the protein purification without Manganese and then prior to gel filtration I exchanged the buffer with the Manganese containing buffer[1]. 5 mL of gel filtration buffer TBS [20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, pH 8.0] were added and spun down at 7500 g for 5 minutes until 250 µl was obtained. Integrin undergoes conformation changes depending on the buffer conditions and presence of ligands or ligand mimics, as outlined in the introduction. To induce this conformation change I added 60 μM GRGDSPK peptide (Sigma-Aldrich, St. Louis, MO) prior to gel filtration to induce the fully extended conformation.During the labelling experiments 5 fold molar excess of NZ-1 was added to the integrin sample and allowed to incubate on ice for 30 minutes before injection into the gel filtration system. This way contaminants introduced during elution (e.g., imidazloe, free PA14 peptide) were removed by dialysis and the final buffer was exchanged to be identical with the gel filtration.

[1]Ni-NTA purification cannot use a Manganese containing buffer as it interferes with the metal chelating interaction that is used to purify the poly-His containing proteins.

2.8 Size Exclusion Chromatography

2.8.1 Filtering protein sample

Gel filtration columns can be very expensive and so before injecting any protein into them the sample should be filtered to remove any large contaminants that may damage the column. To this end, I filtered the protein samples using an Amicon Ultrafree PVDF 0.22 µm filter (Merck Millipore). The PVDF membrane was washed with TBS [20 mM Tris-HCl, 150 mM NaCl, pH 8.0] by spinning at 12,000 g for 4 minutes before adding the protein sample. Further centrifugation at 12,000 g for four minutes filtered the solution. Finally, to reduced protein losses on the membrane a further 100 µl TBS [20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, pH 8.0] was spun into the protein sample.

2.8.2 Size Exclusion Chromatography

Data for the gel filtration was collected with a Superdex 200 Increase 10/300 GL column (GE Healthcare, Chicago, IL) connected to an AKTA Pure running Unicorn version 6.3 (GE Healthcare, Chicago, IL). The Superdex 200 Increase 10/300 GL column is stored in 20% ethanol solution and so initially the column is washed with 3 x column volume of MQ degassed water and then 3 x column volume TBS [20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, pH 8.0] buffer. Since ethanol is slightly more viscous than water the initial flow rate was 0.250 ml/min, increased to 0.5 ml/min only after the pressure had decreased and stabilised (which occured after the first 15 mL of MQ had flown through the column). Gel filtration buffer TBS [20 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, pH 8.0] was prepared by filtering with a 0.45 µm Durapore Membrane Filters (Merck Millipore) and then degassing for 60 minutes. After 250 µl of protein sample was injected into the column and fractions of 200 µl were collected. A gel filtration standard was run for a single set of experiments within 24 hours of collecting the data without removing or changing the column and with identical buffer (section 2.8.2). Absorbance data was determined with the inbuilt U9-L lamp that measures at a fixed wavelength of 280 nm.

Estimating size for gel filtration data

The estimated volume of elution was calculated based on the *Peak integrate* function of Unicorn (ver:6.3) and the estimated mass of each peak was calculated using a linear model (equation: 2.1) in R (Computing, R Foundation for Statistical Vienna, 2013). A Gel Filtration Standard (Bio-Rad Laboratories, Hercules, CA) with ferritin

TABLE 2.2: Gel filtration protein markers and sizes

Marker protein	Size (kDa)
Thyroglobulin (bovine)	670
ferritin[§]	440
γ-globulin (bovine)	158
Ovalbumin (chicken)	44
Myoglobin (horse)	174
Vitamin B_{12}	1.35

[§]Ferritin was added to the Gel Filtration standard to improve resolution for medium size proteins.

added (Table 2.2) was used as a size standard. An identical volume that as injected into the AKTA pure was mixed with the size standard and injected. The natural log of the known marker size was used as a predictor for elution volume:

$$Elutionvolume = \alpha + \beta \ln mass \tag{2.1}$$

A custom script was used to output the estimated mass of the unknown peaks (see for an example output, Figure 2.1).

2.8.3 Nanodrop

A Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific) was used to measure the concentration of soluble protein fractions. 1.5 μL of buffer identical to the protein containing solution was used as a blank, which was confirmed by reading a buffer sample, then 1.0 μL of protein solution was measured at X wavelength. The following calculation was performed to determine the absorbance intensity of the protein solution.

$$Absorbance = -log\left[\frac{I_{sample}}{I_{blank}}\right] \tag{2.2}$$

The Beer-Lambert equation was used to calculate the exact concentration:

$$A = \epsilon \cdot b \cdot c \tag{2.3}$$

The extinction coefficient (ϵ) was determined using the known genetic sequence and the website ProtParam (http://web.expasy.org/protparam/). The calculated extinction coefficients are shown in Table 2.3, which were used to calculate

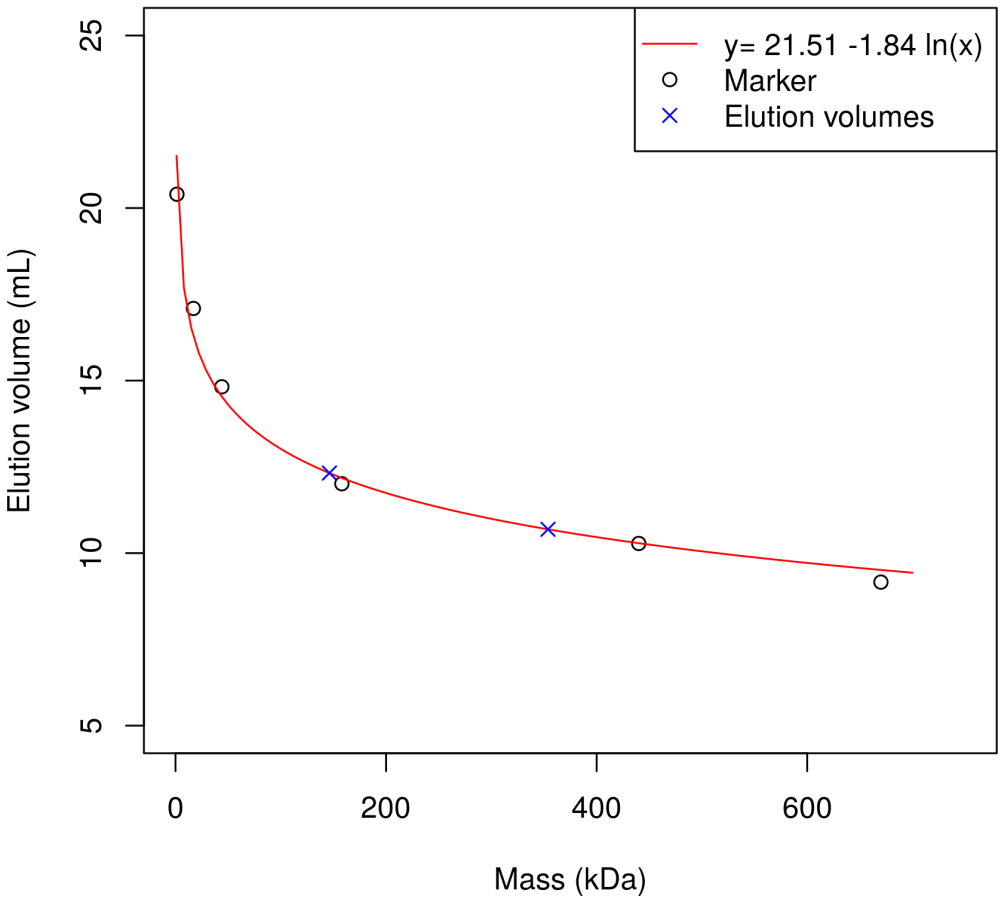


FIGURE 2.1: **Estimated molecular sizes were calculated using the stats package from R and a custom script**. A sample output of this script is show with the linear model for these data, the input marker volumes, and the estimated mass of two example peaks shown.

TABLE 2.3: Estimated extinction coefficients of $\alpha_{IIb}\beta_3$ integrin

Construct	Construct number	non-reduced ϵ	reduced ϵ
Wild type	125 127	1.124	1.099
W2	4326	1.117	1.093
Calf1$_{XZ}$	4334	1.117	1.093
Calf1$_{EF}$	4330	1.117	1.093

the corrected protein concentration.

2.8.4 SDS-PAGE

SDS-PAGE gels were either manually prepared as 7.5% acrylamide or e-PAGEL or c-PAGEL (ATTO Corporation, Tokyo, Japan). Protein fractions were analyzed by SDS-PAGE on 5-20% gradient gels followed by staining with Oriole fluorescent dye (Bio-Rad Laboratories, Hercules, CA) and imaged with a LAS (GE Healthcare, Chicago, IL).

2.8.5 Generating NZ-1 Fab

For EM labelling I used the Fragment antibody binding (Fab) fragment of NZ-1 that I prepared from full length antibody using a process involving digestion with papain, and capturing with PA-peptide bound Sepharose and final purification with size exclusion chromatography (described below). To prepare NZ-1 Fab, 1 mL (2.2 mg/mL) full-length NZ-1 antibody was dialysed in 1x PE buffer () overnight in 10 cm of Spectra/Por dialysis tubing (Spectrum Labs) with a molecular weight cut off of 12-13 kDa. The dialysed NZ-1 was concentrated (Spin X-UF, MWCO: 10 kda) and then added to 0.2 mL of immobilised papain (company) that has been washed with excess PE buffer (20 mM Na/K phosphate, 10 mM EDTA, pH 7.0). 30 µl of 1 M Cys-HCl (pH 7.0) was added and the entire mixture was rotated at 37°C for four hours. To capture the Fab portion 5 mL of PA14 peptide Sepharose was washed with excess TBS (50 mM Tris, 150 mL NaCl, pH 8.0) and filtered into a column. The papain digest was added carefully and the flow through collected, before washing with 9x5 mL TBS (50 mM Tris, 150 mL NaCl, pH 8.0). The bound NZ-1 Fab was eluted using8x5 mL concentrated magnesium buffer (10 mM MES, 3 M $MgCl_2$) and all fractions were measured for protein concentration using the Nanodrop (section). All fractions that contained protein as measured by Nanodrop were pooled and injected into an AKTA prime for size exclusion chromatography with a Superdex 200 16/60 (GE Healthcare, Chicago, IL) column equilibrated with TBS (20 mM, 150 mM

NaCl, pH 7.5) and a flow rate of 0.7 mL/min. Fractions of X were collected and the presence of NZ-1 was confirmed by SDS-PAGE.

2.9 Introduction to EM related methods

In this methods sections I will outline the various techniques that related to the use of the microscopes. Proteins can be imaged either in their native state with low contrast (using cryoEM) or stained with some heavy atoms (such as uranium, tungsten, molybdate etc) to improve contrast but paying a cost in resolution. Samples are prepared on grids made from some conductive material that usually have some carbon membrane to either support the proteins or to provide an adhesive surface. The following section will begin with a description of the preparation of the carbon layers that I used and finish with the various computational analyses.

2.9.1 Carbon Vaporisation

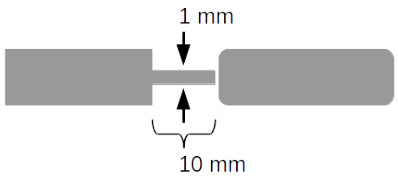


FIGURE 2.2: **Design of carbon rod for vaporisation**. Carbon rods were prepared with the following dimensions using a tabletop grinder.

Carbon was prepared using a JEE-420KM (JEOL Ltd., Tokyo, Japan) carbon evaporator after Bradley, (1954). Graphene rods were ground as shown in **Figure 2.2** with a workbench grinder to be 10 mm in legth and 1 mm wide and set up in the JEE-420KM vacuum chamber as in **figure**. These rods were positioned in a carbon evaporator and a current of 20 mA was applied to cause vaporisation of carbon. Once carbon was evaporating the cover was removed and the carbon began deposition on the mica layer. A 5 mm x 5 mm porcelain tile was used to determine the thickness of the deposited carbon by covering one corner in oil to preent carbon layering that corner and then comparing the color of the remainder of the tile with a standard provided by JEOL until between 5-10 nm was deposited. Once approximately 5-10 nm of carbon were deposited the current was lowered and the heating element was turned off and allowed to cool for 60 minutes before depressurisation and the carbon coated mica was removed and stored at room temperature in preparation for preparing carbon layered grids (**Section 2.9.2**).

2.9.2 Grid Preparation

Copper grids were washed twice with acetone in a sonicator (**brand**) for 10 min before air drying. A water bath containing 10 cm of MQ purified water was prepared with a filter paper as shown in the figure. At least 25 grids were positioned onto a piece of submerged filterpaper (**brand**) using tweezers. The vaporised carbon prepared layering a cut square of mica as described in **Section 2.9.1** was slowly submerged into the waterbath at roughly 45° allowing the carbon layer to separate and float on the water surface. Using tweezers the carbon layer was positioned over the copper girds and the water was drained using a valve. This allowed the carbon to settle over the grids.

2.9.3 Negative Stain

I used negative staining to prepare samples for imaging with electron microscopy. Grids were prepared as outlined in **Section 2.9.2**. Protein purified as in **Section 2.8.2** was diluted with identical buffer to the SEC. Copper grids (**pitch size**) were treated in XX for hydrophilicity for 20 seconds. 5 µl of protein solution were gently pipetted to the surface of the grid and allowed to settle for 1 minute to ensure protein attached to the carbon membrane. After 1 minute the solution was removed using a XXX filter paper and sequentially washed with three drops of 3 µl of ammonium molybdate (hexaammonium heptamolybdate tetrahydrate).

2.9.4 Operation of the Hitachi

Copper grids were stained as outlined in section 2.9.3 and inserted into a Hitachi H-7650 transmission electron microscope (Hitachi High-Technologies, Tokyo, Japan) operating at 200 kV. Images were collected with a 1024 x 1024 CCD camera (TVIPS, Gauting, Germany) at a magnification of 80,000 times corresponding to 3.8 Å/pixel on the specimen. The electron dose was calibrated to 20 $e^-/Å^2$ with a single exposure (i.e., linear data collection) of 1 second. Images recorded with the images were saved as unsigned *tiff* files.

2.9.5 Operation of the JEM 2200-FS

Copper grids were stained as outlined in section 2.9.3 and imaged using a JEM-2200FS (JEOL Ltd., Tokyo, Japan) equipped with a Ω-type energy filter and a field-emission electron gun operated at 200 kV. Zero-loss movies (14 eV) were recorded

TABLE 2.4: Header information stored in the *mrc* file format.

Header name	Value
Number of columns, rows, sections	3710 3838 50
Map mode	2 (32-bit real)
Start cols, rows, sects, grid x,y,z	0 0 0 3710 3838 50
Pixel spacing (Angstroms)	1.941 1.941 1.941
Cell angles	90.000 90.000 90.000
Fast, medium, slow axes	X Y Z
Origin on x,y,z	0.000 0.000 0.000
Minimum density	0.0000
Maximum density	63.505
Mean density	1.4751
tilt angles (original,current)	0.0 0.0 0.0 0.0 0.0 0.0
Space group, extra bytes, IDtype,lens	0 0 0 0

on a K2 Summit direct electron detector (GATAN Inc., Pleasanton, CA) in single-electron counting mode with an underfocus range of 2 to 3 µm at a nominal magnification of 20,000x, corresponding to a pixel size of 1.941 Å on the specimen. Close to focus images may have high frequency information but lack the low frequency that is needed to generate contrast for image alignment (see **Appendix A**). Since $\alpha_{IIb}\beta_3$ integrin is a very small protein I opted to collect data at approximately 3 µm, as this high defocus increased contrast making selecting proteins much easier. Each video had a total accumulated exposure of 80 e-/Å^{-2} fractionated into 50 frames each of 200 ms. An automatic gain correction was applied from the Digital micrograph software. Data was collected as *dm4* and then converted to *mrc* using *dm2mrc* from David Mastronarde's IMOD package. The *mrc* file stores various information from the image collection as part of the header that can be accessed by *header* from the IMOD package. **Table 2.4** shows a sample output. Movies were aligned with MotionCor2 (Zheng et al., 2017) as both whole-frame and 5x5 patch motion correction and CTF parameters estimated with Gctf (Zhang, 2016). A custom script was used to generate graphical representations of the estimated Gctfparameters and used to select resolution and defocus ranges for each data set (**Supplementary Figure 5.2.2**).

2.9.6 Motion Correction

As electrons are highly ionizing radiation that damage the sample and causes heating and motion of the grid (Glaeser, 2016) and so objects image will move during image acquisition. To reduce the motion images are collected initially as videos over a number of frames and then the frames are aligned to improve the signal of each frame using a motion correction program. I used MotionCor2 (Zheng et al., 2017) to align the frames using the settings using the settings shown in **Table 2.5**.

TABLE 2.5: MotionCor2 settings

Setting	Value
Gain	Automatically applied
Mask center	0, 0
Mask size	1.0, 1.0 (full)
Patch	5, 5
Max. iterations	7
Tolerance	0.5 pixel
B-factor	100 $Å^2$
Fourier space binning	1.0 (full)
Initial dose	0
Dose per frame	1.6 e-/$Å^{-2}$
Pixel size	1.941 $Å^2$
kV	200

2.9.7 CTF Estimation

Resolution, information content and CTF parameters (see **Section A**) of EM images can be determined from the Fourier transform of a micrograph, and so the visual checking and parameter estimation of the CTF is an important step. The CTF parameters were estimated using gCTF (Zhang, 2016) with the settings outlined in Table 2.6. The output files were read with a custom script to determine the quality of the collected micrographs as outlines in Section 5.2.2.

2.9.8 Particle Picking

Aligned movies (**Section 2.9.6**) were loaded with *e2boxer.py* from EMAN2.12 (Tang et al., 2007) and particles were selected with a box size of 200 x 200 pixels. Particles were exported from EMAN as IMAGIC particles files with the intensity normalised based on the normalized.edgemean function.

2.9.9 Imagic

Image alignment was performed in the IMAGIC image processing suite (Heel and Keegstra, 1981; Heel et al., 1996). Picked particles (**Section 2.9.8**) were appended into a master file containing all particles. All particles were band-pass filtered to retain frequencies between $\frac{1}{275}$ Å and $\frac{1}{10}$ Åwith a soft Gaussian mask of 0.65 applied. The drop-off parameter for the Gaussian mask was 0.1, which is the halfwidth of the Gaussian drop-off that is used to smooth the edge of the mask to prevent spurious correlations between mask edges. The pixel intensity was normalised with a new

TABLE 2.6: Settings used to estimate the CTF parameters using Gctf.

Setting	**Value**	**Unit**
Pixel size	1.941	Å
FFT box size	512	pixels
C_s	2.0	mm
Astigmatism	100	Å
Maximum defocus	50000	Å
Minimum defocus	5000	Å
Defocus step	500	Å
Voltage	200	kV
Maximum phase shift	180	°
Minimum phase shift	0	°
Phase shift step	10	°
Amplitude contrast	0.1	
Maximum resolution	3	Å
Minimum resolution	30	Å

σ=1.0. Particles were sorted based on highest intensity values to check for any errors (such as dust on the direct detector that causes a direct bright pixel) and removed.

The initial alignment step is to bring the particles to roughly the same starting point. This does not align them rotationally but just uses a total sum of all pixels over all particles and a Cross-Correlation Function (**Section D**) to determine if the alignment is accurate. This was iterated five times to produce a rough alignment that was constrained to 0.2 (or 40 pixels) of the initial frame. Class averages were prepared by multivariate statistical analysis (Heel, 1984) and hierarchical ascendant classification (Ward, 1963) to generate averages and then iterative iterative multi-reference alignments (Frank, 2006).

Chapter 3

Results

3.1 Overview of results

I was able to successfully purify and image both wild type and PA tag containing integrin to show that:

1. 2D averages of purified $\alpha_{IIb}\beta_3$ integrin can be resolved to domain level providing a useful target for testing the utility of PA tag NZ-1 Fab labelling on small, conformationally diverse proteins (**Figure 3.11**).

2. The PA tag containing integrin (W2, Calf1$_{XZ}$, Calf1$_{EF}$) did not show altered hydrodynamics compared with the wild type construct as seen by their identical elution profile and predicted molecular mass during gel filtration chromatography (**Figure 3.2-3.5**).

3. Upon addition of NZ-1 Fab the elution peak remained monodisperse and increased in elution mass for W2 and Calf1$_{EF}$, while Calf1$_{XZ}$ showed no change which corresponds with the predicted binding location of the NZ-1 as internal to the 'legs' of the integrin heterodimer (**Figure 3.2-3.5**).

4. All constructs formed a complex with NZ-1 Fab that could be confirmed by SDS-PAGE (**Figure 3.2-3.5**).

5. Raw whole field micrographs showed that PA tag insertion did not disrupt global conformation as these integrins had a similar conformation to the wild type construct (**Figure 3.6-3.9**).

6. 2D averages (**Figure 3.11**) of the wild type and integrin constructs confirmed that:

   (a) The design strategy produces native-like reconstituted integrin.

   (b) The PA tag insertion does not disrupt the global conformation.

(c) An additional density is present that corresponds to the predicted NZ-1 Fab binding location based on the PA tag insertion site.

7. Overall this shows that the PA tag and NZ-1 Fab can be used as a marker for subunits or domains even for small and heterogeneous proteins.

3.2 EM imaging of wild-type integrin

To demonstrate the efficacy of the PA tag and NZ-1 as a novel EM label I selected the type I glycoprotein $\alpha_{IIb}\beta_3$ integrin and produced soluble ectodomain fragments comprised of residues 1-963 of the α_{IIb}-subunit and residues 1-692 of the β_3-subunit in Expi293F cells. This corresponds with the ectodomain fragment of the heterodimer and was based on an established design (Takagi et al., 2002) that uses a disulphide-linked coiled-coil 'velcro' linker (O'Shea et al., 1993) to mimic the native ectodomain heterodimer (Takagi et al., 2002). $\alpha_{IIb}\beta_3$ integrin undergoes conformational changes between a bent and highly compact conformation in the resting state to an extended conformation when activated. Extension can be induced by addition of manganese ions (Mould et al., 1995) which causes the lower portion of the legs to extend (Takagi et al., 2002). Binding of ligands that contain Arginine-Glycine-Aspartate (RGD) sequences complete the extension of integrin by opening the upper legs (Zhu et al., 2008). The extended-open conformation has the largest surface area for PA tag insertion and NZ-1 fab binding, making this conformation ideal for demonstrating the labelling efficacy of the PA tag and NZ-1. However, the highest resolution model of $\alpha_{IIb}\beta_3$ integrin is not in the open-extended conformation, rather it is a bent-compact crystal structure (Zhu et al., 2008). By deforming this crystal structure to match the extended-open conformation (**Figure 3.1**) this exposed the largest surface area of $\alpha_{IIb}\beta_3$ integrin to best demonstrate the labelling of PA tag containing constructs with NZ-1 Fab. Furthermore, the extended-open conformation of integrin is expected to produce more uniform projections during EM imaging as attachment to the carbon film should be more uniform.

Given the well characterised biochemistry of integrin I could induce the extended-open conformation by incubating the protein sample with 1 mM Mn^{2+} and 60 μM RGD peptide for 30 minutes prior to size exclusion chromatography (SEC) with a Superdex 200 column equilibrated with 1 mM Mn^{2+} in Tris-buffered saline (TBS). The majority of the expressed and purified wild-type integrin eluted as a single monodisperse peak with an estimated molecular size of 393 kDa (**Figure 3.2**). The presence of a single disulphide-linked band at approximately 250 kDa that decomposed under reducing conditions to three bands corresponding to the α_{IIb} heavy chain (110 kDa), β_3 subunit (90 kDa) and the α_{IIb} light chain (20 kDa) confirmed the presence of integrin as the sole protein (**Figure 3.2**, inset). Fractions collected

from SEC were diluted in identical buffer (detailed in **Appendix 5.1.1**) and placed on glow-discharge treated carbon-film layered copper grids and stained with ammonium molybdate for imaging with TEM. Whole field micrographs showed that α_{IIb} integrin formed the predicted extended-open conformation characterised by a ring-like projections (**Figure 3.6**). From these micrographs over 2000 particles (**Figure 3.10**) were selected and after CTF estimation and correction I used multivariate-statistical analysis to generate 2D classes (**Figure 3.11**). The 2D classes closely resembled both the previously published structures (Takagi et al., 2002) and the predicted atomic model of extended-open integrin created using the crystal structure of $\alpha_{IIb}\beta_3$ integrin (**Figure 3.1**). In particular, all four domains (β-propeller, thigh, calf1, calf2) could be resolved in most of the class averages (**Figure 3.11**). The design strategy I used originality devised by Takagi et al., (2002) preserves the 'natural' flexibility native $\alpha_{IIb}\beta_3$ integrin has on the cell surface and only constrains the subunits to within approximately 40 Å of each other. This can be seen in reduced density of the lower portion of the β_3 subunit in some of the averages which suggests either that the β-subunit has more intrinsic variability and so is averaged away during alignment.

3.3 PA tag inserted integrin and Fab labelling

After obtaining the SEC elution profile and confirming the quality of 2D averages of wild-type $\alpha_{IIb}\beta_3$ integrin to domain resolution I then aimed to demonstrate the utility of labelling various sites with PA tag. Fujii et al., (2016b) had previously shown that up to eight different sites on $\alpha_{IIb}\beta_3$ integrin can accept the PA tag insertion and remain reactive with NZ-1. Since I wanted to show that NZ-1 Fab can be used as a label, I selected sites based on the following criteria. 1) they had high affinity to NZ-1 (Fujii et al., 2014), and 2) they were on the side of the α_{IIb} subunit when integrin is in the extended-open conformation (**Figure 1.6b**) so that projections and 2D averages would show the Fab adjacent to the integrin density (c.f., overlapping the subunits and being hidden). The three sites with surface exposed unstructured loops that fulfilled these criteria were in the W2 domain (residues 136-137), and Calf1$_{XZ}$ (residues 680-681) and Calf1$_{EF}$ domains (residues 711-712). I expressed these constructs in Expi293F cells and purified them using NZ-1 bound sepharose (Fujii et al., 2014). After dialysis, concentration, and treatment with 1 mM Mn^{2+} and 60 μM RGD peptide these constructs were were subjected to SEC purification and showed elution profiles that closely matched the wild-type integrin in both peak outline and estimated molecular mass (W2: 403 kDa, Calf1$_{XZ}$: 398 kDa, Calf1$_{EF}$: 396 kDa, **Figure 3.3-3.5**, dotted lines). The slight difference in estimated mass compared with the wild-type construct can be accounted for by the different insertion locations of the PA tag as either being externally (W2) or internally (Calf1$_{XZ}$, Calf1$_{EF}$) oriented relative to the

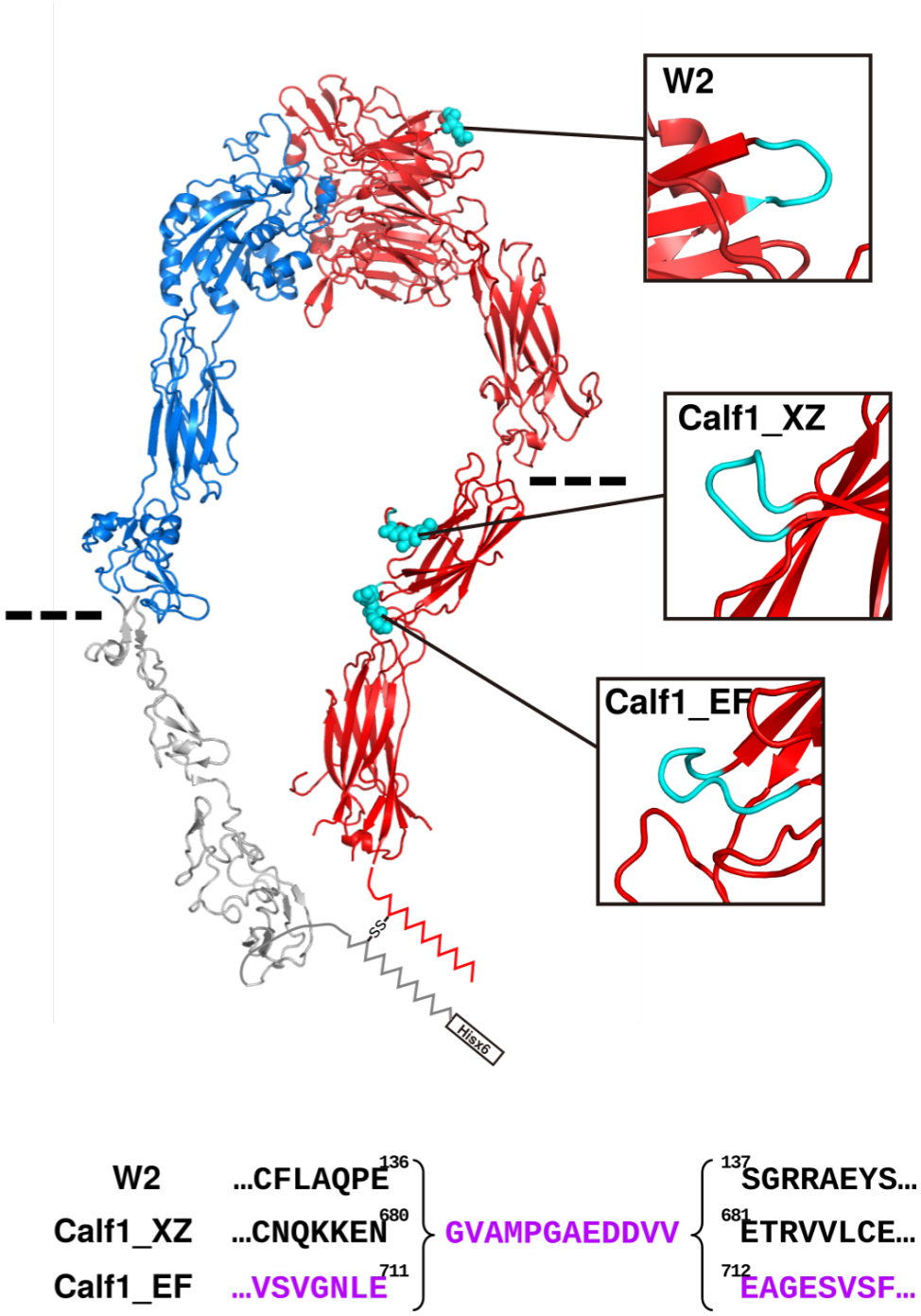


FIGURE 3.1: **Structural model of $\alpha_{IIb}\beta_3$ integrin and PA tag insertion sites**. This model was created using the crystal structure (PDB: 3FCS) as detailed in **Section 2.2.2** and shows the α_{IIb} subunit (red) and β_3 subunit (blue and gray) as cartoon representations. The lower portion of the β_3 subunit is marked in gray as it is highly flexible compared to the other domains of this protein. The coiled-coil 'velcro' linker as the C-terminal of each subunit is shown schematically with the disulphide bond marked, and Hisx6 tag marked. The domain nomenclature defined by (Fujii et al., 2016b; Xiao et al., 2004) used to identify the PA tag constructs is shown in the expanded view, with the insertion locations marked as either cyan spheres or loops. The amino acid sequence of each construct is shown at the bottom with the tag inserted between the primary sequence of the α_{IIb}-subunit.

centre of the heterodimer (**Figure 3.1**). An additional 12 residues at the externally oriented W2 site contributes to an increase in the hydrodynamic radius of the protein, whereas the Calf1$_{XZ}$ and Calf1$_{EF}$ sites are internally positioned on the α_{IIb}-subunit and oriented towards the central gap between the two legs of the heterodimer, thus causing a much smaller increase in hydrodynamic radius. These data along with the already discussed 2D averages confirm the accuracy of the deformed crystal structure. SDS-PAGE also confirmed the presence of integrin that closely matched the wild-type construct (**Supplementary Figure 5.1**).

Exogenously added tags that are used for domain level localisation during EM must form stable complexes with their targets to minimise the data collection requirements[1]. To test the stability and occupancy of the PA tag insertion site by NZ-1 I incubated the various constructs in excess NZ-1 Fab and then subjected the complex to SEC. The W2 integrin showed a uniform shift from the 403 kDa to a higher 455 kDa position as well as the presence of a new peak at 42 kDa (**Figure 3.3**, solid line). The identity of the 455 kDa peak (peak C) is likely the W2 integrin-NZ-1 Fab complex (403 kDa integrin and 42 kDa NZ-1 Fab) and the later 42 kda peak (peak F) the excess unbound NZ-1. This was indeed confirmed using SDS-PAGE, peak C contained both integrin bands and NZ-1 Fab heavy and light chains whereas peak F only contained Fab heavy and light chains (**Figure 3.3**, inset). The profile shift to a higher mass of the integrin-NZ-1 complex as a single monodisperse peak is very important as it suggests that there is high occupancy of the PA tag site, a property that is necessary for effective EM labelling technologies as outlined above. The other PA tag containing constructs showed changed to their gel filtration elution profile that should be interpreted in conjuction with the SDS-PAGE results and the predicted binding site of the NZ-1 Fab. While Calf1$_{XZ}$ integrin showed no increase (remaining at 398 kDa) in mass upon binding with NZ-1 the SDS-PAGE results (**Figure 3.4**, inset) confirm the presence of NZ-1 Fab in the integrin containing peak (peak C). The lack of a peak shift is not anomalous when considered with the predicted insertion site of the PA tag on the inside of the α_{IIb} subunit leg and oriented internally. Presumably the NZ-1 Fab is 'buried' within the legs of the heterodimer and the binding of NZ-1 does not increase the hydrodynamic radius of the complex. Calf1$_{EF}$ integrin does show a small increase in mass upon complexing with NZ-1 Fab (changing from 396 kDa to 432 kDa) and the presence of NZ-1 Fab can be confirmed by SDS-PAGE (**Figure 3.9**). Taken together, these results show that the addition of NZ-1 Fab to PA tag containing proteins allows the formation of a stable complex in a stoichiometric manner, and importantly this can be done in a temporally controlled manner.

[1]Labelling systems that don't have high occupancy also have another potential drawback. If partially labelled proteins are not be easily distinguishable from unlabelled proteins, and they have a different conformation they may be incorrectly averaged together, which will limit the obtained resolution.

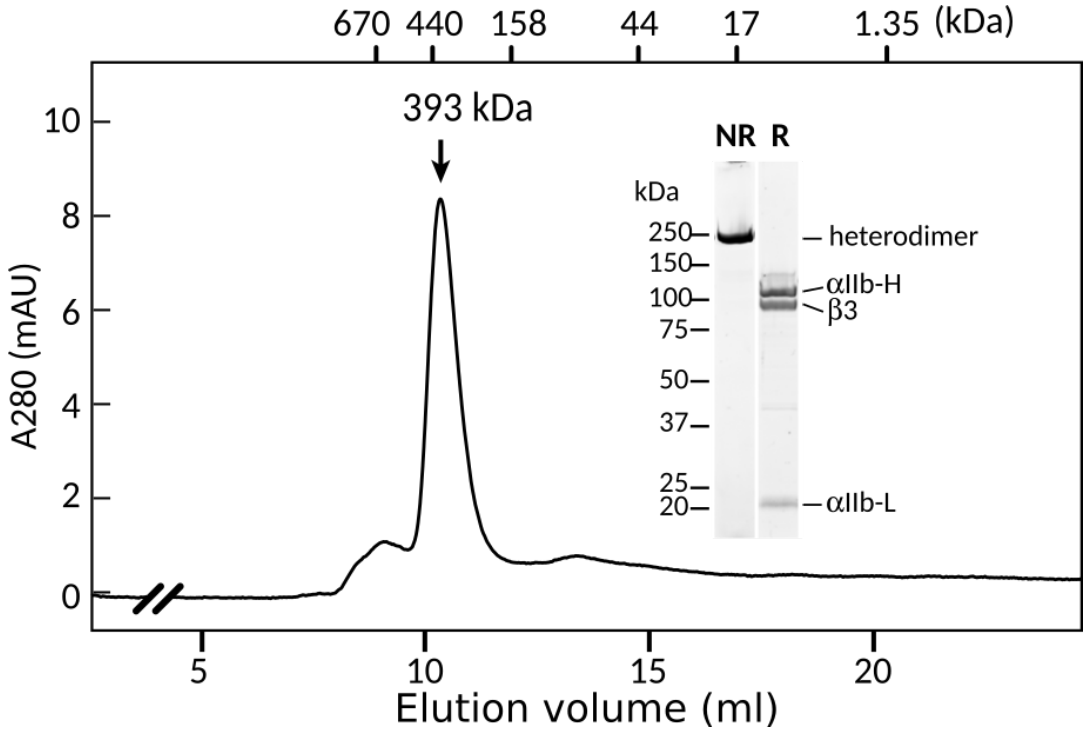


FIGURE 3.2: **Gel filtration of wild-type $\alpha_{IIb}\beta_3$ integrin**. Estimated molecular mass was calculated from a protein size standard (shown at top). SDS-PAGE gel of the peak fraction under non-reducing (NR) and reducing (R) conditions is shown in *inset*, with marker size and predicted band identity shown. Estimated molecular mass was calculated from a protein size standard (shown at top) that included thyroglobulin, ferritin, γ-globulin, ovalbumin, myoglobin, and vitamin B12.

3.4 EM imaging of Fab labelled integrin

I collected the peak containing the integrin-Fab complex (peak C in **Figure 3.3-3.5**) and prepared them for TEM imaging as with the wild-type protein. Raw micrographs of the PA tag containing constructs show features similar to the wild-type protein with both subunits of integrin heterodimer visible (**Figure 3.7-3.9**). Importantly, many of the particles in the raw micrographs showed an additional density corresponding with the predicted binding location of the NZ-1 Fab. Particles were picked for each of the constructs (examples shown in **Figure 3.10**) and subjected to alignment and 2D averaging (**Figure 3.11**). A majority of the classes had a density that corresponded with the predicted binding location of the NZ-1 Fab (**Figure 3.11**, comparison with right panel). While the W2 integrin had an additional density at the 'head' of the heterodimer, the Calf1$_{XZ}$ and Calf1$_{EF}$ integrin constructs showed the Fab density between the α_{IIb}-, and β_3 subunits. The central location of the Fab for these two constructs supports the predicted hydrodynamics and explains the minimal change in elution mass during gel filtration chromatography (i.e., the Fab is

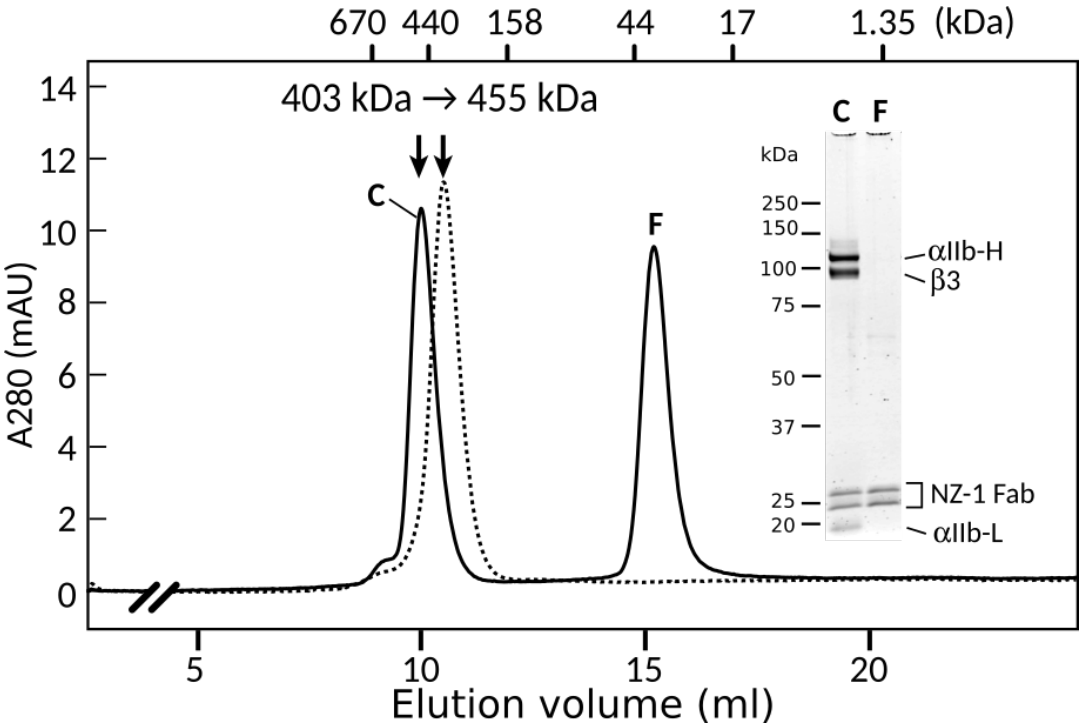


FIGURE 3.3: **Gel filtration of W2 $\alpha_{IIb}\beta_3$ integrin**. Integrin-Fab complex (peak C) excess Fab (peak F) are marked, with SDS-PAGE gel of these fractions shown under reducing conditions with estimated band identity marked (*inset*). Gel filtration of integrin prior to Fab addition is also shown as dashed lines. Estimated molecular mass was calculated from a protein size standard (shown at top) that included thyroglobulin, ferritin, γ-globulin, ovalbumin, myoglobin, and vitamin B12, with the change in estimated mass between *apo* and complex marked.

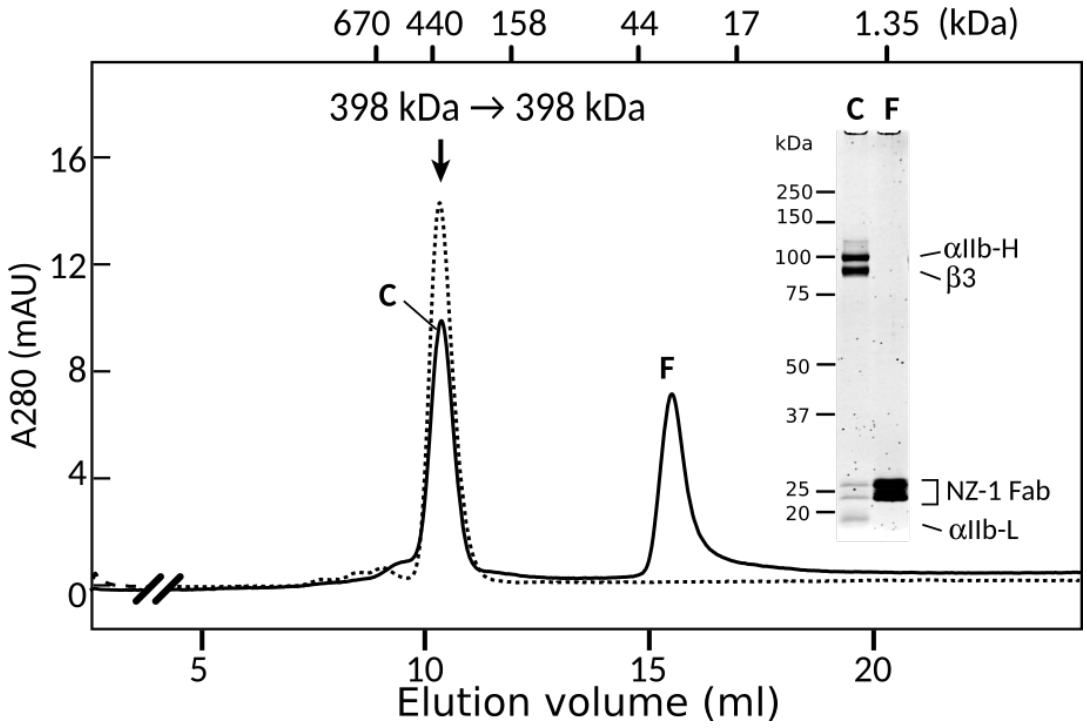


FIGURE 3.4: **Gel filtration of Calf1$_{XZ}$ $\alpha_{IIb}\beta_3$ integrin**. Integrin-Fab complex (peak C) excess Fab (peak F) are marked, with SDS-PAGE gel of these fractions shown under reducing conditions with estimated band identity marked (*inset*). Gel filtration of integrin prior to Fab addition is also shown as dashed lines. Estimated molecular mass was calculated from a protein size standard (shown at top) that included thyroglobulin, ferritin, γ-globulin, ovalbumin, myoglobin, and vitamin B12, with the change in estimated mass between *apo* and complex marked.

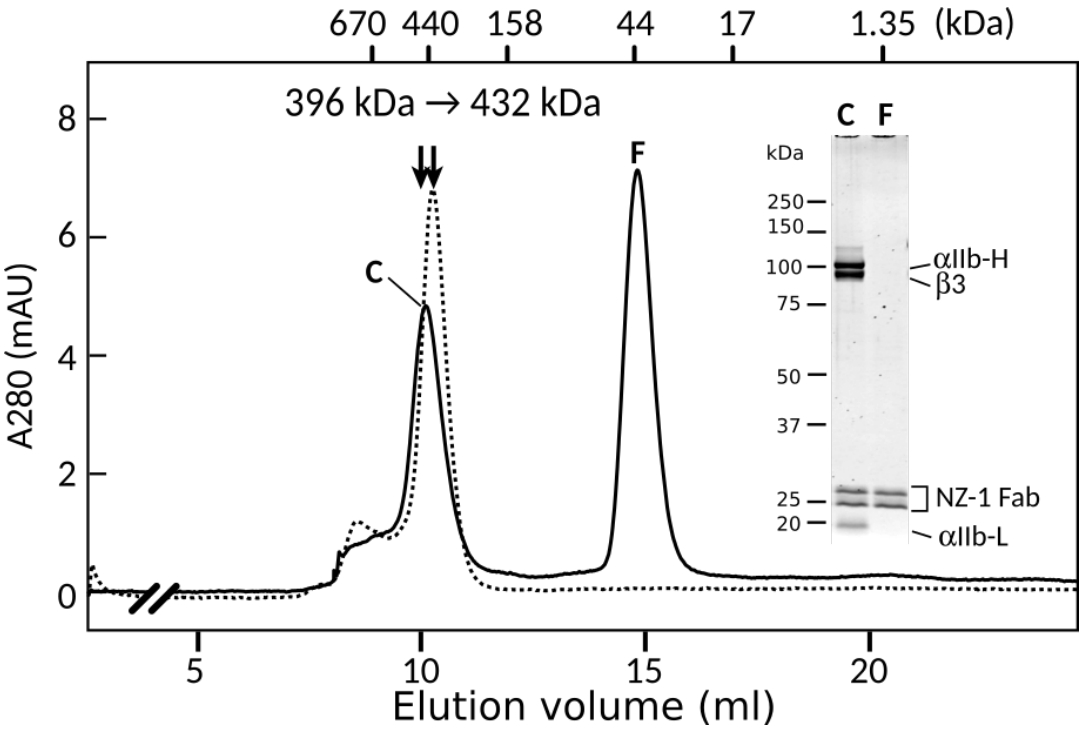


FIGURE 3.5: **Gel filtration of Calf1$_{EF}$ $\alpha_{IIb}\beta_3$ integrin**. Integrin-Fab complex (peak C) excess Fab (peak F) are marked, with SDS-PAGE gel of these fractions shown under reducing conditions with estimated band identity marked (*inset*). Gel filtration of integrin prior to Fab addition is also shown as dashed lines. Estimated molecular mass was calculated from a protein size standard (shown at top) that included thyroglobulin, ferritin, γ-globulin, ovalbumin, myoglobin, and vitamin B12, with the change in estimated mass between *apo* and complex marked.

TABLE 3.1: Prevalence of Fab-bound integrin estimated from class averages

Mutant	insertion location	total particles[1]	number of class averages with Fab[2]	Fab-bound particles[3]	prevalence (%)
Wild-type	–	2407	–	–	–
W2	β-hairpin	1700	7/12	961	57
Calf1$_{XZ}$	β-hairpin	3126	9/12	2822	90
Calf1$_{EF}$	inter-sheet loop	3311	7/12	1836	55

[1]Total picked particles.
[2]Number of classes with visible Fab density.
[3]Number of particles contained within Fab classes.

hidden between the subunits and doesn't increase the hydrodynamic radius of the protein).

While a majority of the 2D classes and many of the particles in the raw micrographs showed a density that clearly corresponded with bound NZ-1 Fab there were some that lacked such a density. In an attempt to estimate the prevalence of Fab binding I compared the total particles in classes that had a Fab density with the total of all picked particles. This showed that classes with a visible Fab density accounted for 55-90% of the total particles for the PA tag containing constructs (**Table 3.1**). Although 100% occupancy would be ideal the occupancy that is seen here is not unreasonable due to several factors that may result in a decreased number of identifiable Fab. In particular during alignment particles may become misaligned and cause the Fab density to be averaged away. Another causes may be dissociation during the negative staining process[2], and perhaps use of other staining reagents (or sample preparation techniques) may improve this aspect. Regardless, these data show that the PA tag can function as a novel 12-reside tag that can be inserted into surface exposed loops and bind with high affinity to NZ-1 Fab. The resulting complex is stable during purification and the NZ-1 Fab can be identified in both raw micrographs and 2D averages, showing that this method is applicable for labelling subunits or domains or small and highly flexible proteins.

3.5 Summary

I was able to successfully purify and image wild-type $\alpha_{IIb}\beta_3$ integrin to show that 2D averages of purified $\alpha_{IIb}\beta_3$ integrin can be resolved to domain level providing a

[2]Ammonium molybdate can be a relatively acidic stain (pH 6) and perhaps this contributed to some of the dissociation. This lab does use other stains (such as those based on tungsten or uranium) but in my experience they suffer from reduced contrast or have very complicated waste management procedures and given the small and highly flexible nature of integrin I opted to use ammonium molybdate

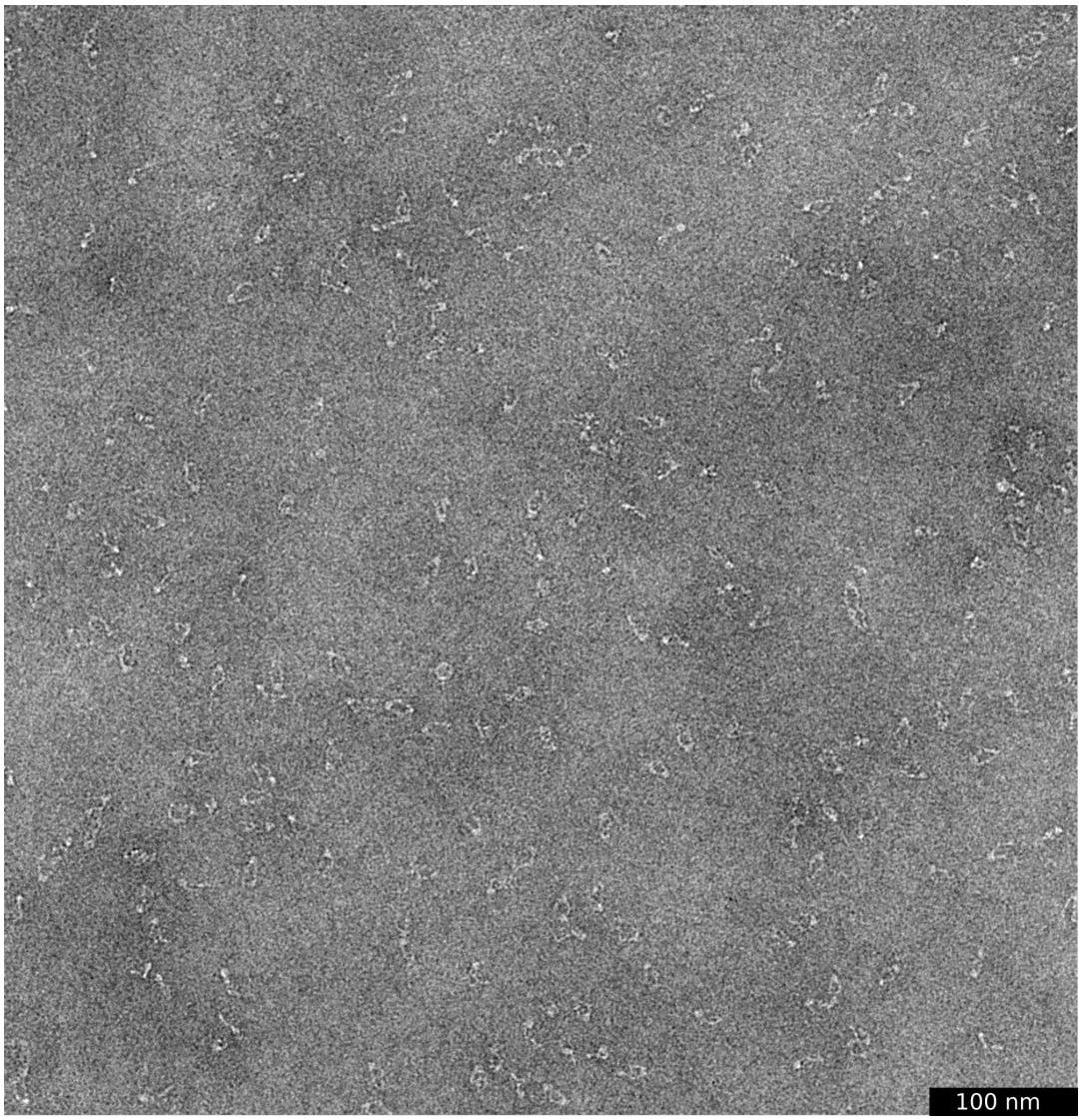


FIGURE 3.6: **Micrograph of purified wild-type $\alpha_{IIb}\beta_3$ integrin**. Representative image collected using a JEM-2200FS. Image has been low-pass filtered with a 0.05 (20 Å) cut-off. Scale bar is 100 nm.

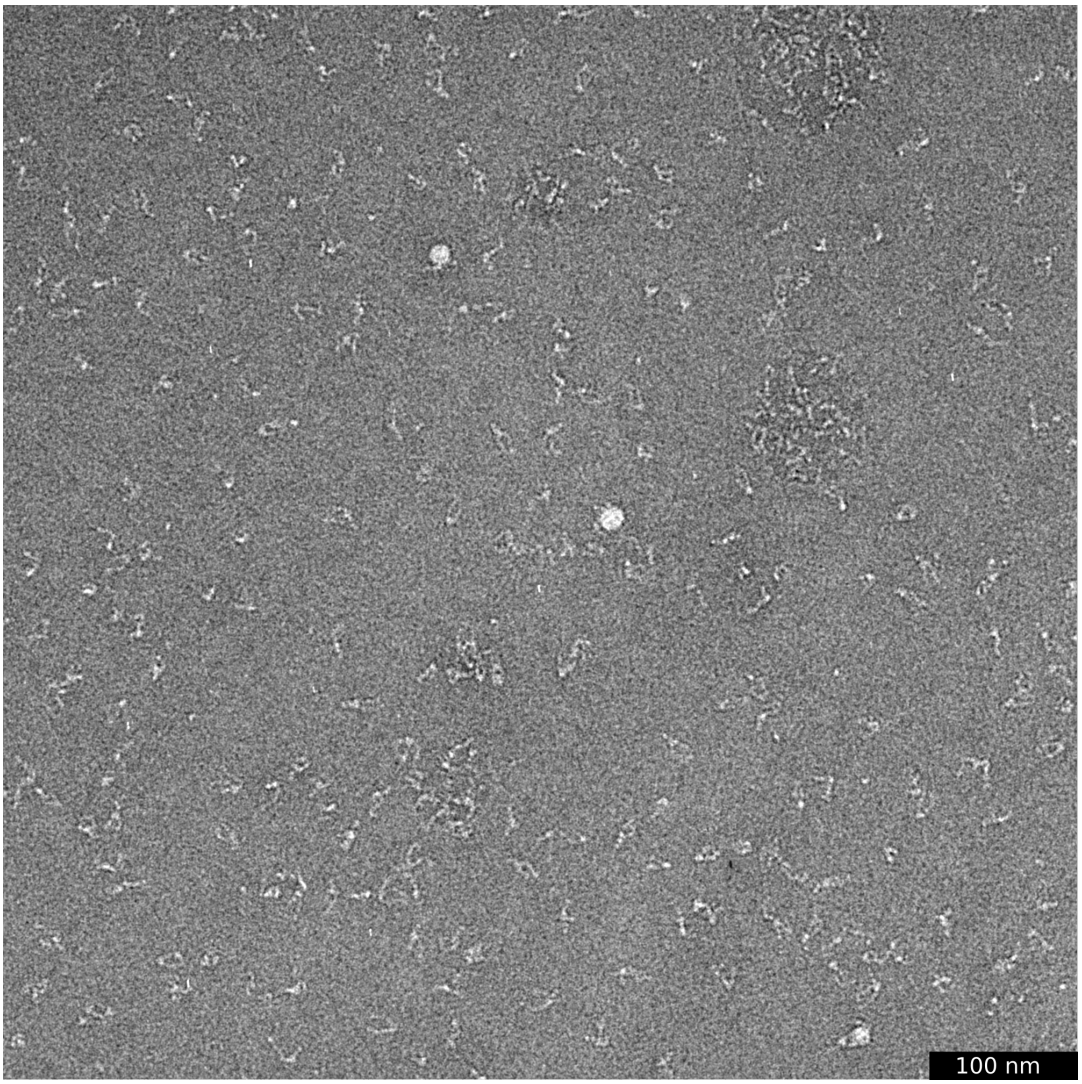


FIGURE 3.7: **Micrograph of SEC purified W2$\alpha_{IIb}\beta_3$ integrin**. Representative image collected using a JEM-2200FS. Image has been low-pass filtered with a 0.1 (10 Å) cut-off. Scale bar is 100 nm.

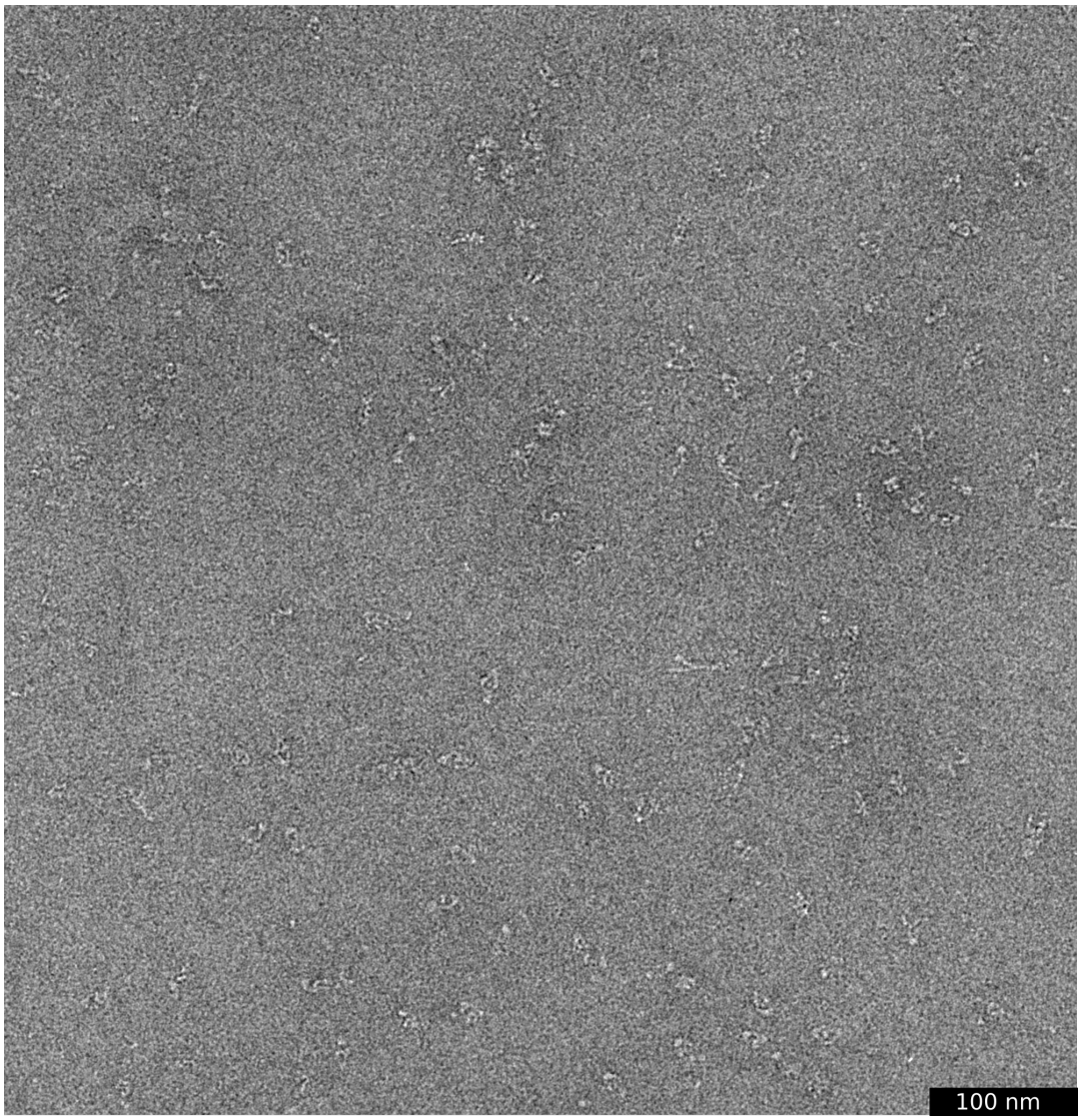


FIGURE 3.8: **Micrograph of SEC purified Calf1$_{XZ}\alpha_{IIb}\beta_3$ integrin**. Representative image collected using a JEM-2200FS. Image has been low-pass filtered with a 0.1 (10 Å) cut-off. Scale bar is 100 nm.

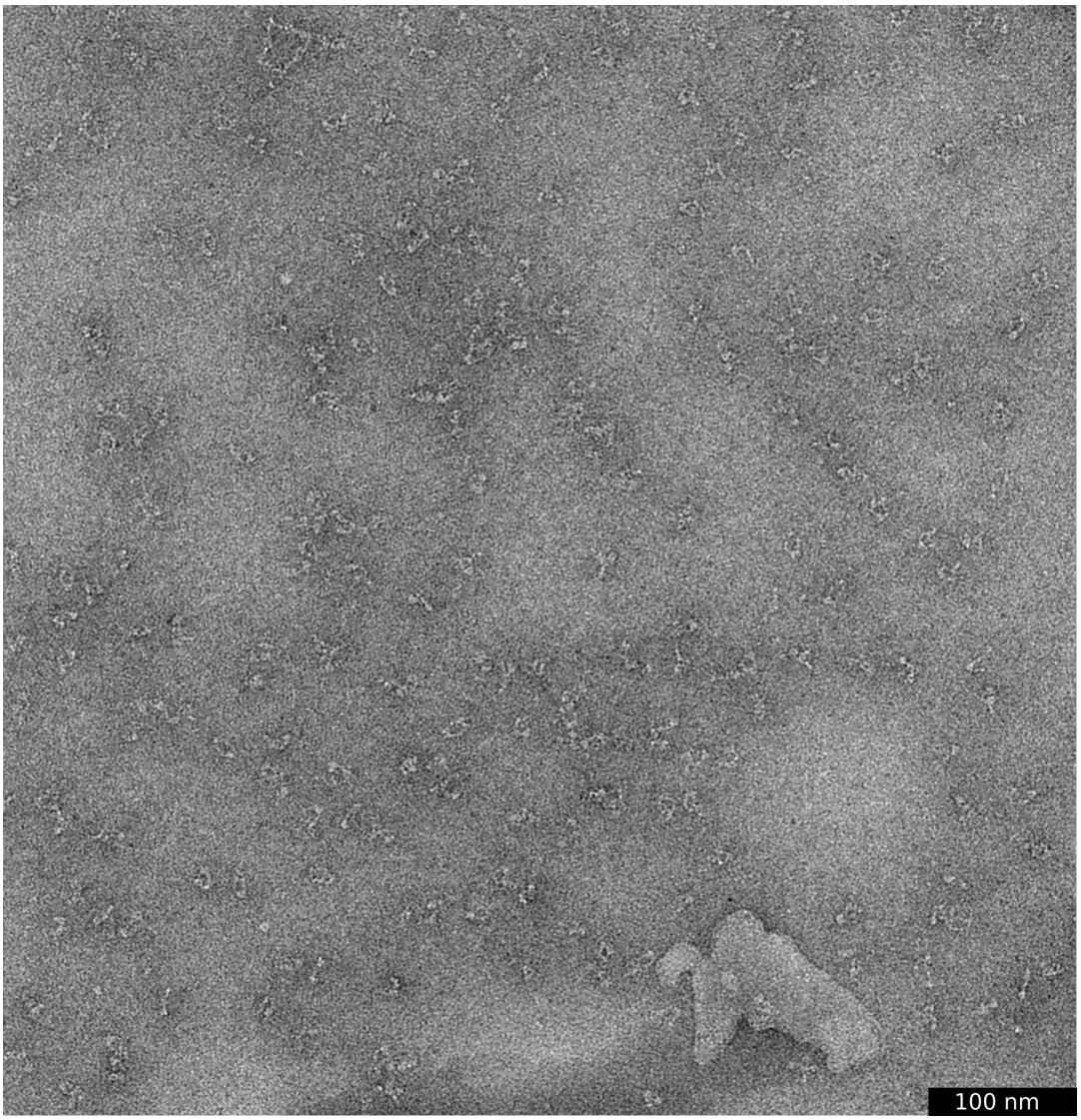
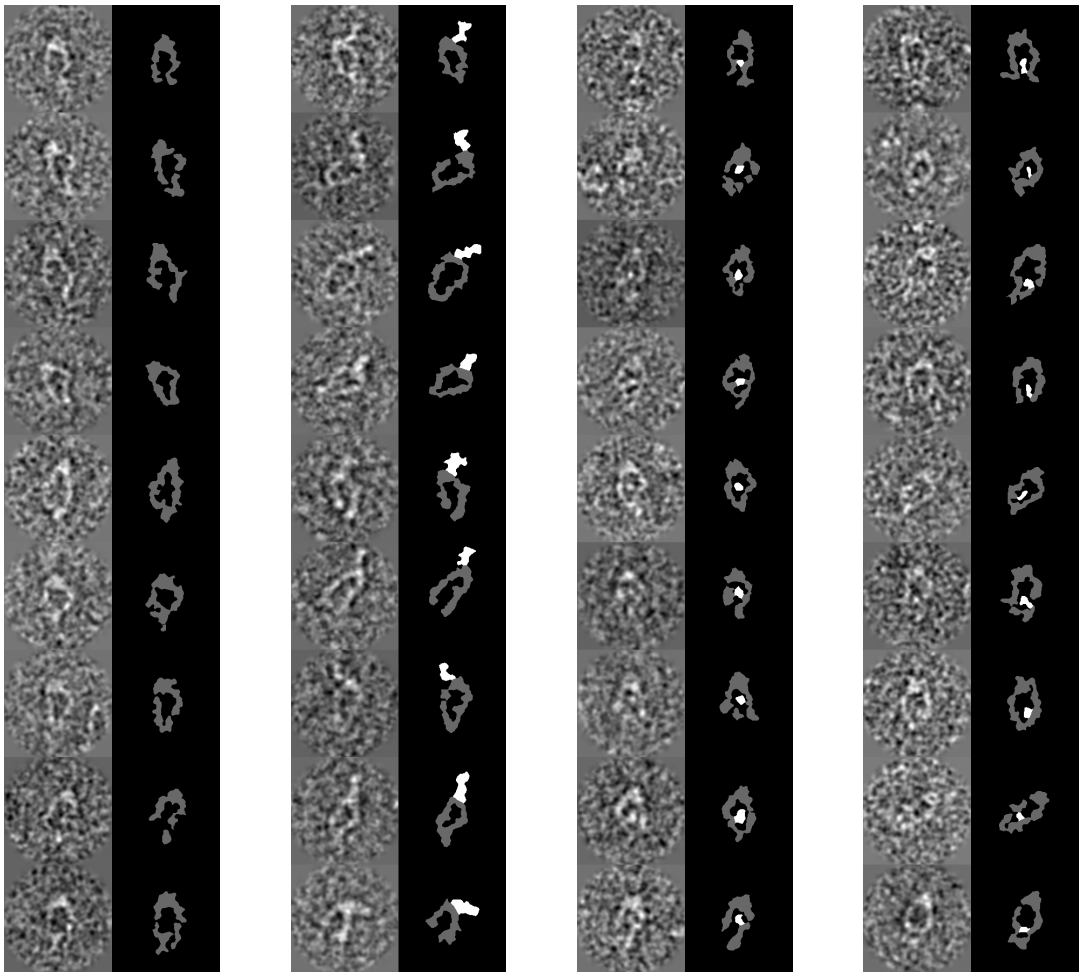


FIGURE 3.9: **Micrograph of SEC purified Calf1$_{EF}\alpha_{IIb}\beta_3$ integrin**. Representative image collected using a JEM-2200FS. Image has been low-pass filtered with a 0.05 (20 Å) cut-off. Scale bar is 100 nm.



(A) (B) (C) (D)

FIGURE 3.10: Randomly selected particles for the wild type integrin (3.10a), and each of the PA tag containing constructs

useful target for testing the utility of PA tag NZ-1 Fab labelling on small, conformationally diverse proteins (**Figure 3.11**). By creating several PA tag containing integrin constructs (PA tag inserted into W2, Calf1$_{XZ}$, or Calf1$_{EF}$ domains) I could demonstrate that these proteins showed very similar hydrodynamics with the wild-type protein suggesting that the PA tag insertion does not disrupt the overall conformation of the protein (**Figure 3.2-3.5**.) Upon addition of excess NZ-1 Fab the elution peak remained monodisperse and increased in elution mass for W2 and Calf1$_{EF}$, while Calf1$_{XZ}$ showed no change which corresponds with the predicted binding location of the NZ-1 as internal to the 'legs' of the integrin heterodimer (**Figure 3.1**). The peak shift suggested that the NZ-1 Fab forms a stable complex with the PA tag integrin and fully occupies the PA tag site. Further, the presence of NZ-1 Fab heavy and light chains were confirmed for all PA tag containing constructs during SDS-PAGE (**Figure 3.2-3.5**, *inset*). Raw whole field micrographs showed that PA tag insertion did not disrupt global conformation as these integrins had a similar conformation to the wild type construct (**Figure 3.6-3.9**), showing that the PA tag does

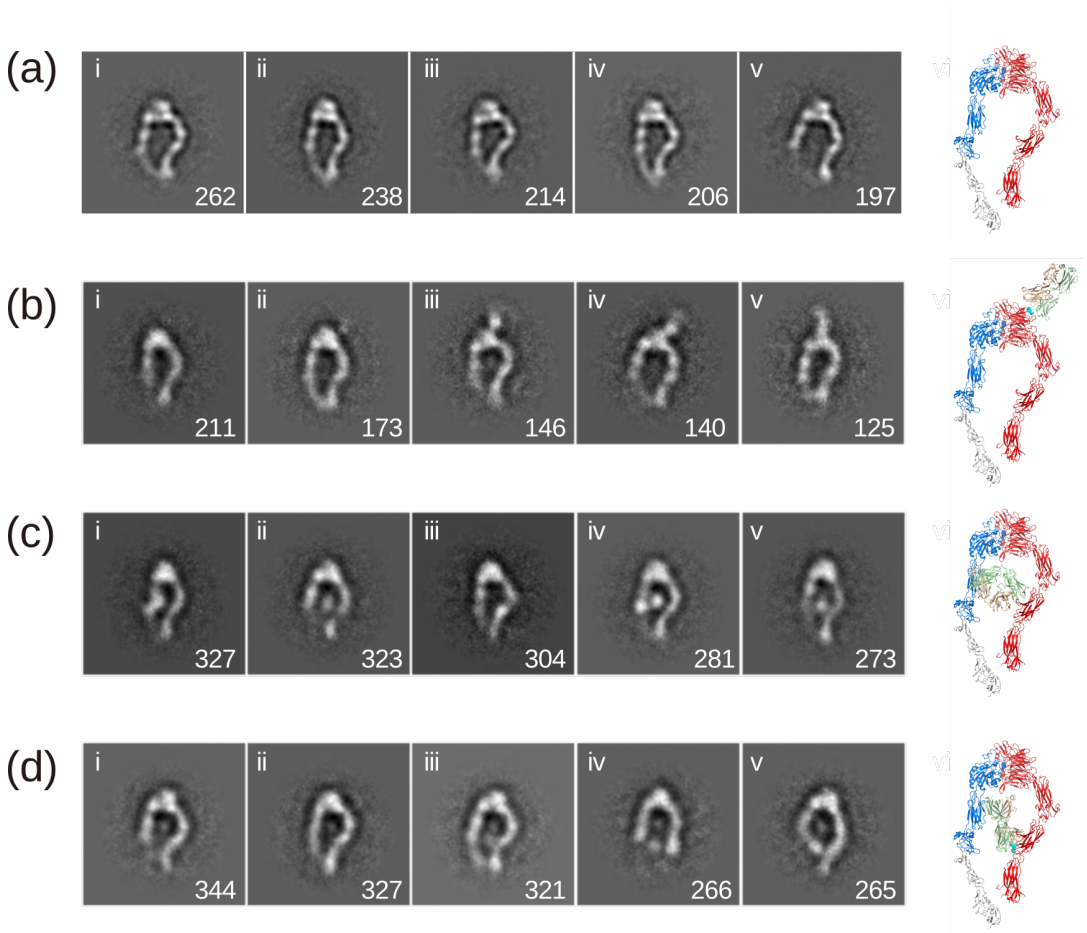


FIGURE 3.11: **Top five 2D classes and predicted atomic model showing predicted binding location of NZ-1 Fab**. 2D classes. Full classes are shown in **Supplementary Figure 5.14**.

not disrupt the global conformation. An additional density was present that corresponded with the predicted binding location of the NZ-1 Fab based on the insertion site of the PA tag (**Figure 3.11**). Overall, this shows the utility of the PA tag and NZ-1 as a novel EM label.

Chapter 4

Discussion

The PA tag and NZ-1 antibody were originally designed as a protein purification system (Fujii et al., 2014), but due to the very high affinity[1] of the PA tag/NZ-1 interaction and its useful properties[2] further biochemical and structural characterisation was completed (Fujii et al., 2016b). After obtaining the crystal structure and testing the insertion compatibility of the the PA tag, Professor Takagi and co-workers (Fujii et al., 2016b) remarked that:

> [O]wing to the very slow dissociation rate of the PA-tag–NZ-1 interaction ... labeling with the Fab or the single chain variable fragment (scFv) of NZ-1 could be used to visualize individual domains or subunits in electron microscopy analyses ...

Here, I demonstrated the accuracy of their prediction and showed that the PA tag and NZ-1 Fab can function as a high quality EM label, even for small and very flexible proteins such as $\alpha_{IIb}\beta_3$ integrin. Comparisons between the gel filtration data of wild-type and PA tag containing integrins showed little difference in their elution profile suggesting that the PA tag insertion does not cause disruption to the protein hydrodynamics (**Figure 3.2-3.5**). This is an important property for a genetically encoded EM label, as unwanted structural (or functional) changes could result in projections of particles that do no resemble the wild-type protein. If a meaningful comparison between the wild-type and tagged construct is not possible then identification of the various domains cannot be made with any confidence. Further evidence for the structural stability of PA tag containing proteins was provided by examining the raw micrographs (**Figure 3.6-3.9**) and 2D class averages (**Figure 3.11**), which showed little difference in the global architecture between the wild type and PA tag containing constructs. This is hardly surprising as there is a growing body of

[1] The K_D of the interaction between the NZ-1 antibody and the PA peptide was $K_D = 4.00\text{e}{-10}$ when measured using Bio-layer interferometry as shown in **Figure 1.1**, which was one to two orders of magnitude higher than other tested affinity purification tags (Fujii et al., 2014).

[2] Two useful properties that stand out are the very mild elution conditions and the easy regeneration of the antibody bound sepharose.

work that has used the PA tag as a purification tool for both biochemical and structural based experiments (Hirai et al., 2017; Lammert et al., 2017; Matsunaga et al., 2016; Meng et al., 2017; Mihara et al., 2016; Suzuki et al., 2016; Umitsu et al., 2016; Ushioda et al., 2016), which presumably would not be the case if this 12 amino acid insertion causes gross structural or functional distortions for target proteins. Previous work that aimed to compare PA tag containing constructs and show they did not suffer from functional deformities compared with the wild-type protein has already been completed (Fujii et al., 2016b), and now the stability of the PA tag insertion has been further supported with the first explicit comparison between wild-type and PA tag containing constructs at intermediate resolution using EM.

Upon addition of NZ-1 Fab a stable complex formed between that caused a uniform increase in elution mass for the externally oriented W2 PA tag insertion (**Figure 3.3**), strongly suggesting that there was full occupancy of the PA tag site. The other constructs also formed a stable complex with NZ-1 but there elution mass did not uniformly increase as was seen with the W2 construct. The presence of NZ-1 Fab in the elution peaks as confirmed with SDS-PAGE suggests that the choice of insertion site has an effect on whether the NZ-1 Fab binding causes and increase in the hydrodynamic radius of the protein. Indeed, once the location of the insertion site is examined in the predicted structural model it can be seen that the W2 site is externally oriented whereas the Calf1$_{XZ}$, and to a lesser extent the Calf1$_{EF}$ sites are internally oriented and binding of the NZ-1 Fab is between the two subunit 'legs'[3]. The presence of a density very likely corresponding to NZ-1 Fab was confirmed for all constructs by transmission electron microscopy. The images showed an additional density visible in both the raw micrographs (**Figure 3.6-3.9**) and 2D class averages (**Figure 3.11**) that corresponds with the predicted binding location of the NZ-1 Fab. The Calf1$_{XZ}$ construct that showed no change in estimated mass upon binding of NZ-1 during gel filtration had a clear density between the lower legs of the subunit, supporting the hypothesis that this binding location is protected by the two subunits and would not increase the hydrodynamic radius. Analysis of the members of each class suggested that between 50-90% of all particles have an NZ-1 Fab density present at the PA tag site (**Table 3.1**) which correlates with the high occupancy seen from the W2 gel filtration data. *In toto*, these data strongly demonstrate the utility of the PA tag and NZ-1 Fab as a mobile EM label that can be used to map subunits or domains of proteins using electron microscopy. The PA tag can function as a mobile epitope that requires only a single high affinity antibody to effect high affinity and occupancy tagging of a protein, thereby removing the need to develop potentially costly antibodies or use other labelling strategies.

[3]In this case the hydrodynamic radius of the protein will not change, much in the same way the total radius of a donut does not change if you fill the hole with strawberries and cream.

4.1 Criteria for a successful EM label

The quality of a labelling methodology can be judged by four factors: 1) specificity of the labelling reaction, 2) occupancy of labelling site, 3) signal intensity, and 4) precision of localisation information. I will recapitulate previously published reports, combined with the the data I presented in this thesis to judge the quality of the PA tag and NZ-1 based on these criteria.

1) *Specificity*. High specificity is important as it prevents unwanted domains being labelled, which reduces the variation in the labelled proteins. The original paper that first demonstrated the use of the PA tag/NZ-1 as a protein purification system showed that it could purify proteins from whole cell lysate with minimal contamination indicating the high specificity of the PA tag/NZ-1 interaction (Fujii et al., 2014). This has been demonstrated repeatedly with many other papers that have used this system (Hirai et al., 2017; Lammert et al., 2017; Matsunaga et al., 2016; Meng et al., 2017; Mihara et al., 2016; Suzuki et al., 2016; Umitsu et al., 2016; Ushioda et al., 2016) and from my own experience using the PA tag as a purification handle in this and other projects.

2) *Occupancy*. Occupancy refers to the level of tagging sites that have a label present. The occupancy of a tag used for labelling is an important property as this influences both the data collection requirements and may have some impact on the maximum resolution that can be obtained. Data collection needs to be continued until some desired level of signal can be obtained that can be used to identify a feature of interest. For low occupancy tags the minimum dataset size will logically increase as some subset of the data will contain particles with no visible tag. Low occupancy of a tagging site is potentially even more deleterious for data collection than just increasing the required amount[4] of data collected. If there is some conformational change that is caused by tag binding, and tagged and untagged projections can not be easily distinguished then it is possible that there will be incorrect alignment and lowering of the obtainable resolution. This second issue can be addressed by using tags with a strong signal intensity, as will be discussed shortly. Regardless, the occupancy of the PA tag by the NZ-1 Fab as measured by complex formation was very high during gel filtration (**Figure 3.3-3.5**) and analysis of the 2D averages suggested that there was an occupancy of 50-90% (**Table 3.1**). This shows that the NZ-1 Fab has a very high occupancy of the PA tag site which should reduce the data collection requirements for a project using this system as a label.

[4]Given the increasing usage demands for data collection low occupancy tags that require larger datasets could easily be a 'lethal' factor in deciding if an experimental question can be feasibly answered or not. When high end machines are in such demand that researchers have access only one day a month the quality of the sample is paramount.

3) *Signal intensity*. Signal intensity relates to how clearly the tagging signal can be identified from background, structural, or other sources of noise[5]. Fab labels have been used extensively in EM experiments (Boisset et al., 1993; Boisset et al., 1995; Jiang et al., 2004; Prasad et al., 1990; Wu et al., 2012) and so clearly they have a strong enough signal (mass) to be effective labels. Indeed, even smaller proteins such as GFP can be used (Ciferri et al., 2015) and so it is clear that the signal intensity of the NZ-1 Fab is high enough to be a successful labelling strategy. In cases were the background noise is very high, or for experiments that for some reason or other require a stronger signal intensity then one presently untested possibility is the conjugation of gold labels to NZ-1 Fab to greatly increase the contrast (Hainfeld and Powell, 2000). Although this has not been tested with the NZ-1 Fab there should be no technical reason this isn't feasible, and provides a method to greatly increase the contrast in any EM imaging experiment.

At this point in our assessment of the PA tag and NZ-1 Fab as a labelling tool it is hopefully clear that the specificity, occupancy, and signal intensity is of sufficient quality to consider this labelling system to be effective by some measure. The final criterion for a labelling strategy is the precision of the localisation signal, or in other words, how accurately the label can be used to identify some portion of a target protein.

4) *Precision of localisation signal*. How precisely does the NZ-1 Fab bind to the PA tag insertion site? The data presented in this thesis demonstrates that it has sufficient accuracy for subunit and domain level identification, but presently there is no published data about the precision at a higher resolution. To address this question I will outline some potential sources of localisation inaccuracy and detail the presently available data and what it suggests about the the PA tag and NZ-1 system. However, it should be noted that without additional experimentation this question can not be answered in full, and this discussion is fundamentally hypothetical.

There are many potential sources of precision inaccuracy for EM labelling systems, some of which may not be related to the label itself, but to the protein (i.e., high conformational variability could make it difficult to distinguish between a flexibility intrinsic to the label or the protein). If we exclude protein conformational variability then we can limit the two sources to either 1) tag flexibility, or 2) antibody flexibility. **Figure 4.1a** presents either the flexibility of the tag or flexibility of the antibody schematically. The EM data presented in this thesis has some limitations when trying to address the question of precision of the PA tag and NZ-1 Fab as it is not high enough resolution to confidently determine the precision or cause of any apparent variability in label placement. Take note of the difference in

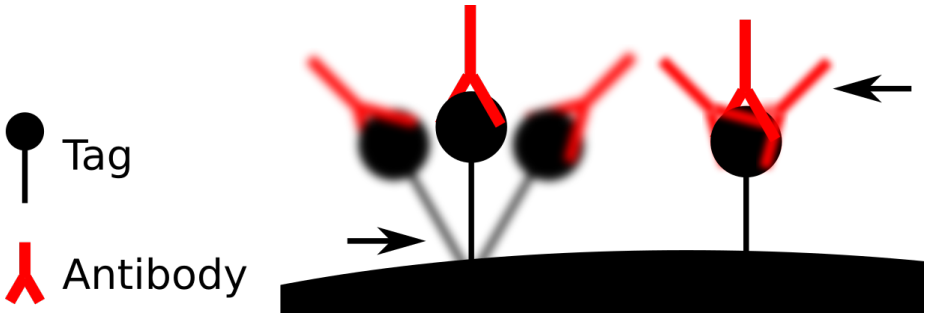
[5]Noise in this sense does not have to be a strict addition of Gaussian signal to the projection, rather, any signal that is mixed or indistinguishable from the label.

length of the Fab density in **Figure 3.11** (b)iii,iv and v. This could be caused by either conformational variability of $\alpha_{IIb}\beta_3$ integrin, flexibility of the PA tag, flexibility of the NZ-1, or problems with misalignment of particles. The available data does not support one analysis over another and so the question is left open when examining these data. I address why $\alpha_{IIb}\beta_3$ integrin was used and the reasons behind not obtaining an atomic resolution representation of integrin in complex with NZ-1 Fab in **Section 4.4.1** but for the present discussion we have access to other sources of information on the potential precision of the PA tag and NZ-1 Fab. While the EM data cannot be used to answer this question we have atomic level structural information in the crystal structure of the PA peptide in the NZ-1 binding pocket (Fujii et al., 2016b). As discussed in **Section 1.3.1** the NZ-1 Fab binding pocket does not undergo a serious conformation change between the bound and unbound states, and there was high conformational similarity between both PA peptides observed in the crystal structure. This indicates that perhaps the NZ-1 Fab will not have too much flexibility when binding to the PA peptide. On the other hand the terminal regions of the PA peptide are not involved in the antibody recognition reaction and despite being spaced 10.2 Å apart (**Figure 4.1b**) were able to be inserted into unstructured loops with as little as 5.2 Å spacing (the W2 loop in **Figure 3.1**). This indicates that the terminal regions of the PA tag may be flexible and contribute to a lowering of the precision of this tag. Finally, it should be stressed again that this cannot be answered with any confidence given the limited data available, and additional experiments are needed to build a better model and understanding of this aspect of the PA tag and NZ-1 Fab. I will discuss some of these potential experiments in **Section 4.4**.

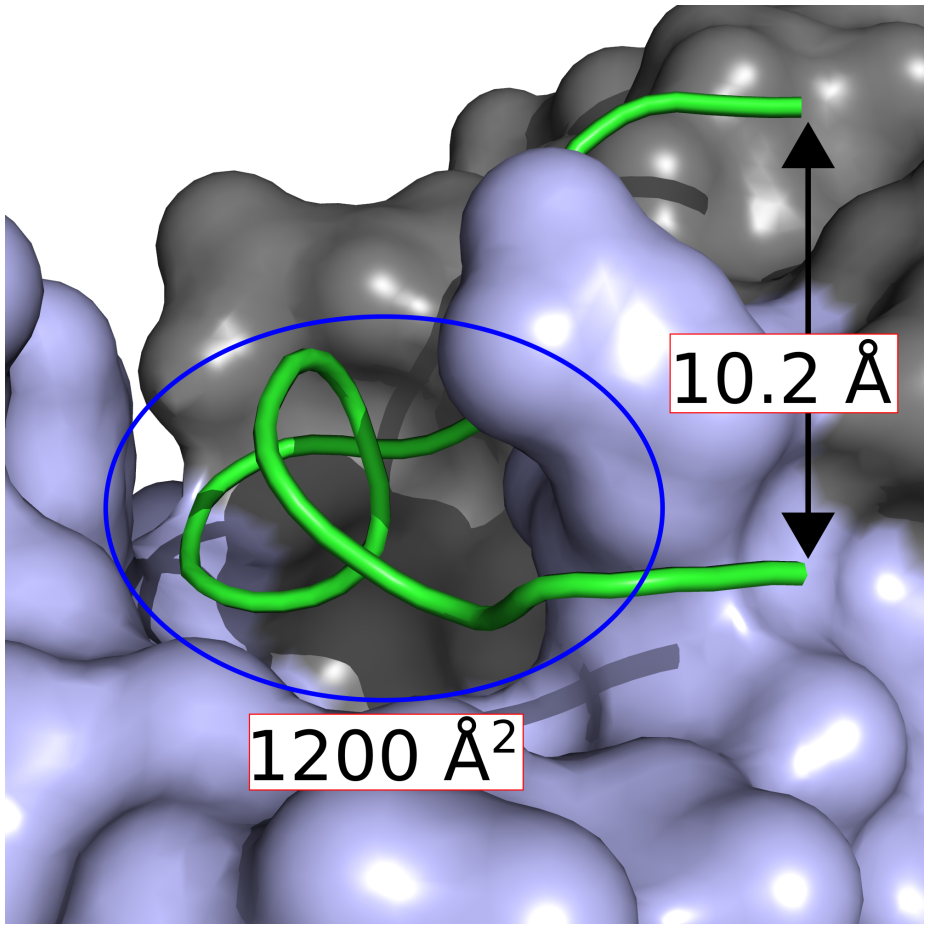
As outlined above the high performance of the PA tag and NZ-1 in regards to specificity, occupancy, and signal intensity is clearly supported, and there is no reason to suspect that the precision of localisation information does not also perform to a high standard, although before making such a judgement further experimental evidence is needed.

4.2 Comparison to other EM labels

Based on the four criteria of specificity, occupancy, signal intensity and precision to PA tag and NZ-1 seems to be an objectively high quality tagging system, but how does the PA tag and NZ-1 Fab compare to other EM labelling systems? Compared with other labelling systems the PA tag and NZ-1 Fab has a number of advantages. 1) The insertion is very small (only 12 residues), 2) there is a high degree of topological freedom with epitope placement, 3) the label is not genetically encoded giving temporal control, and 4) it relies on a single readily available antibody with proven high affinity and high occupancy.



(A) Two sources of precision error



(B) Carbon backbone of PA peptide in NZ-1 Fab binding pocket

FIGURE 4.1: **Precision of PA tag and NZ-1 Fab**. (Top) Two potential sources of label imprecision. Either the tag is flexible (left), or the antibody can bind in multiple conformations and has some flexibility. (Bottom) Surface representation of the NZ-1 Fab binding pocket with heavy chain (light blue) and light chain (grey). C_α line representation of PA14 peptide is shown with surface contact surface area between peptide and antibody marked. Prepared in Pymol from PDB file 4yo0.

1) *Small insertion size*. Some other labelling strategies have used either strategic deletion of subunits[6] (Bui et al., 2008; Heuser et al., 2009; Pigino et al., 2011) or the insertion of some additional domain such as GFP (Ciferri et al., 2012; Ciferri et al., 2015; Lees et al., 2010), or MBP (Ciferri et al., 2012; Mercogliano and DeRosier, 2007). While some of these results have been impressive the removal of a structurally important region or the insertion of an additional domain into a protein could interfere with the correct folding or final conformation. This could be an issue and may limit the widespread applicability of such techniques. In contrast, the PA tag is only 12 residues and has a terminal spacing of 10.2 Å (**Figure 4.1b**) and could even function when inserted into loops of 5.2 Å spacing (W2 loop in **Figure 3.1**). Furthermore, EM imaging showed no structural change between the wild-type and PA tag containing constructs of integrin, suggesting that the PA tag insertion does not cause any structural deformation.

2) *PA tag insertion site has a high degree of freedom*. Many EM localisation tools are limited in their placement to the C-, or N-terminals of a protein (**Table 1.3**). As discussed above epitope tags that are used for EM labelling often have this limitation (Kelly et al., 2010; Yip et al., 2010), but so do some domain insertion techniques (Mercogliano and DeRosier, 2007) and even some novel labelling systems (Flemming et al., 2010). If a central region is needed to be labelled then there are number of choices including GFP domain insertion (Ciferri et al., 2012; Ciferri et al., 2015; Lees et al., 2010), DOLORS (Domain localization by RCT sampling) labelling (Lau et al., 2012), and biotin-streptavidin labelling strategies (Oda and Kikkawa, 2013). As discussed elsewhere (**Section 1.4.5**) these strategies may not be ideal as they can introduce conformational variability if they are inserted into some structurally sensitive region of a target protein. Using smaller epitopes that are recognised by some affinity purification antibody (such as FLAG, Myc, or HA) then the insertion site is usually limited to the C-, or N-terminals of a polypeptide chain due to the linear recognition requirements of their cognate antibodies, or are placed at the end of long linker peptides to preserve their reactivity with their antibodies. In contrast, the freedom that the PA tag has when selecting an insertion site is a major advantage of this system, as it is quite resilient when selecting an insertion site. Both in this work and elsewhere (Fujii et al., 2016b) it has been demonstrated that the PA tag insertion site can be chosen with a high degree of freedom and it still remains reactive with the NZ-1 antibody.

3) *Temporal control of labelling*. The ability to time the 'activation' of the labelling signal by addition of some reagent could be important in proteins that undergo a high level of structural change. There are various tag systems that use antibodies (Kelly et al., 2010; Yip et al., 2010), small proteins such as streptavidin (Lau et al., 2012), or Maltose binding protein that can be conjugated to an unnatural amino acid

[6]More accurately they identified mutants that lacked some subunit, but functionally there is no difference between finding such a mutant and generating one in the lab.

(Dambacher and Lander, 2015). The PA tag and NZ-1 Fab also allow temporal control over binding by timing the addition of the NZ-1 to after the target macromolecular complex has formed.

4) *Antibody based technology*. Many tagging systems use some kind of chemical reaction to induce tagging (Lau et al., 2012; Oda and Kikkawa, 2013), or rely on the incorporation of an unnatural amino acid to allow tagging (Dambacher and Lander, 2015), or the addition of multiple protein components to induce labelling (Flemming et al., 2010). However, the PA tag is recognised by a single high affinity antibody in a relatively well understood manner (Fujii et al., 2016b). There are many antibody based technologies to improve contrast (Koeck et al., 1996), or to improve antibody design to reduce flexibility and reduce the overall size of the Fab (Arimori et al., 2017).

4.3 Future uses of the PA tag and NZ-1 antibody

The technological advances in the electron microscopy field have been staggering with improvements in direct electron detectors (McMullan et al., 2014), phase plate technology for increased contrast (Danev et al., 2014; Danev et al., 2017), CTF estimation (Rohou and Grigorieff, 2015; Zhang, 2016), movie alignment (Zheng et al., 2017), and analysis software (Lyumkis et al., 2013; Scheres, 2012). In line with this smaller and more heterogenous proteins are being determined to atomic resolution[7] (Danev et al., 2017; Merk et al., 2016). However,Henderson, (1995) predicted that there is a theoretical limit to the lower size of a protein that can have an atomic resolution structure determined using electron microscopy due to the ionising nature of electrons. Indeed, it was shown that as proteins decrease in size the alignment errors increase (Henderson et al., 2011). Incorrect alignment acts as a truncation of signal in Fourier space (see **Appendix D**,**Equation D.3** and **Section 1.4.2**). One method of increasing the size of a target complex to improve the alignment and obtainable resolution is the use of antibody Fab to increase the size of the target complex by roughly 50 kDa. This was first used by Jiang et al., (2004) to find the structure of the voltage-dependent K(+) channel to 10.2 Å and then more systematically demonstrated by (Wu et al., 2012). Obviously the PA and NZ-1 Fab could also be used in this way, as a tool to increase the mass of a target complex and increase the resolution that could be obtained. In particular, since the binding affinity and occupancy of the PA tag site are very high it could even be possible to use multiple PA tag insertions to greatly increase the mass of a small protein. This strategy could be combined with new antibody designs that improve stability (Arimori et al., 2017) and could have real utility for high resolution cryo-EM. Other techniques such as antibody affinity grids that

[7] See some criticism of these structures in (Wlodawer et al., 2017)

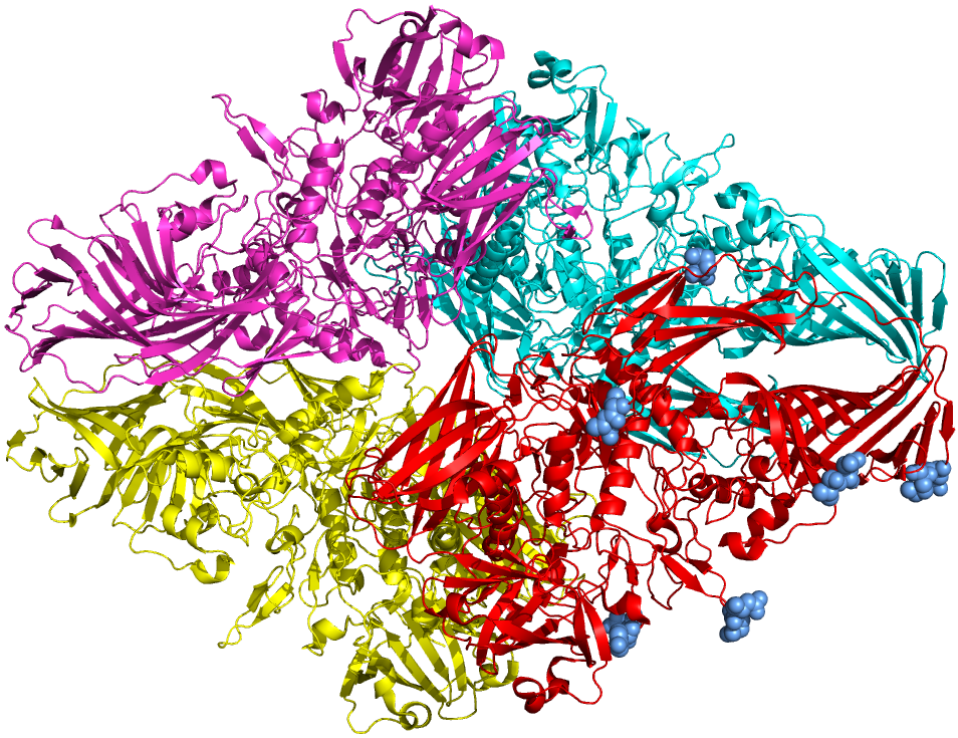
increase the particle density by binding proteins have an obvious potential use for the PA tag and NZ-1 Ab (Yu et al., 2016)

4.4 Alternative experiments

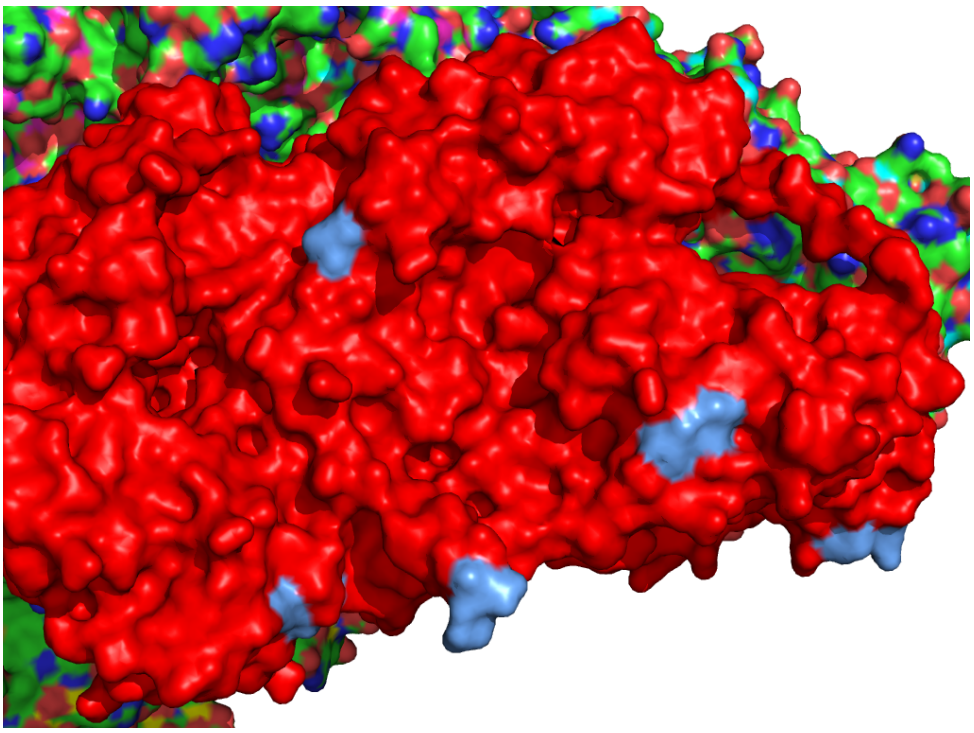
During the design phase of these experiments there was a trade-offs between demonstrating the labelling system on a small, flexible protein that could be very difficult to determine at a high resolution (see the following **Section 4.4.1**), or use a rigid 'easy' model protein that could potentially reach very high resolution as the target. I elected to use the PA tag on integrin, which is a relatively small and flexible protein and show that even for these difficult cases labelling with PA/NZ-1 is still effective. However, now that I have these data the next set of experiments should be on a more conventional model system to try and determine to atomic resolution the binding interface of the NZ-1 with PA tag. One potential target is the bacterial protein β-galactosidase as this is already a model target for cryoEM experiments (Henderson and McMullan, 2013). β-galactosidase is a homotetramer of 1024 residues and 464 kDa, and would be a useful target for further testing the PA tag as it has many surface exposed unstructured loops that could serve as insertion sites for the epitope. **Figure 4.2** shows some selected sites at the end of various loops of differing lengths. Insertion sites in loops of different lengths, or with different levels of access to the surface (**Figure 4.2b**) would help demonstrate the level of structural variability and binding flexibility of the NZ-1 Fab and PA tag. Although I have picked β-galactosidase there are other proteins that could also be used, such as the 20S proteasome which also is stable and may provide a good target system.

4.4.1 Why I didn't collect more data and go to 3D

Perhaps during reading this thesis the question of "*Why not just get more data and find the structure at atomic resolution*?" has come up. I presented the experimental choice above as between either a conformationally stable protein and a difficult protein like integrin, but is there a good reason to think this? While it is theoretically possible that we could improve the image quality by increasing the number of particles integrin (and other flexible proteins) are particularly difficult due to the high flexibility of the C-terminal ends of the heterodimer legs. Unpublished data provided by Professor Junichi Takagi and Associate Professor Kenji Iwasaki showed that the 3D density from over 20,000 particles still cannot be resolved into a clear structure. **Figure 4.3** is one example 3D class, where the identity of each subunit can be inferred by not with a high degree of confidence. Integrin is quite conformationally flexible (see **Section 1.5.1**) and if there are only five total conformations then the data collection requirements would be at least 100,000 particles to determine each conformation to



(A) Cartoon representation



(B) Surface representation

FIGURE 4.2: **β-galactose**. (Top) Cartoon representation of β-galactosidase chain A (red), chain B (turquoise), chain C (pink), and chain D (yellow) are shown. Potential insertion sites in chain A for PA tag are marked with blue spheres in six locations with unstructured loops between residues 95-96, 580-581, 633-634, 706-707, 744-745, and 890-891. (Bottom) Surface representation of chain A with potential insertion sites marked in blue (resi. 633-634 are hidden from view).

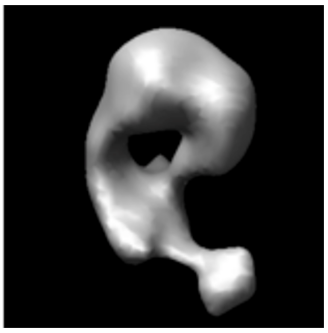


FIGURE 4.3: **3D model of integrin**. Representative class of integrin built from 20,000 particles. Data provided by Professor Junichi Takagi and Associate Professor Kenji Iwasaki.

some reasonable level. That could be over 1000 micrographs per PA tag construct which would greatly increase the data collection time. If the purpose of this project was to object was to identify the global architecture of integrin then I would collect a much larger datasets, however, the goal of these experiments was to demonstrate that the PA tag could be used as a mobile label to identify subunits (or domains) for small, flexible proteins such as integrin. Since the bent-form crystal structure (Zhu et al., 2008) and global architecture from 2D EM averages is already known (Takagi et al., 2002) I reasoned that greatly increasing the data collection and trying to find a 3D model was unnecessary for the purpose of these experiments.

4.5 Conclusion

Here I have shown that the PA tag and NZ-1 antibody can be used as a mobile EM label that can tag subunits or domains for proteins where other high quality antibodies are not available. This functionality can be added to the ever growing list of methods that the PA tag and NZ-1 can be used for, which now includes: 1) protein purification (Fujii et al., 2014; Hirai et al., 2017; Lammert et al., 2017; Matsunaga et al., 2016; Meng et al., 2017; Mihara et al., 2016; Suzuki et al., 2016; Umitsu et al., 2016; Ushioda et al., 2016), 2) Western blotting, flow cytometry, and immunoprecipitation (Fujii et al., 2014; Fujii et al., 2016b), 3) single molecule reporting of conformation and modification of cell surface receptors (Fujii et al., 2016b), and 4) EM domain labelling. As discussed in above there may be applications for the PA tag as a crystallisation chaperone (**section**), or as a tool to improve the resolution of small proteins using cryo-electron microscopy (**Section 4.3**). The work presented in this thesis has contributed to the body of knowledge on the utility of 12 residues, GVAMPGAEDDVV, and a high affinity antibody.

Chapter 5

Supplementary Data

5.1 Overview of Supplementary Data

This is the supplemental data section that has additional data that complements the main thesis.

5.1.1 Determining the optimal density for EM imaging

Since use of high-end microscopes is very competitive initially I determined the optimal particle density using a smaller Hitachi microscope (**Section 2.9.4**). From my experience using the Nanodrop to determine the protein concentration and then diluting the sample to some optimal has not been very accurate, and so I used the absorbance values directly from the size exclusion chromatogram to determine the optimal denisty. From experience dilutions to 1.5-1.6 mAu are ideal for imaging with TEM. For illustration an undiluted 6 mAu protein sample (**Figure 5.4**) can be compared with a dilution to 1.5 mAu (**Figure 5.5**). Take note that the density of particles is decreased and should allow selection (if imaging conditions are optimised) of individual particles for image processing.

TABLE 5.1: Concentration of $\alpha_{IIb}\beta_3$ integrin: 350 μL

Construct	**Protein concentration [mg/mL]**
wild-type	0.195
W24326	0.293
Calf1$_{XZ}$4329	0.177
Calf1$_{EF}$4330	0.035

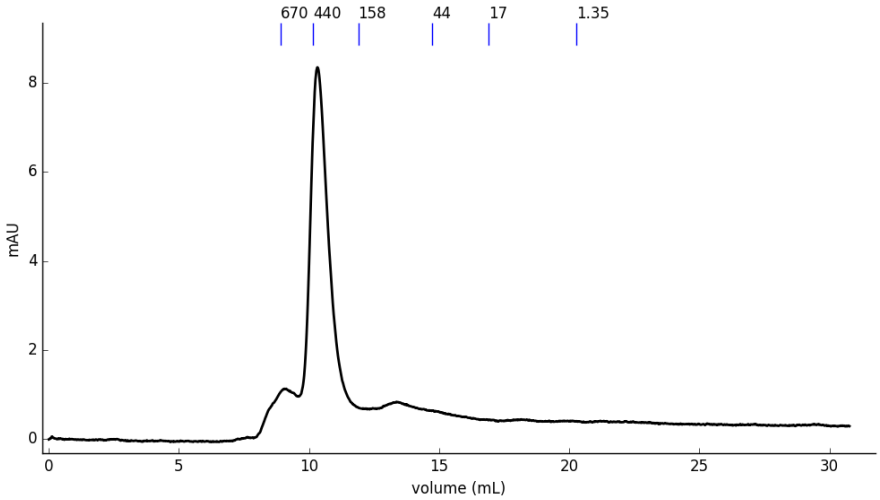
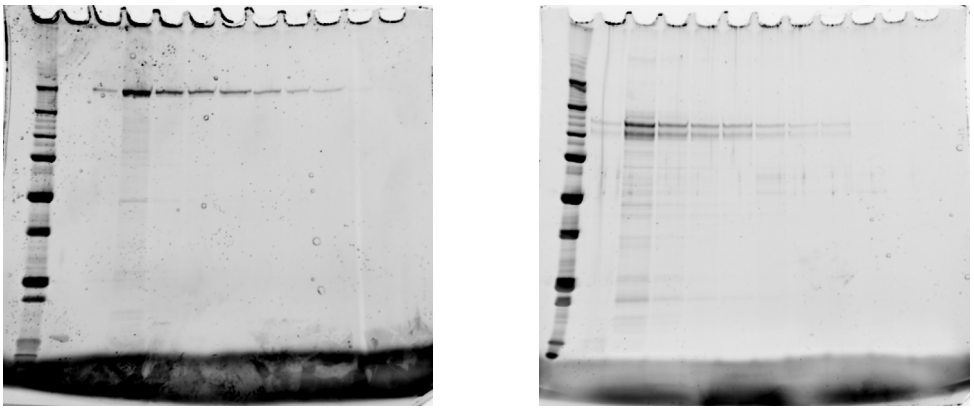
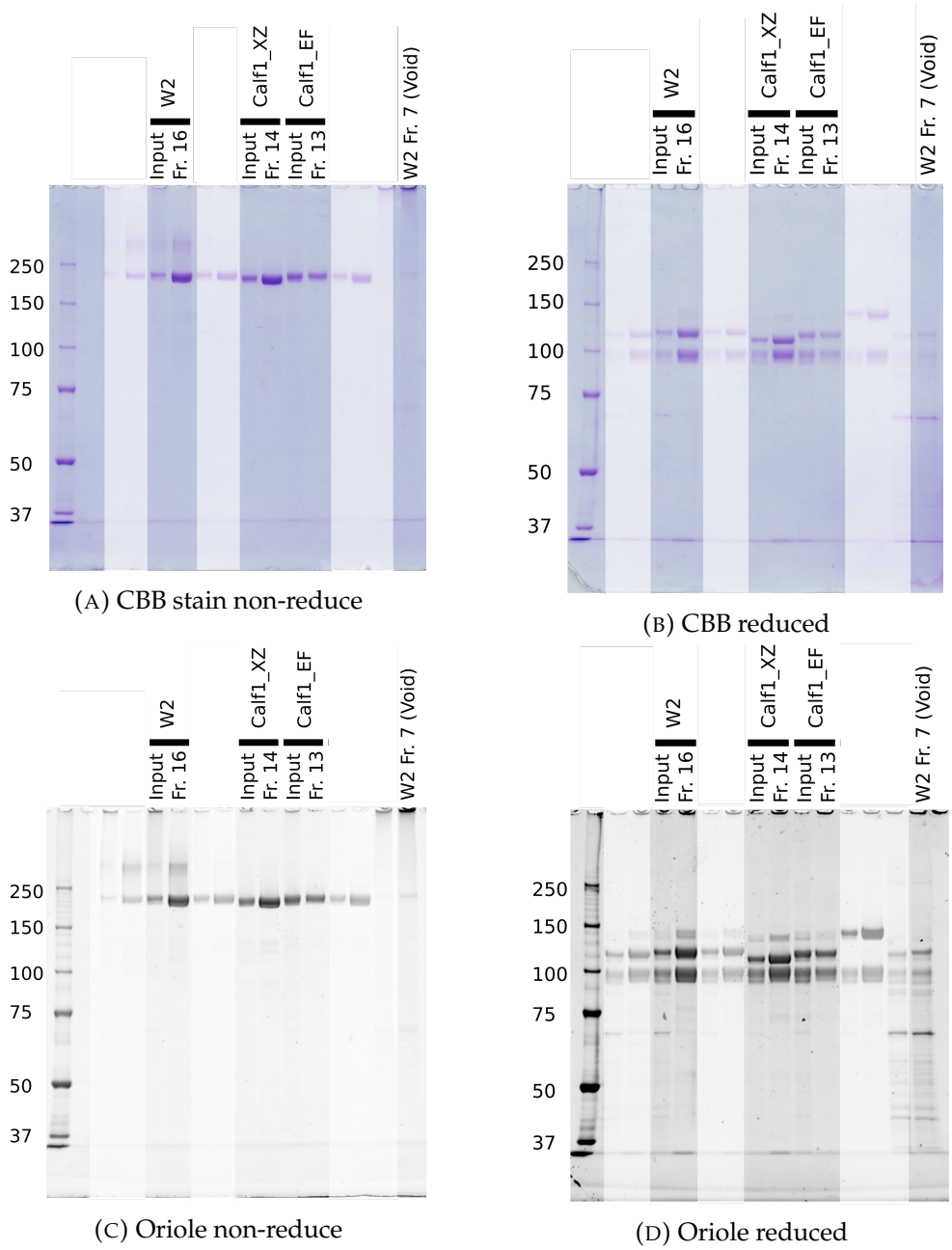


FIGURE 5.1: **SEC data of wild-type $\alpha_{IIb}\beta_3$integrin**. Gel filtration marker is from a protein size standard (shown at top) that included thyroglobulin, ferritin, γ-globulin, ovalbumin, myoglobin, and vitamin B12.



(A) Oriole stain non-reduce (B) Oriole stain reduced

FIGURE 5.2: **Purification of wild-type integrin**. SDS-PAGE of wild-type integrin purified using Ni-NTA and eluted with TBS containing 300 mM imidazole. Elutions 1-10 are shown in both non-reduced and reduced gels.



(A) CBB stain non-reduce

(B) CBB reduced

(C) Oriole non-reduce

(D) Oriole reduced

FIGURE 5.3: **Example SDS-PAGE data of integrin from book5 p134.** $\alpha_{IIb}\beta_3$integrin was purified using NZ-1 bound to sepharose and concentrated before injection into SEC. The purified fractions were run on an SDS-PAGE and stained with oriole and then CBB.

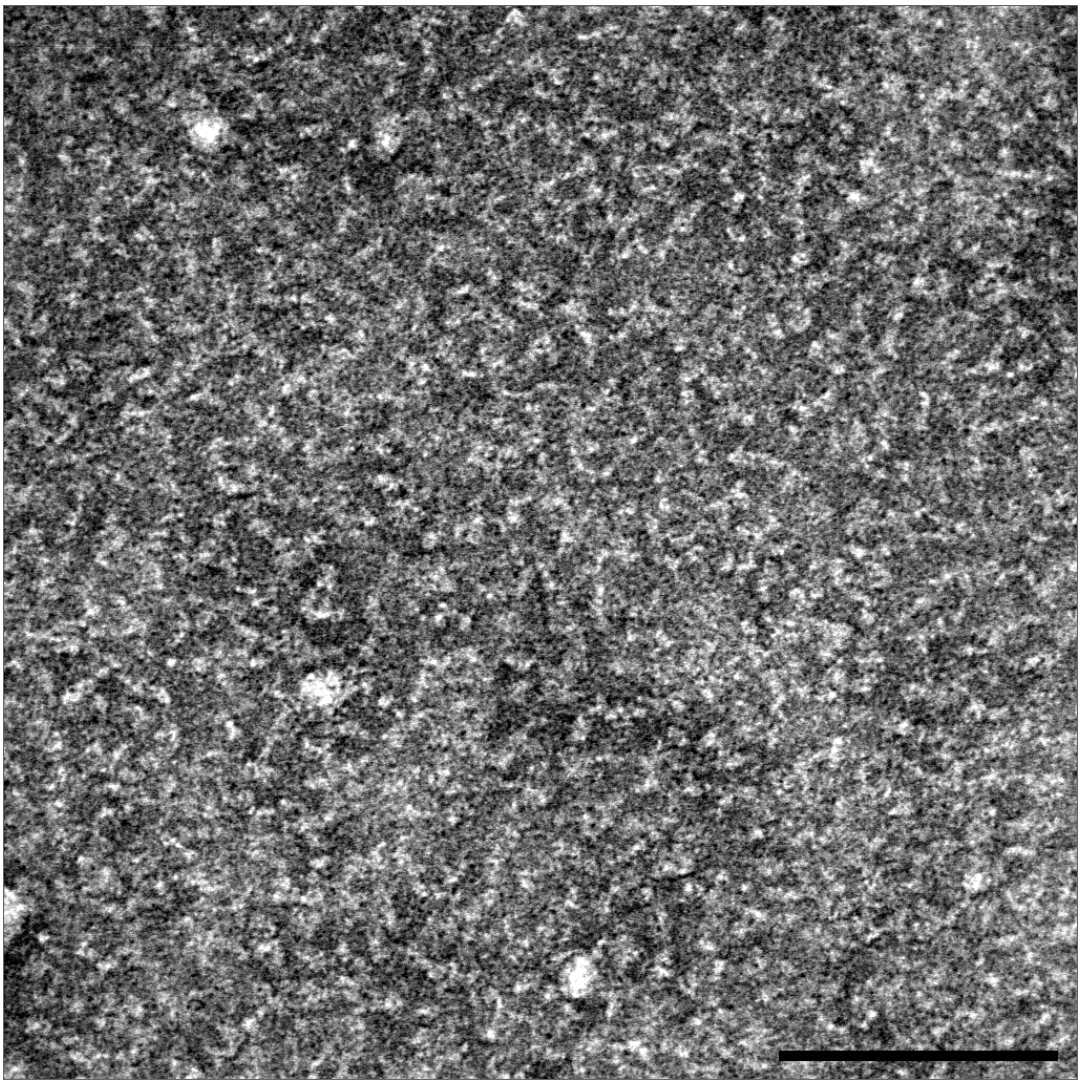


FIGURE 5.4: **Micrograph of SEC purified $\alpha_{IIb}\beta_3$integrin containing a PA tag insertion in Calf1$_{XZ}$domain**. The Calf1$_{XZ}$peak of SEC purified $\alpha_{IIb}\beta_3$integrin was with stained ammonium molybdate before imaging with EM using a Hitachi H-7650. Scale bar is 100 nm.

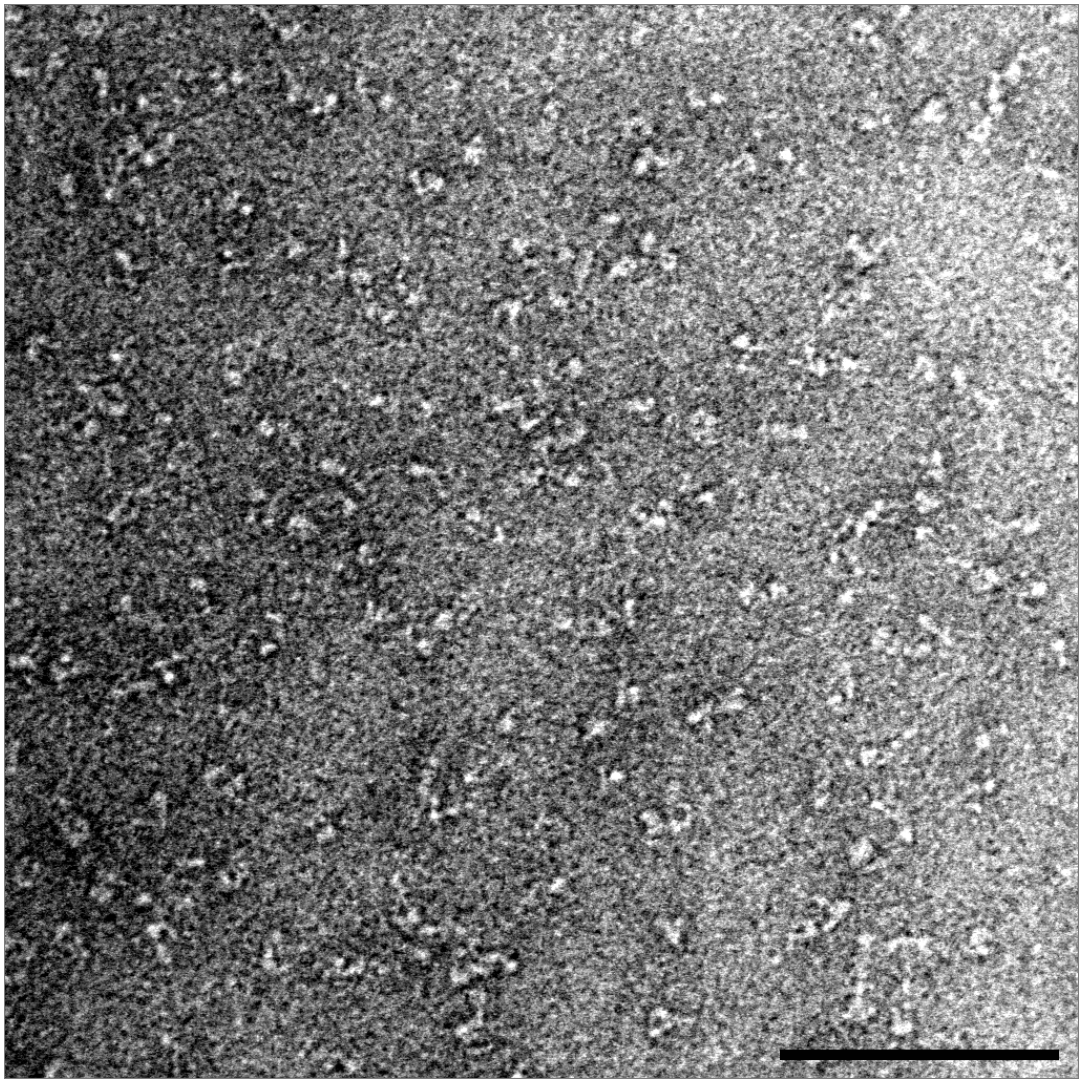


FIGURE 5.5: **Micrograph of SEC purified $\alpha_{IIb}\beta_3$integrin containing a PA tag insertion in Calf1$_{XZ}$domain**. The $\frac{1}{4}$ dilution main peak of SEC purified $\alpha_{IIb}\beta_3$integrin was with stained ammonium molybdate before imaging with EM using a Hitachi H-7650. Scale bar is 100 nm.

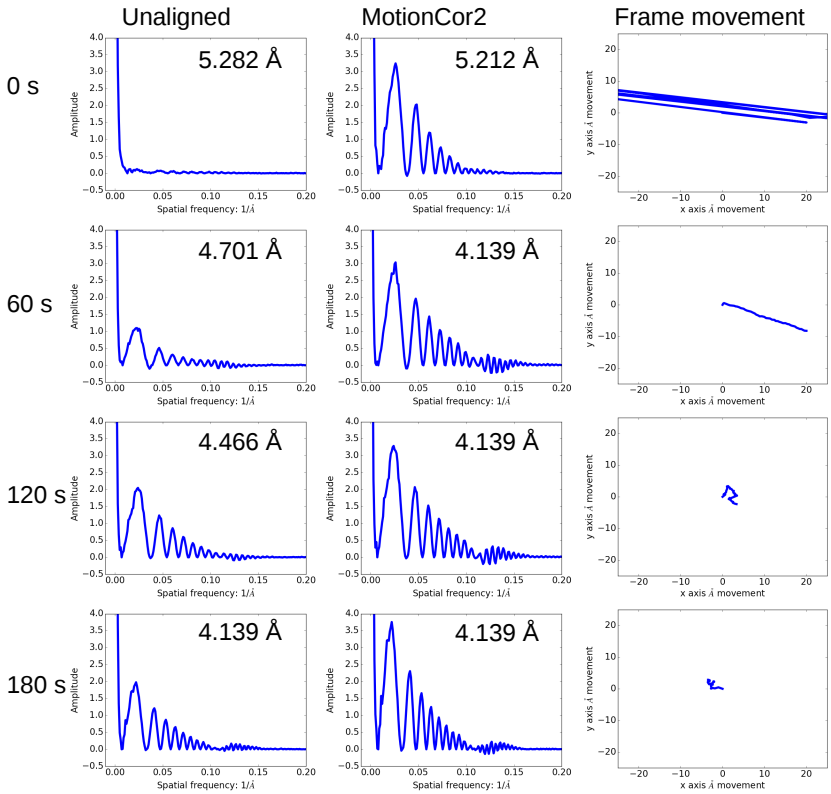


FIGURE 5.6: **2D amplitude of the unaligned, aligned micrographs and total stage movement**. Micrographs were collected at 20,000x magnification corresponding to a pixel size of 1.941 Å. After the indicated (left) time in seconds an image was recorded in counting mode with a dose rate of 8.0 e^-/pixel/s for a total dose of 41.2 e^-/pixel. For each time period the frames were either directly averaged, *Unaligned*, or aligned using MotionCor2 as described in the methods, *MotionCor2*. The CTF parameters were calculated using gCTF (Zhang, 2016) and the maximum resolution is shown in the top right for each unaligned and aligned micrograph. The total x-y movement during alignment was measured using the output files from gCTF and shown on an identical scale, *Frame Movement*.

5.2 Movement of the JEOL stage

I measured the total movement between frames after moving the goniometer as a proxy for residual movement in the gonimoter (Fig. 5.6). There is a *higher* level of movement as I did not roll the tracking ball back half a turn, as is common to reduce movement during imaging as I was concerned that I couldn't replicate that. Instead I moved the trackball one turn to the right and waited the indicated number of seconds.

5.2.1 JEOL optimal imaging conditions

During EM imaging after moving the sample holder there is some unwanted drift and residual motion. To this end I tried to determined in a relatively qualitative manner the residual motion over a number of time frames (0s, 30s, 60s, 90s, 120s). I recorded videos after moving the stage and waiting for the indicated time and then compared the direct average image, and the aligned image with the highest estimated resolution using Gctf. The unaligned CTF estimation remains has a very low estimated resolution until 180s, which is not surprising as even for a perfectly stable stage there will always be some motion caused by the heating from the electron beam (Zheng et al., 2017). Rather, a better measure of residual frame movement would be the total frame movement (right portion of the figure). This shows that after 60s the residual frame movement has completely stopped the remaining movement is caused by the beam heating. From these data I tried to wait 60s to minimize the residual stage movement during data collection using the JEM-2200FS.

5.2.2 Selecting good images

Aligned images can be assessed for quality based on a number of criteria. Essentially the real micrograph and the the Fourier transform can be checked for consistency based on a number of criteria. Initially a visual assessment of the micrograph can provide some quality control, in particular if the beam is note centred then it is possible to see Fresnel fringes across the field of view. As the resolution of a micrograph (or conceptually the information content) can be assessed from checking the Fourier domain this is usually used to check the quality of the micrographs.

Evaluating the quality of micrographs can be based on a number of parameters. In particular the image should be checked in Fourier space to assess the resolution. As I used the estimated micrograph CTF parameters to reverse the phases, I also checked the alignment of the CTF estimation. I prepared a number of example micrographs to illustrate these points. **Figure 5.7** shows the result of a misalignment of the beam relative to the optical axis of the microscope. It is clear that the resolution is severely truncated at approx. 6 Angstroms and the alignment of the CTF is meaningless from that point. The following figure shows apparently a correctly aligned micrograph but the CTF estimation has not been very successful and the phases are misaligned from —Angstroms. Finally, for comparison I include a correctly aligned and CTF estimated micrograph. I need to get the CTF estimation parameters for sure.

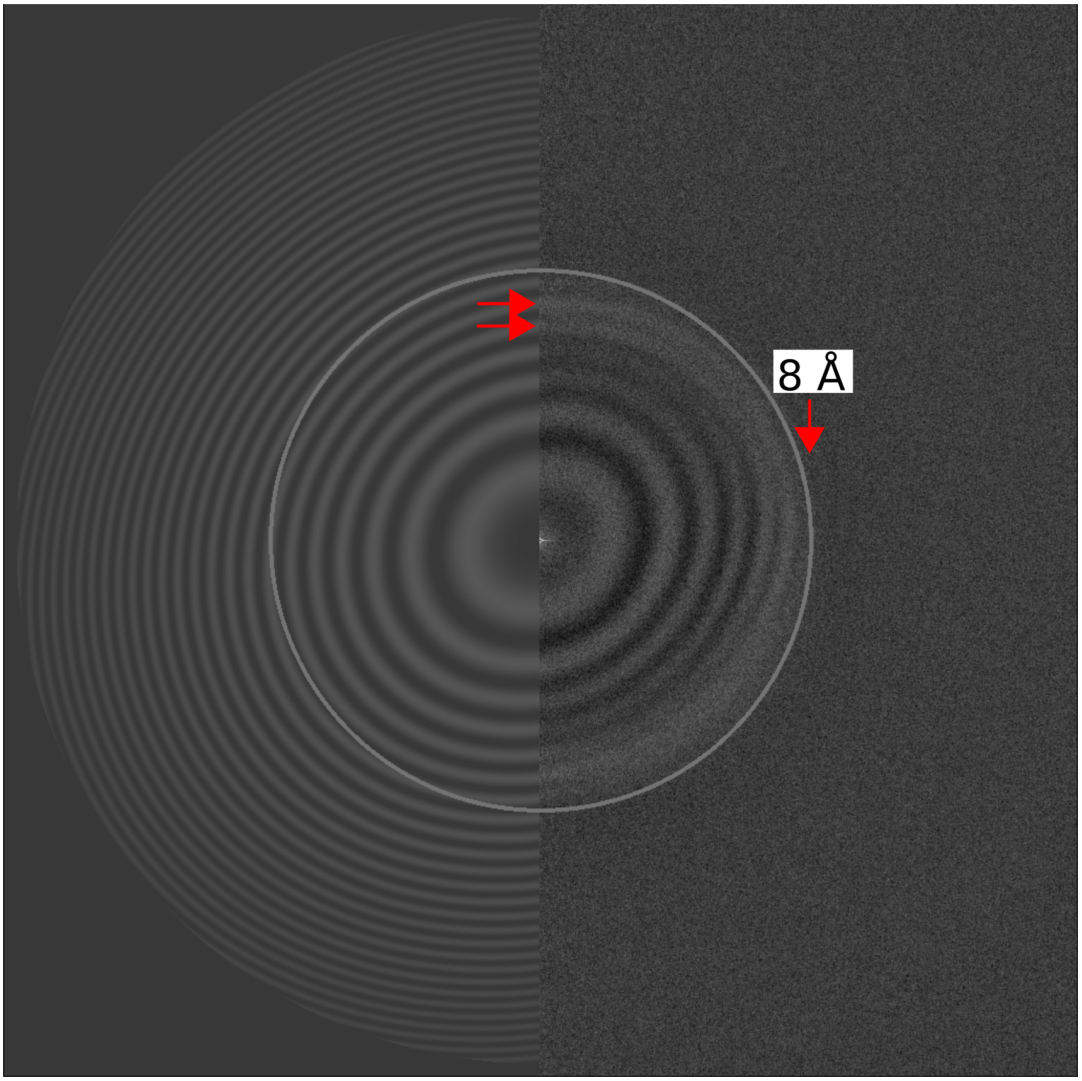


FIGURE 5.7: **CTF estimation: misaligned EM**. The Fourier transform (right) and Gctf (Zhang, 2016) estimation (left) of an image recorded when the EM beam alignment was not completed correctly on the JEM-2200FS. Take note of the astigmatic character of the Fourier transform, the low maximum estimated resolution (8 Å), and poor correlation between the real data and the estimated phases of the CTF (red arrows).

TABLE 5.2: Summary of data collection

Construct	**Movies**[A]	**visual**[B]	**Gctf**[C]	**Particles**[D]
Wild-type	182	173	173	2407
W2	116	116	113	1700
Calf1$_{XZ}$	155	113	113	3126
Calf1$_{EF}$	187	173	173	3311

[A]Total collected movies.
[B]After visual assessment.
[C]After checking estimated Gctf parameters.
[D]Total picked particles.

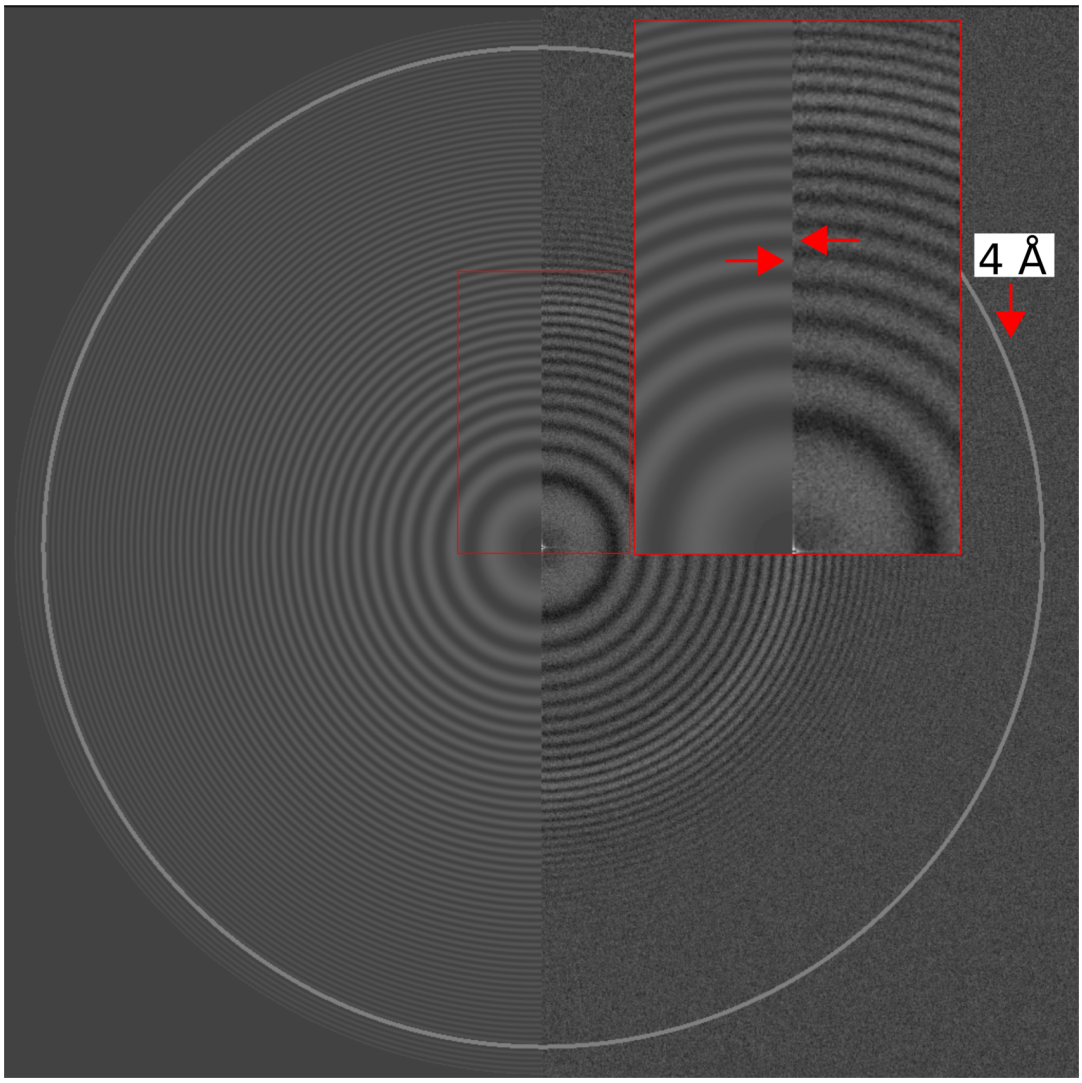


FIGURE 5.8: **CTF estimation: misaligned phase estimates**. The Fourier transform (right) and Gctf (Zhang, 2016) estimation (left) of an image recorded on the JEM-2200FS. Fourier space shows signal out to 4 Å but the CTF estimation was unable to correctly estimate the phases (inset, note mismatch between red arrows).

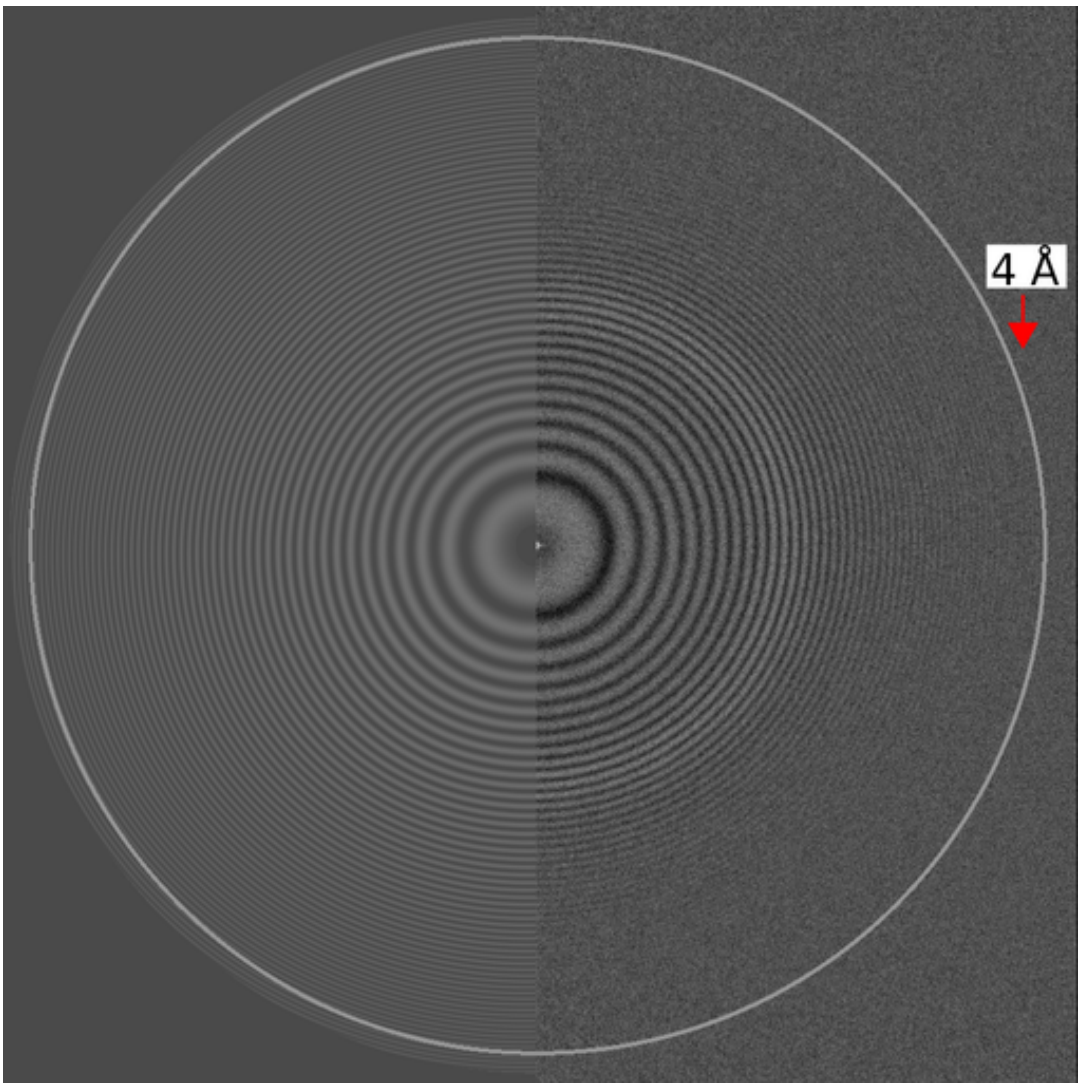


FIGURE 5.9: **CTF estimation: correct CTF parameters**. The Fourier transform (right) and Gctf (Zhang, 2016) estimation (left) of an image recorded on the JEM-2200FS. Fourier space shows signal out to 4 Å and the CTF estimation correctly matches the signal out to that range.

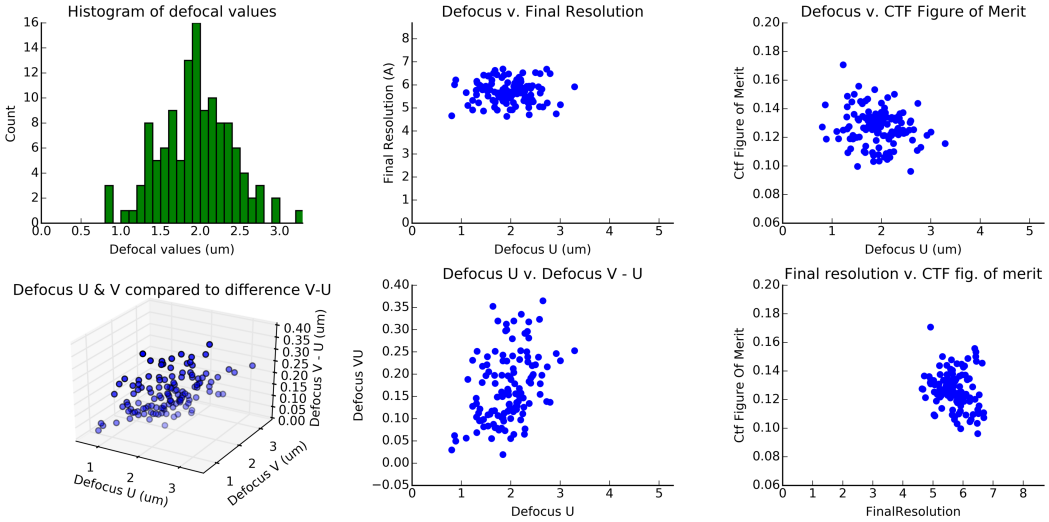


FIGURE 5.10: **Gctfestimation values of wildtype integrin collected on the JEM-2200FS**. A custom script was used to generate these outputs and select resolution and defocus ranges for the data as detailed below. (*top,left*) Histogram of the estimated defocus values for each aligned movie. (*top, middle*) Estimated defocus values compared with estimated resolution. (*top, right*) Defocus values compared with the CTF figure of merit. (*bottom, left*) Three dimensional plot of defocusU and defocusV compared with difference between defocusU and defocusV. (*bottom, middle*) 2D slice of the previous plot. (*bottom, right*) CTF figure of merit compared with final resolution.

5.3 All class averages

Figure 5.14 shows all 12 class averages.

5.4 Generating NZ-1 Fab

For EM labelling I used NZ-1 Fab that was prepared from full length antibody by digestion with papain, captured with PA-peptide bound sepharose and then purified with size exclusion chromatography as outlined in methods section 2.8.5. Papain digestion was completed over four hours and a sample collected every hour to confirm complete digestion of the NZ-1 antibody (Figure **??**). After the reaction time had proceeded past two hours there as no visible full length antibody, which shows that the Fab portion should be the predominant species.

Digested Fab was purified using a Superdex 200 16/60 (GE Healthcare, Chicago, IL) column and showed two peaks corresponding to estimated mass of 92 kDa and 31 kDa. Previous data suggested that the first 91 kDa peak was incorrectly folded Fab. SDS-PAGE of both peaks (**Figure 5.18**) confirmed that heavy and light

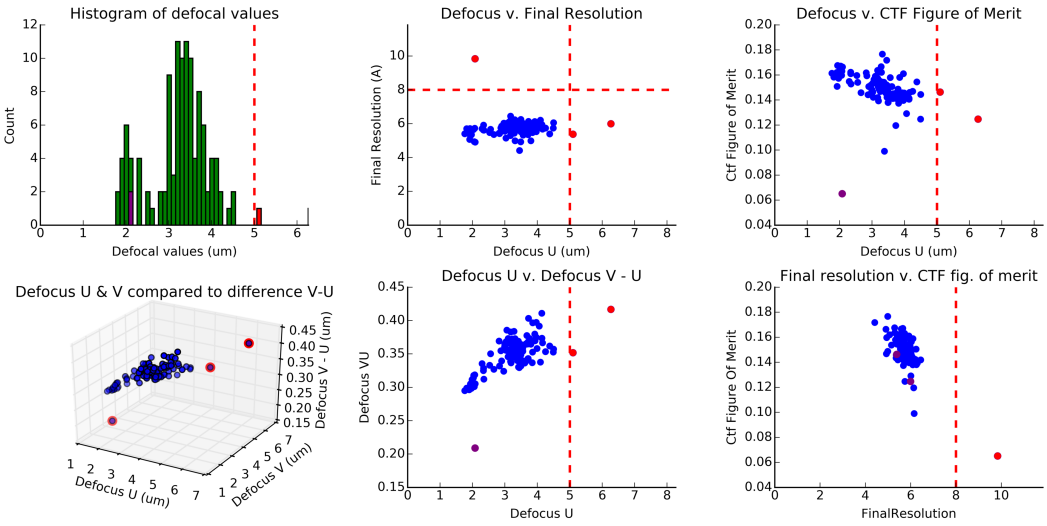


FIGURE 5.11: **Gctfestimation values of W2PA insert dataset collected on the JEM-2200FS.** A custom script was used to generate these outputs and select resolution and defocus ranges for the data as detailed below. In all cases purple indicates a movie removed based on some other criteria. A red-dashed line indicates some cut-off value that has been input. (*top,left*) Histogram of the estimated defocus values for each aligned movie. Movies with estimated defocus above 5 µm were removed (red). The purple column are movies that were removed in based on some other criteria. (*top, middle*) Estimated defocus values compared with estimated resolution. (*top, right*) Defocus values compared with the CTf figure of merit. (*bottom, left*) Three dimensional plot of defocusU and defocusV compared with difference between defocusU and defocusV. (*bottom, middle*) 2D slice of the previous plot. (*bottom, right*) CTF figure of merit compared with final resolution.

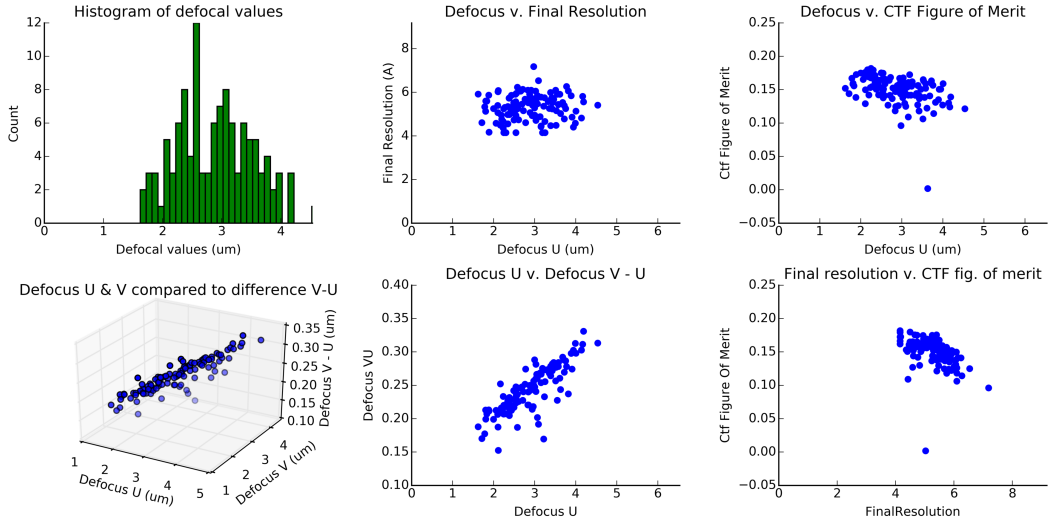


FIGURE 5.12: **Gctfestimation values of Calf1$_{XZ}$PA insert dataset collected on the JEM-2200FS.** A custom script was used to generate these outputs and select resolution and defocus ranges for the data as detailed below. (*top,left*) Histogram of the estimated defocus values for each aligned movie. (*top, middle*) Estimated defocus values compared with estimated resolution. (*top, right*) Defocus values compared with the CTF figure of merit. (*bottom, left*) Three dimensional plot of defocusU and defocusV compared with difference between defocusU and defocusV. (*bottom, middle*) 2D slice of the previous plot. (*bottom, right*) CTF figure of merit compared with final resolution.

chain from the Fab was present in both, but the first 91 kDa peak was oligomerised antibody.

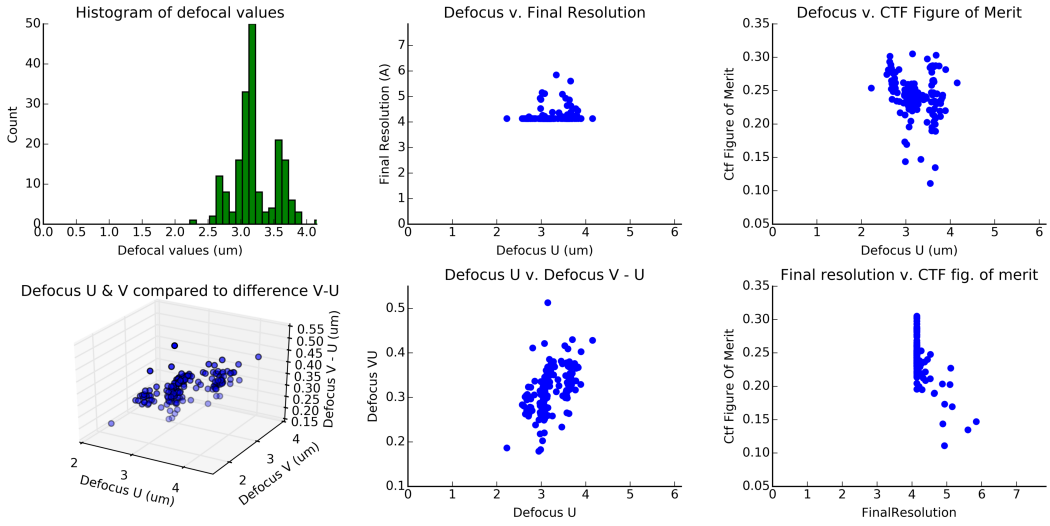


FIGURE 5.13: **Gctfestimation values of Calf1$_{EF}$ PA insert dataset collected on the JEM-2200FS.** A custom script was used to generate these outputs and select resolution and defocus ranges for the data as detailed below. (*top,left*) Histogram of the estimated defocus values for each aligned movie. (*top, middle*) Estimated defocus values compared with estimated resolution. (*top, right*) Defocus values compared with the CTF figure of merit. (*bottom, left*) Three dimensional plot of defocusU and defocusV compared with difference between defocusU and defocusV. (*bottom, middle*) 2D slice of the previous plot. (*bottom, right*) CTF figure of merit compared with final resolution.

(A) (B) (C) (D)

FIGURE 5.14: **Full class averages.** All 12 classes generated using reference free alignment as outlined in **Section 2.9.9**. (a) wild-type, (b) W2,(c) Calf1$_{XZ}$, (d) Calf1$_{EF}$

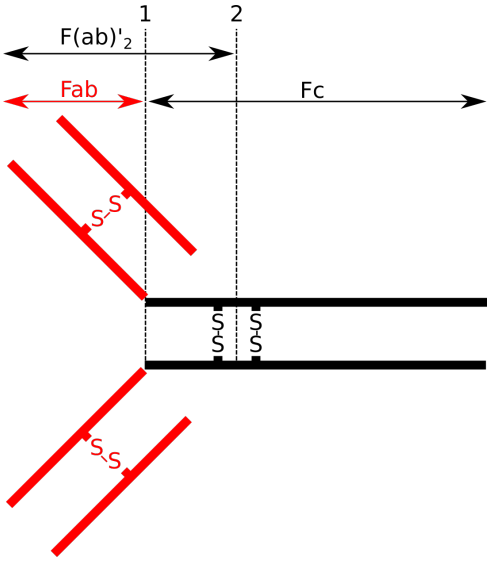
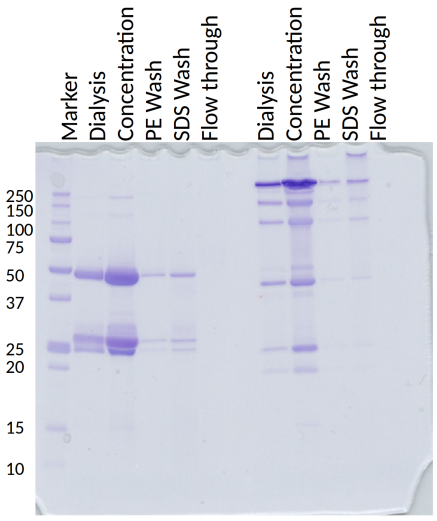
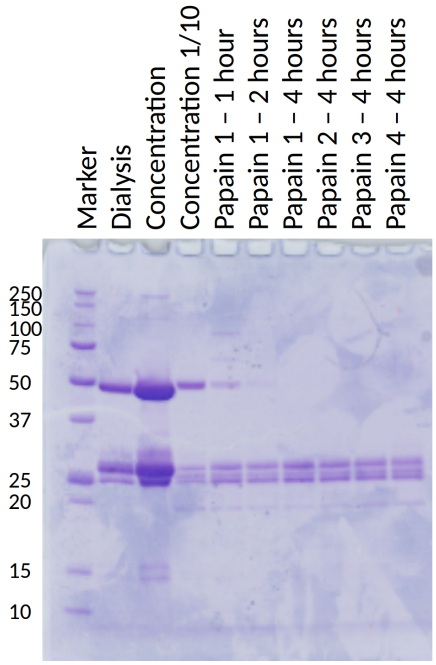


FIGURE 5.15: **Schematic of an antibody** showing the papain (1) and pepsin (2) cleavage sites that yield either Fab and Fc or F(ab)'$_2$ and truncated Fc.



(A) CBB stain non-reduce



(B) Oriole reduced

FIGURE 5.16: **NZ-1**. $\alpha_{IIb}\beta_3$integrin was purified using NZ-1 bound to sepharose and concentrated before injection into SEC. The purified fractions were run on an SDS-PAGE and stained with oriole and then CBB.

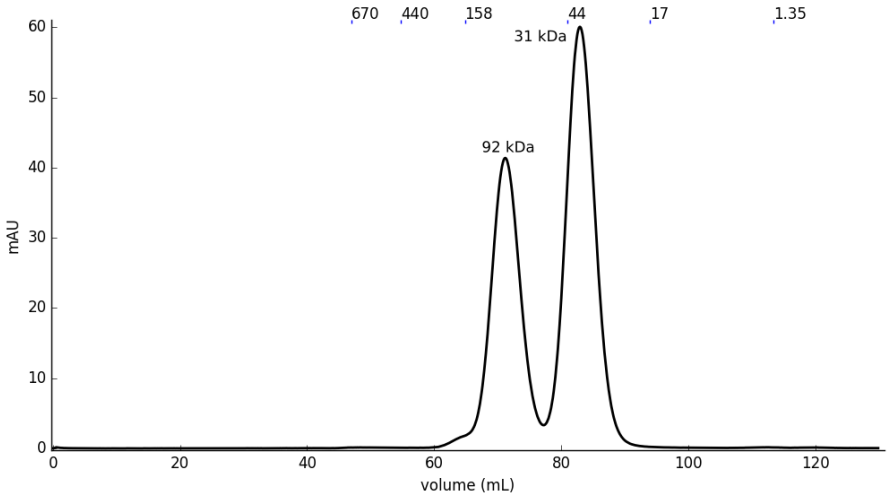
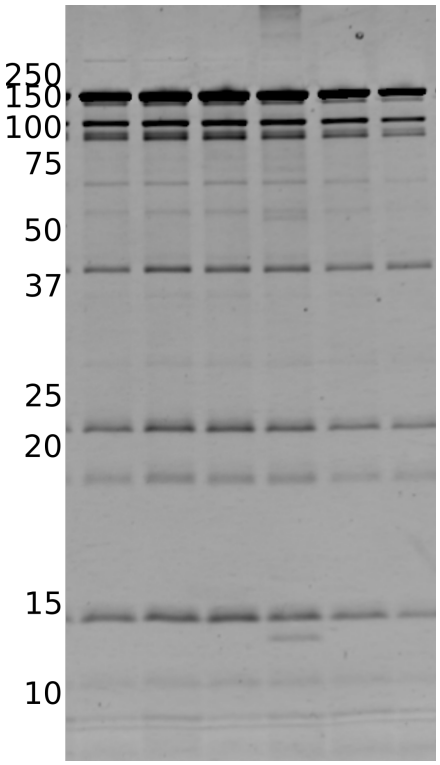
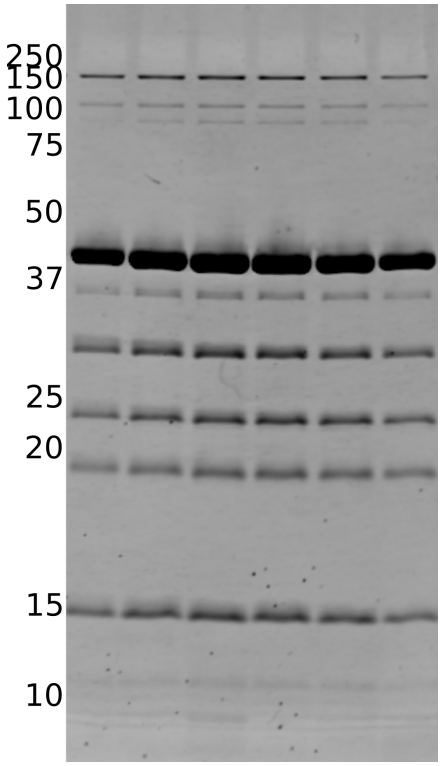


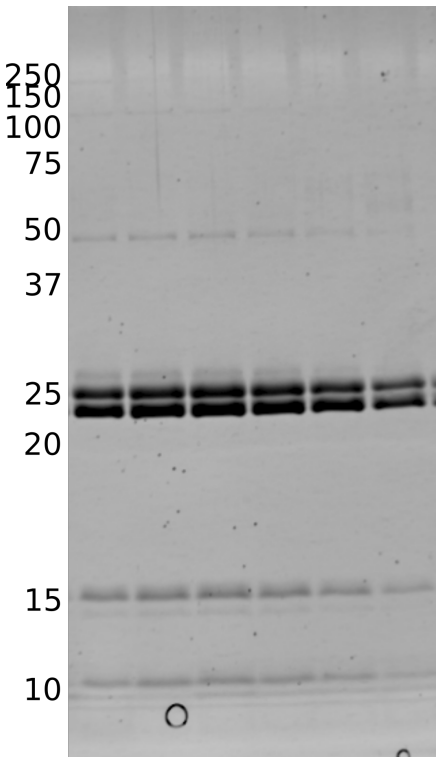
FIGURE 5.17: **AKTA prime of NZ-1 Fab purification**. Papain digested NZ-1 was purified using gel filtration to collect the correctly folded portion.



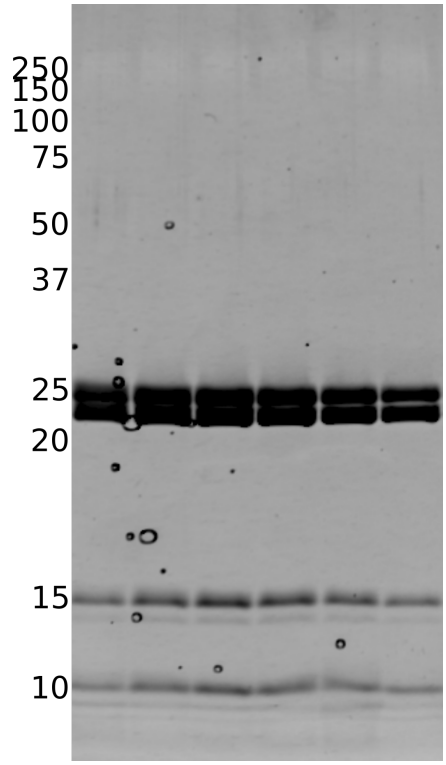
(A) First peak, 92 KDa (NR)



(B) Second peak, 31 kDa (NR)



(C) First peak, 92 KDa (R)



(D) Second peak, 31 kDa (R)

FIGURE 5.18: **From the AKATA prime**. The non-reduced (NR) and reduced (R) SDS-PAGE gels of SEC purified NZ-1 Fab.

Bibliography

Absolom, D R and C J van Oss (1986). "The nature of the antigen-antibody bond and the factors affecting its association and dissociation." In: *CRC Critical reviews in immunology* 6.1, pp. 1–46. ISSN: 01973355.

Arimori, Takao et al. (2017). "Fv-clasp: An Artificially Designed Small Antibody Fragment with Improved Production Compatibility, Stability, and Crystallizability". In: *Structure* 25.10, 1611–1622.e4. ISSN: 18784186. DOI: 10.1016/j.str.2017.08.011. URL: https://doi.org/10.1016/j.str.2017.08.011.

Boisset, N et al. (1995). "Three-dimensional reconstruction of Androctonus australis hemocyanin labeled with a monoclonal Fab fragment." In: *Journal of structural biology* 115.1, pp. 16–29. ISSN: 1047-8477. DOI: 10.1006/jsbi.1995.1025. URL: http://www.ncbi.nlm.nih.gov/pubmed/7577229.

Boisset, Nicolas et al. (1993). *Three-Dimensional Immunoelectron Microscopy of Scorpion Hemocyanin Labeled with a Monoclonal Fab Fragment*. DOI: 10.1006/jsbi.1993.1053. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847783710531.

Bonasio, R. et al. (2007). "Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots". In: *Proceedings of the National Academy of Sciences* 104.37, pp. 14753–14758. ISSN: 0027-8424. DOI: 10.1073/pnas.0705201104. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0705201104.

Bradley, D. E. (1954). "Evaporated carbon films for use in electron microscopy". In: *British Journal of Applied Physics* 5.2, p. 65. ISSN: 0508-3443. DOI: 10.1088/0508-3443/5/2/304. URL: http://stacks.iop.org/0508-3443/5/i=2/a=304?key=crossref.e3494c22302856e2334abb72e3117c78.

Brown, Zuben P. et al. (2017). "Development of a new protein labeling system to map subunits and domains of macromolecular complexes for electron microscopy". In: *Journal of Structural Biology*. ISSN: 10478477. DOI: 10.1016/j.jsb.2017.11.006. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847717301958.

Büchel, Claudia et al. (2001). "Localisation of the PsbH subunit in photosystem II: a new approach using labelling of his-tags with a Ni2+-NTA gold cluster and single particle analysis". In: *Journal of Molecular Biology* 312.2, pp. 371–379. ISSN: 00222836. DOI: 10.1006/jmbi.2001.4951. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022283601949514.

Bui, Khanh Huy et al. (2008). "Molecular architecture of inner dynein arms in situ in Chlamydomonas reinhardtii flagella". In: *Journal of Cell Biology* 183.5, pp. 923–

932. ISSN: 00219525. DOI: 10.1083/jcb.200808050. URL: http://www.jcb.org/lookup/doi/10.1083/jcb.200808050.

Calleja, V et al. (2003). "Monitoring conformational changes of proteins in cells by fluorescence lifetime imaging microscopy." In: *Biochemical Journal* 40, pp. 33–40. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1223389/.

Cantor, Eric J. and Shaorong Chong (2001). "Intein-Mediated Rapid Purification of Cre Recombinase". In: *Protein Expression and Purification* 22.1, pp. 135–140. ISSN: 10465928. DOI: 10.1006/prep.2001.1428. URL: http://linkinghub.elsevier.com/retrieve/pii/S1046592801914287.

Chen, Jieming, Nicholas Sawyer, and Lynne Regan (2013). "Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area". In: *Protein Science* 22.4, pp. 510–515. ISSN: 09618368. DOI: 10.1002/pro.2230.

Ciferri, Claudio et al. (2012). "Molecular architecture of human polycomb repressive complex 2". In: *eLife* 2012.1, pp. 1–22. ISSN: 2050084X. DOI: 10.7554/eLife.00005.

Ciferri, Claudio, Gabriel C. Lander, and Eva Nogales (2015). "Protein domain mapping by internal labeling and single particle electron microscopy". In: *Journal of Structural Biology* 192.2, pp. 159–162. ISSN: 10958657. DOI: 10.1016/j.jsb.2015.09.016. URL: http://www.sciencedirect.com/science/article/pii/S104784771530071X.

Computing, R Foundation for Statistical Vienna, Austria (2013). *R: A language and environment for statistical computing*. URL: http://www.r-project.org/.

Connelly, Patrick R. et al. (1990). "Thermodynamics of Protein-Peptide Interactions in the Ribonuclease S System Studied by Titration Calorimetry". In: *Biochemistry* 29.25, pp. 6108–6114. ISSN: 15204995. DOI: 10.1021/bi00477a031.

Dambacher, Corey M. and Gabriel C. Lander (2015). "Site-specific labeling of proteins for electron microscopy". In: *Journal of Structural Biology* 192.2, pp. 151–158. ISSN: 10958657. DOI: 10.1016/j.jsb.2015.09.010. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847715300654.

Danev, R. et al. (2014). "Volta potential phase plate for in-focus phase contrast transmission electron microscopy". In: *Proceedings of the National Academy of Sciences* 111.44, pp. 15635–15640. ISSN: 0027-8424. DOI: 10.1073/pnas.1418377111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1418377111.

Danev, Radostin, Dimitry Tegunov, and Wolfgang Baumeister (2017). "Using the Volta phase plate with defocus for cryo-EM single particle analysis". In: *eLife* 6. ISSN: 2050-084X. DOI: 10.7554/eLife.23006. URL: http://elifesciences.org/lookup/doi/10.7554/eLife.23006https://elifesciences.org/articles/23006.

Dennison, S. M. et al. (2014). "Vaccine-Induced HIV-1 Envelope gp120 Constant Region 1-Specific Antibodies Expose a CD4-Inducible Epitope and Block the Interaction of HIV-1 gp140 with Galactosylceramide". In: *Journal of Virology* 88.16,

pp. 9406–9417. ISSN: 0022-538X. DOI: 10.1128/JVI.01031-14. URL: http://jvi.asm.org/cgi/doi/10.1128/JVI.01031-14.

Dinculescu, Astra et al. (2002). "Insertional mutagenesis and immunochemical analysis of visual arrestin interaction with rhodopsin". In: *Journal of Biological Chemistry* 277.14, pp. 11703–11708. ISSN: 00219258. DOI: 10.1074/jbc.M111833200.

Dyson, H Jane, Richard A Lerner, and Peter E Wright (1988). "The Physical Basis for Induction of Protein-Reactive Antipeptide Antibodies". In: *Annual Review of Biophysics and Biophysical Chemistry* 17.1, pp. 305–324. ISSN: 0883-9182. DOI: 10.1146/annurev.bb.17.060188.001513. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.bb.17.060188.001513.

Eng, Edward T. et al. (2011). "Intact α IIbβ 3 integrin is extended after activation as measured by solution x-ray scattering and electron microscopy". In: *Journal of Biological Chemistry* 286.40, pp. 35218–35226. ISSN: 00219258. DOI: 10.1074/jbc.M111.275107. URL: http://www.jbc.org/lookup/doi/10.1074/jbc.M111.275107.

Erickson, H. P. and A. Klug (1971). "Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 261.837, pp. 105–118. ISSN: 0962-8436. DOI: 10.1098/rstb.1971.0040. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.1971.0040.

Evan, G I et al. (1985). "Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product." In: *Molecular and Cellular Biology* 5.12, pp. 3610–3616. ISSN: 0270-7306. DOI: 10.1128/MCB.5.12.3610. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=369192{\&}tool=pmcentrez{\&}rendertype=abstracthttp://mcb.asm.org/lookup/doi/10.1128/MCB.5.12.3610.

Facey, Sandra J. and Andreas Kuhn (2003). "The sensor protein KdpD inserts into the Escherichia coli membrane independent of the Sec translocase and YidC". In: *European Journal of Biochemistry* 270.8, pp. 1724–1734. ISSN: 00142956. DOI: 10.1046/j.1432-1033.2003.03531.x.

Fellouse, Frederic A. et al. (2007). "High-throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-displayed Libraries". In: *Journal of Molecular Biology* 373.4, pp. 924–940. ISSN: 00222836. DOI: 10.1016/j.jmb.2007.08.005. arXiv: NIHMS150003.

Field, J et al. (1988). "Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method." In: *Molecular and cellular biology* 8.5, pp. 2159–65. ISSN: 0270-7306. DOI: 10.1128/MCB.8.5.2159. URL: http://www.ncbi.nlm.nih.gov/pubmed/2455217{\%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC363397.

Flemming, Dirk et al. (2010). "Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label". In: *Nature Structural & Molecular Biology* 17.6, pp. 775–778. ISSN: 1545-9993. DOI: 10.1038/nsmb.1811. URL: http://www.nature.com/doifinder/10.1038/nsmb.1811.

Frank, Joachim (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, pp. 1–456. ISBN: 9780195182187. DOI: 10.1093/acprof:oso/9780195182187.001.0001. arXiv: arXiv:1011.1669v3. URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195182187.001.0001/acprof-9780195182187.

Fujii, Yuki et al. (2014). "PA tag: A versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin". In: *Protein Expression and Purification* 95, pp. 240–247. ISSN: 10465928. DOI: 10.1016/j.pep.2014.01.009. URL: http://www.ncbi.nlm.nih.gov/pubmed/24480187.

Fujii, Yuki, Mika K. Kaneko, and Yukinari Kato (2016a). "MAP Tag: A Novel Tagging System for Protein Purification and Detection". In: *Monoclonal Antibodies in Immunodiagnosis and Immunotherapy* 35.6, pp. 293–299. ISSN: 2167-9436. DOI: 10.1089/mab.2016.0039. URL: http://online.liebertpub.com/doi/10.1089/mab.2016.0039.

Fujii, Yuki et al. (2016b). "Tailored placement of a turn-forming PA tag into the structured domain of a protein to probe its conformational state". In: *Journal of Cell Science* 129.7, pp. 1512–1522. ISSN: 0021-9533. DOI: 10.1242/jcs.176685. URL: http://jcs.biologists.org/lookup/doi/10.1242/jcs.176685.

Fujii, Yuki et al. (2017). "Development of RAP Tag, a Novel Tagging System for Protein Detection and Purification". In: *Monoclonal Antibodies in Immunodiagnosis and Immunotherapy* 36.2, pp. 68–71. ISSN: 2167-9436. DOI: 10.1089/mab.2016.0052. URL: http://online.liebertpub.com/doi/10.1089/mab.2016.0052.

Glaeser, R. M. (2016). "Specimen Behavior in the Electron Beam". In: *Methods in Enzymology*, pp. 19–50. ISBN: 9780128053829. DOI: 10.1016/bs.mie.2016.04.010. URL: http://linkinghub.elsevier.com/retrieve/pii/S0076687916300283.

Guillén, D. et al. (2013). "The starch-binding domain as a tool for recombinant protein purification". In: *Applied Microbiology and Biotechnology* 97.9, pp. 4141–4148. ISSN: 14320614. DOI: 10.1007/s00253-013-4778-0.

Guruprasad, K and S Rajkumar (2000). "Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials." In: *Journal of biosciences* 25.2, pp. 143–156. ISSN: 0250-5991.

Hainfeld, James F. and Richard D. Powell (2000). "New Frontiers in Gold Labeling". In: *Journal of Histochemistry & Cytochemistry* 48.4, pp. 471–480. ISSN: 0022-1554. DOI: 10.1177/002215540004800404. URL: http://journals.sagepub.com/doi/10.1177/002215540004800404.

Hancock, David C and Nicola J O'Reilly (2005). "Synthetic peptides as antigens for antibody production." In: *Methods in molecular biology (Clifton, N.J.)* Methods in Molecular Biology™ 295.C. Ed. by John D. Pound, pp. 13–26. ISSN: 1064-3745. DOI: 10.1007/978-1-59259-257-9. URL: http://www.springerlink.com/index/10.1007/978-1-59259-257-9http://link.springer.com/10.1007/978-1-59259-257-9http://www.ncbi.nlm.nih.gov/pubmed/15596885.

Hawkes, Peter W. (2006). "The electron microscope as a structure projector". In: *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, pp. 83–111. ISBN: 9780387690087. DOI: 10.1007/978-0-387-69008-7_4. arXiv: arXiv:1011.1669v3.

Heel, M van (1984). "Multivariate statistical classification of noisy images (randomly oriented biological macromolecules)." In: *Ultramicroscopy* 13.1-2, pp. 165–83. ISSN: 0304-3991. URL: http://www.ncbi.nlm.nih.gov/pubmed/6382731.

Heel, Marin van and Wilko Keegstra (1981). "IMAGIC: A fast, flexible and friendly image analysis software system". In: *Ultramicroscopy* 7.2, pp. 113–129. ISSN: 03043991. DOI: 10.1016/0304-3991(81)90001-2.

Heel, Marin van et al. (1996). "A New Generation of the IMAGIC Image Processing System". In: *Journal of Structural Biology* 116.1, pp. 17–24. ISSN: 10478477. DOI: 10.1006/jsbi.1996.0004. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847796900040.

Heel, Marin van, Rodrigo V. Portugal, and Michael Schatz (2016). "Multivariate Statistical Analysis of Large Datasets: Single Particle Electron Microscopy". In: *Open Journal of Statistics* 06.04, pp. 701–739. ISSN: 2161-718X. DOI: 10.4236/ojs.2016.64059. URL: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojs.2016.64059.

Hegerl, R and W Hoppe (1976). "Influence of electron noise on three-dimensional image reconstruction". In: *Zeitschrift für Naturforschung Teil A* 31.4, pp. 1717–1721. ISSN: 1865-7109. DOI: 10.1515/zna-1976-1241. URL: http://adsabs.harvard.edu/abs/1976ZNatA..31.1717H.

Henderson, Richard (1995). "The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules". In: *Quarterly Reviews of Biophysics* 28.02, p. 171. ISSN: 0033-5835. DOI: 10.1017/S003358350000305X. arXiv: arXiv:1011.1669v3. URL: http://www.journals.cambridge.org/abstract{_}S003358350000305X.

Henderson, Richard and Greg McMullan (2013). "Problems in obtaining perfect images by single-particle electron cryomicroscopy of biological structures in amorphous ice". In: *Journal of Electron Microscopy* 62.1, pp. 43–50. ISSN: 00220744. DOI: 10.1093/jmicro/dfs094. URL: https://academic.oup.com/jmicro/article-lookup/doi/10.1093/jmicro/dfs094.

Henderson, Richard et al. (2011). "Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy". In: *Journal of Molecular Biology* 413.5, pp. 1028–1046. ISSN: 00222836. DOI: 10.1016/j.jmb.2011.09.008. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022283611009983.

Hernan, Ron, Ken Heuermann, and Bill Brizzard (2000). "Multiple epitope tagging of expressed proteins for enhanced detection". In: *BioTechniques* 28.4, pp. 789–793. ISSN: 07366205.

Heuser, Thomas et al. (2009). "The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella". In: *Journal of Cell Biology* 187.6,

pp. 921–933. ISSN: 00219525. DOI: 10.1083/jcb.200908067. URL: http://www.jcb.org/lookup/doi/10.1083/jcb.200908067.

Heuser, Thomas et al. (2012). "Cryoelectron tomography reveals doublet-specific structures and unique interactions in the I1 dynein". In: *Proceedings of the National Academy of Sciences* 109.30, E2067–E2076. ISSN: 0027-8424. DOI: 10.1073/pnas.1120690109. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1120690109.

Hirai, Hidenori et al. (2017). "Structural basis for ligand capture and release by the endocytic receptor ApoER2". In: *EMBO reports* 18.6, pp. 982–999. ISSN: 1469-221X. DOI: 10.15252/embr.201643521. URL: http://embor.embopress.org/lookup/doi/10.15252/embr.201643521.

Hochuli, E. et al. (1988). "Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent". In: *Nature Biotechnology* 6.11, pp. 1321–1325. ISSN: 1087-0156. DOI: 10.1038/nbt1188-1321. URL: http://www.nature.com/doifinder/10.1038/nbt1188-1321.

Hopp, Thomas P. et al. (1988). *A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification*. DOI: 10.1038/nbt1088-1204. arXiv: arXiv:1011.1669v3. URL: http://www.nature.com/doifinder/10.1038/nbt1088-1204.

Humphries, Martin J., Emlyn J.H. Symonds, and A. Paul Mould (2003). "Mapping functional residues onto integrin crystal structures". In: *Current Opinion in Structural Biology* 13.2, pp. 236–243. ISSN: 0959440X. DOI: 10.1016/S0959-440X(03)00035-6.

Irannejad, Roshanak et al. (2013). "Conformational biosensors reveal GPCR signalling from endosomes". In: *Nature* 495.7442, pp. 534–538. ISSN: 0028-0836. DOI: 10.1038/nature12000. URL: http://www.nature.com/doifinder/10.1038/nature12000.

Jiang, Qiu-Xing, Da-Neng Wang, and Roderick MacKinnon (2004). "Electron microscopic analysis of KvAP voltage-dependent K+ channels in an open conformation". In: *Nature* 430.7001, pp. 806–810. ISSN: 0028-0836. DOI: 10.1038/nature02735. URL: http://www.nature.com/doifinder/10.1038/nature02735.

Kato, Yukinari et al. (2006). "Inhibition of tumor cell-induced platelet aggregation using a novel anti-podoplanin antibody reacting with its platelet-aggregation-stimulating domain." In: *Biochemical and biophysical research communications* 349.4, pp. 1301–7. ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2006.08.171. URL: http://www.ncbi.nlm.nih.gov/pubmed/16979138.

Keefe, Anthony D. et al. (2001). "One-Step Purification of Recombinant Proteins Using a Nanomolar-Affinity Streptavidin-Binding Peptide, the SBP-Tag". In: *Protein Expression and Purification* 23.3, pp. 440–446. ISSN: 10465928. DOI: 10.1006/prep.2001.1515. URL: http://linkinghub.elsevier.com/retrieve/pii/S1046592801915153.

Kelly, Deborah F. et al. (2010). "Molecular structure and dimeric organization of the notch extracellular domain as revealed by electron microscopy". In: *PLoS ONE* 5.5, pp. 1–6. ISSN: 19326203. DOI: 10.1371/journal.pone.0010532.

Kendall, Ryan T. and Susan E. Senogles (2006). "Investigation of the alternatively spliced insert region of the D 2L dopamine receptor by epitope substitution". In: *Neuroscience Letters* 393.2-3, pp. 155–159. ISSN: 03043940. DOI: 10.1016/j.neulet.2005.09.057.

Koeck, P. J.B. et al. (1996). "Unconventional immuno double labelling by energy filtered transmission electron microscopy". In: *Ultramicroscopy* 62.1-2, pp. 65–78. ISSN: 03043991. DOI: 10.1016/0304-3991(95)00088-7.

Lammert, Dawn B. et al. (2017). "The de novo autism spectrum disorder RELN R2290C mutation reduces Reelin secretion and increases protein disulfide isomerase expression". In: *Journal of Neurochemistry* 142.1, pp. 89–102. ISSN: 00223042. DOI: 10.1111/jnc.14045. URL: http://doi.wiley.com/10.1111/jnc.14045.

Lau, Pick Wei et al. (2012). "DOLORS: Versatile strategy for internal labeling and domain localization in electron microscopy". In: *Structure* 20.12, pp. 1995–2002. ISSN: 09692126. DOI: 10.1016/j.str.2012.10.019. URL: http://dx.doi.org/10.1016/j.str.2012.10.019.

Lees, Joshua A et al. (2010). "Molecular organization of the COG vesicle tethering complex". In: *Nature Structural & Molecular Biology* 17.11, pp. 1292–1297. ISSN: 1545-9993. DOI: 10.1038/nsmb.1917. arXiv: NIHMS150003. URL: http://www.nature.com/doifinder/10.1038/nsmb.1917.

Lin, Jianfeng et al. (2012). "One of the Nine Doublet Microtubules of Eukaryotic Flagella Exhibits Unique and Partially Conserved Structures". In: *PLoS ONE* 7.10, pp. 27–30. ISSN: 19326203. DOI: 10.1371/journal.pone.0046494.

Lyumkis, Dmitry et al. (2013). "Likelihood-based classification of cryo-EM images using FREALIGN". In: *Journal of Structural Biology* 183.3, pp. 377–388. ISSN: 10478477. DOI: 10.1016/j.jsb.2013.07.005. arXiv: NIHMS150003. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847713001858.

Maina, Claude V. et al. (1988). "An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein". In: *Gene* 74.2, pp. 365–373. ISSN: 03781119. DOI: 10.1016/0378-1119(88)90170-9.

Mallick, Satya P. et al. (2005). "ACE: Automated CTF estimation". In: *Ultramicroscopy* 104.1, pp. 8–29. ISSN: 03043991. DOI: 10.1016/j.ultramic.2005.02.004. URL: http://www.ncbi.nlm.nih.gov/pubmed/15935913.

Matoba, Kyoko et al. (2017). "Conformational Freedom of the LRP6 Ectodomain Is Regulated by N-glycosylation and the Binding of the Wnt Antagonist Dkk1". In: *Cell Reports* 18.1, pp. 32–40. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.12.017. URL: http://linkinghub.elsevier.com/retrieve/pii/S2211124716317077http://dx.doi.org/10.1016/j.celrep.2016.12.017.

Matsunaga, Yukiko et al. (2016). "Allosteric Inhibition of a Semaphorin 4D Receptor Plexin B1 by a High-Affinity Macrocyclic Peptide". In: *Cell Chemical Biology* 23.11,

pp. 1341–1350. ISSN: 24519456. DOI: 10.1016/j.chembiol.2016.09.015. URL: http://linkinghub.elsevier.com/retrieve/pii/S2451945616303555.

McMullan, G. et al. (2014). "Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy". In: *Ultramicroscopy* 147, pp. 156–163. ISSN: 03043991. DOI: 10.1016/j.ultramic.2014.08.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S030439911400151X.

Meng, Hongrui et al. (2017). "Loss of Parkinson's disease-associated protein CHCHD2 affects mitochondrial crista structure and destabilizes cytochrome c". In: *Nature Communications* 8, p. 15500. ISSN: 2041-1723. DOI: 10.1038/ncomms15500. URL: http://www.nature.com/doifinder/10.1038/ncomms15500.

Mercogliano, Christopher P. and David J. DeRosier (2007). "Concatenated metallothionein as a clonable gold label for electron microscopy". In: *Journal of Structural Biology* 160.1, pp. 70–82. ISSN: 10478477. DOI: 10.1016/j.jsb.2007.06.010. arXiv: NIHMS150003.

Merk, Alan et al. (2016). "Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery". In: *Cell* 165.7, pp. 1698–1707. ISSN: 10974172. DOI: 10.1016/j.cell.2016.05.040. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867416305918.

Mihara, Emiko et al. (2016). "Active and water-soluble form of lipidated Wnt protein is maintained by a serum glycoprotein afamin/α-albumin". In: *eLife* 5. ISSN: 2050-084X. DOI: 10.7554/eLife.11621. URL: http://elifesciences.org/lookup/doi/10.7554/eLife.11621.

Morlacchi, Simona et al. (2012). "Insertion of a myc-tag within α-dystroglycan domains improves its biochemical and microscopic detection". In: *BMC Biochemistry* 13.1, p. 14. ISSN: 1471-2091. DOI: 10.1186/1471-2091-13-14. URL: http://www.biomedcentral.com/1471-2091/13/14.

Mould, A P, S K Akiyama, and M J Humphries (1995). "Regulation of integrin alpha 5 beta 1-fibronectin interactions by divalent cations. Evidence for distinct classes of binding sites for Mn2+, Mg2+, and Ca2+." In: *The Journal of biological chemistry* 270.44, pp. 26270–7. ISSN: 0021-9258. DOI: 10.1074/JBC.270.44.26270. URL: http://www.ncbi.nlm.nih.gov/pubmed/7592835.

Nishino, Yuri, Takuo Yasunaga, and Atsuo Miyazawa (2007). "A genetically encoded metallothionein tag enabling efficient protein detection by electron microscopy". In: *Journal of Electron Microscopy* 56.3, pp. 93–101. ISSN: 00220744. DOI: 10.1093/jmicro/dfm008.

Nogi, Terukazu et al. (2008). "Novel affinity tag system using structurally defined antibody-tag interaction: application to single-step protein purification." In: *Protein science : a publication of the Protein Society* 17.12, pp. 2120–2126. ISSN: 0961-8368. DOI: 10.1110/ps.038299.108.

Oda, Toshiyuki and Masahide Kikkawa (2013). "Novel structural labeling method using cryo-electron tomography and biotin-streptavidin system". In: *Journal of*

Structural Biology 183.3, pp. 305–311. ISSN: 10478477. DOI: 10.1016/j.jsb.2013.07.003. URL: http://dx.doi.org/10.1016/j.jsb.2013.07.003.

Orloff, Jon (2008). *Handbook of Charged Particle Optics, Second Edition (Google eBook)*. CRC Press, p. 666. ISBN: 1420045555. DOI: 10.1201/9781420045550. URL: http://books.google.com/books?id=y0FF19lud5YC{\&}pgis=1.

O'Shea, Erin K., Kevin J. Lumb, and Peter S. Kim (1993). "Peptide 'Velcro': Design of a heterodimeric coiled coil". In: *Current Biology* 3.10, pp. 658–667. ISSN: 09609822. DOI: 10.1016/0960-9822(93)90063-T.

Pazour, Gregory J. et al. (2005). "Proteomic analysis of a eukaryotic cilium". In: *Journal of Cell Biology* 170.1, pp. 103–113. ISSN: 00219525. DOI: 10.1083/jcb.200504008. URL: http://www.jcb.org/lookup/doi/10.1083/jcb.200504008.

Pigino, Gaia et al. (2011). "Cryoelectron tomography of radial spokes in cilia and flagella". In: *Journal of Cell Biology* 195.4, pp. 673–687. ISSN: 00219525. DOI: 10.1083/jcb.201106125.

Porter, Mary E. and Winfield S. Sale (2000). "The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility". In: *Journal of Cell Biology* 151.5, F37–F42. ISSN: 00219525. DOI: 10.1083/jcb.151.5.F37. URL: http://www.jcb.org/lookup/doi/10.1083/jcb.151.5.F37.

Prasad, B V et al. (1990). "Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy." In: *Nature* 343.6257, pp. 476–9. ISSN: 0028-0836. DOI: 10.1038/343476a0. URL: http://www.ncbi.nlm.nih.gov/pubmed/2153941.

Radon, Johann (1986). "On the determination of functions from their integral values along certain manifolds". In: *IEEE Transactions on Medical Imaging* 5.4, pp. 170–176. ISSN: 0278-0062. DOI: 10.1109/TMI.1986.4307775. URL: http://ieeexplore.ieee.org/document/4307775/.

Reimer and H Kohl (2008). *Transmission Electron Microscopy*. Vol. 36. Springer Series in Optical Sciences. New York, NY: Springer New York. ISBN: 978-0-387-40093-8. DOI: 10.1007/978-0-387-40093-8. URL: http://link.springer.com/10.1007/978-0-387-40093-8.

Rohou, Alexis and Nikolaus Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs". In: *Journal of Structural Biology* 192.2, pp. 216–221. ISSN: 10958657. DOI: 10.1016/j.jsb.2015.08.008. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847715300460.

Roosild, Tarmo P., Samantha Castronovo, and Senyon Choe (2006). "Structure of anti-FLAG M2 Fab domain and its use in the stabilization of engineered membrane proteins". In: *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 62.9, pp. 835–839. ISSN: 17443091. DOI: 10.1107/S1744309106029125.

Samsó, Montserrat et al. (1999). "Three-dimensional location of the imperatoxin A binding site on the ryanodine receptor". In: *Journal of Cell Biology* 146.2, pp. 493–499. ISSN: 00219525. DOI: 10.1083/jcb.146.2.493.

Sangawa, Takeshi et al. (2013). "A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein." In: *Protein science : a publication of the Protein Society* 22.6, pp. 840–50. ISSN: 1469-896X. DOI: 10.1002/pro.2254. URL: http://www.ncbi.nlm.nih.gov/pubmed/23526492.

Sassenfeld, Helmut M and Stephen J Brewer (1984). "A polypeptide fusion designed for the purification of recombinant proteins". In: *Nature biotechnology* 2, pp. 76–81. ISSN: 0733-222X. DOI: 10.1038/nbt0184-76.

Scheres, S. H.W. (2016). "Processing of Structurally Heterogeneous Cryo-EM Data in RELION". In: *Methods in Enzymology*, pp. 125–157. ISBN: 9780128053829. DOI: 10.1016/bs.mie.2016.04.012. URL: http://linkinghub.elsevier.com/retrieve/pii/S0076687916300301.

Scheres, Sjors H W (2012). "RELION: Implementation of a Bayesian approach to cryo-EM structure determination". In: *Journal of Structural Biology* 180.3, pp. 519–530. ISSN: 10478477. DOI: 10.1016/j.jsb.2012.09.006. URL: http://dx.doi.org/10.1016/j.jsb.2012.09.006.

Scherzer, O. (1949). "The theoretical resolution limit of the electron microscope". In: *Journal of Applied Physics* 20.1, pp. 20–29. ISSN: 00218979. DOI: 10.1063/1.1698233.

Schmidt, Thomas GM and Arne Skerra (2007). "The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins". In: *Nature Protocols* 2.6, pp. 1528–1535. ISSN: 1754-2189. DOI: 10.1038/nprot.2007.209. URL: http://www.nature.com/doifinder/10.1038/nprot.2007.209.

Shannon, Claude E. (1949). "Communication in the Presence of Noise". In: *Proceedings of the IRE* 37.1, pp. 10–21. ISSN: 00968390. DOI: 10.1109/JRPROC.1949.232969.

Sheth, Lisa K., Angela L. Piotrowski, and Neil R. Voss (2015). "Visualization and quality assessment of the contrast transfer function estimation". In: *Journal of Structural Biology* 192.2, pp. 222–234. ISSN: 10958657. DOI: 10.1016/j.jsb.2015.06.012. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847715300162.

Smith, Donald B. and Kevin S. Johnson (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase". In: *Gene* 67.1, pp. 31–40. ISSN: 03781119. DOI: 10.1016/0378-1119(88)90005-4.

Smith, W C et al. (2004). "The surface of visual arrestin that binds to rhodopsin". In: *Mol Vis* 10.July, pp. 392–398. ISSN: 1090-0535. DOI: v10/a49[pii]. URL: http://www.ncbi.nlm.nih.gov/pubmed/15215746{\%}5Cnhttp://www.molvis.org/molvis/v10/a49/v10a49-smith.pdf.

Stofko-Hahn, Renata E., Daniel W. Carr, and John D. Scott (1992). "A single step purification for recombinant proteins Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase". In: *FEBS Letters* 302.3, pp. 274–278. ISSN: 00145793. DOI: 10.1016/0014-5793(92)80458-S. URL: http://www.ncbi.nlm.nih.gov/pubmed/1318232{\%}5Cn{\%}5C{\%}5Cji-cfs2{\%}5Cslproj{\%}5Cins{\%}5CSL{_}SHARED{\%}5Cpdfpapers{\%}5C5827.pdf.

Suzuki, Kei et al. (2016). "Structure of the Plexin Ectodomain Bound by Semaphorin-Mimicking Antibodies". In: *PLOS ONE* 11.6. Ed. by Juan Carlos Pizarro, e0156719. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0156719. URL: http://dx.plos.org/10.1371/journal.pone.0156719.

Tabata, Sanae et al. (2010). "A rapid screening method for cell lines producing singly-tagged recombinant proteins using the "TARGET tag" system." In: *Journal of proteomics* 73.9, pp. 1777–85. ISSN: 1876-7737. DOI: 10.1016/j.jprot.2010.05.012. URL: http://www.ncbi.nlm.nih.gov/pubmed/20566373.

Takagi, Junichi and Timothy a Springer (2002). "Integrin activation and structural rearrangement." In: *Immunological reviews* 186, pp. 141–163. ISSN: 0105-2896. DOI: 10.1034/j.1600-065X.2002.18613.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/12234369.

Takagi, Junichi et al. (2002). "Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling." In: *Cell* 110.5, pp. 599–11. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(02)00935-2. URL: http://www.ncbi.nlm.nih.gov/pubmed/12230977.

Takeda, Hiroyuki et al. (2017). "CP5 system, for simple and highly efficient protein purification with a C-terminal designed mini tag". In: *PLoS ONE* 12.5. Ed. by Ing-Feng Chang, e0178246. ISSN: 19326203. DOI: 10.1371/journal.pone.0178246. URL: http://dx.plos.org/10.1371/journal.pone.0178246.

Tang, Guang et al. (2007). "EMAN2: An extensible image processing suite for electron microscopy". In: *Journal of Structural Biology* 157.1, pp. 38–46. ISSN: 10478477. DOI: 10.1016/j.jsb.2006.05.009.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". In: *Applied Microbiology and Biotechnology* 60.5, pp. 523–533. ISSN: 0175-7598. DOI: 10.1007/s00253-002-1158-6. URL: http://link.springer.com/10.1007/s00253-002-1158-6.

Tomme, Peter et al. (1998). "Characterization and affinity applications of cellulose-binding domains". In: *Journal of Chromatography B: Biomedical Applications* 715.1, pp. 283–296. ISSN: 15726495. DOI: 10.1016/S0378-4347(98)00053-X.

Umitsu, Masataka et al. (2016). "Probing conformational and functional states of human hepatocyte growth factor by a panel of monoclonal antibodies". In: *Scientific Reports* 6.1, p. 33149. ISSN: 2045-2322. DOI: 10.1038/srep33149. URL: http://www.nature.com/articles/srep33149.

Ushioda, Ryo et al. (2016). "Redox-assisted regulation of Ca 2+ homeostasis in the endoplasmic reticulum by disulfide reductase ERdj5". In: *Proceedings of the National Academy of Sciences* 113.41, E6055–E6063. ISSN: 0027-8424. DOI: 10.1073/pnas.1605818113. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1605818113.

Van Regenmortel, Marc H. V. (1992). *Sructure of Antigens, Volume I*. Ed. by Marc H. V. Van Regenmortel. 1st ed. Boca Raton, Fla: CRC Press, p. 432. ISBN: 9780849388651.

Velázquez-Muriel, J. A. et al. (2003). "A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment". In: *Ultramicroscopy* 96.1, pp. 17–35. ISSN: 03043991. DOI: 10.1016/S0304-3991(02)00377-7. URL: http://www.ncbi.nlm.nih.gov/pubmed/12623169http://linkinghub.elsevier.com/retrieve/pii/S0304399102003777.

Vulović, Miloš et al. (2012). "Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy". In: *Ultramicroscopy* 116, pp. 115–134. ISSN: 03043991. DOI: 10.1016/j.ultramic.2012.03.004. URL: http://www.sciencedirect.com/science/article/pii/S030439911200040X.

Wade, R. H. (1992). "A brief look at imaging and contrast transfer". In: *Ultramicroscopy* 46.1-4, pp. 145–156. ISSN: 03043991. DOI: 10.1016/0304-3991(92)90011-8. URL: http://linkinghub.elsevier.com/retrieve/pii/0304399192900118.

Walker, Francesca et al. (2004). "CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor". In: *Journal of Biological Chemistry* 279.21, pp. 22387–22398. ISSN: 00219258. DOI: 10.1074/jbc.M401244200.

Wang, S and T Hazelrigg (1994). "Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis." In: *Nature* 369.6479, pp. 400–403. ISSN: 0028-0836. DOI: 10.1038/369400a0.

Ward, J (1963). "Hierarchical Grouping to Maximize an Objective Function". In: *Journal of the American Statistical Association* 58.301, pp. 236–244.

Williams, David B. and C. Barry Carter (2009). *Transmission electron microscopy: A textbook for materials science*. Boston, MA: Springer US, pp. 1–760. ISBN: 9780387765006. DOI: 10.1007/978-0-387-76501-3. arXiv: arXiv:1011.1669v3. URL: http://link.springer.com/10.1007/978-0-387-76501-3.

Wlodawer, Alexander, Mi Li, and Zbigniew Dauter (2017). "High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective". In: *Structure* 25, pp. 1–9. ISSN: 18784186. DOI: 10.1016/j.str.2017.07.012. URL: http://linkinghub.elsevier.com/retrieve/pii/S0969212617302460.

Wu, Shenping et al. (2012). "Fabs enable single particle cryoEM studies of small proteins". In: *Structure* 20.4, pp. 582–592. ISSN: 09692126. DOI: 10.1016/j.str.2012.02.017. arXiv: NIHMS150003.

Xiao, Tsan et al. (2004). "Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics." In: *Nature* 432.7013, pp. 59–67. ISSN: 1476-4687. DOI: 10.1038/nature02976. URL: http://www.ncbi.nlm.nih.gov/pubmed/15378069.

Xiong, J.-P. (2001). "Crystal Structure of the Extracellular Segment of Integrin alpha Vbeta 3". In: *Science* 294.5541, pp. 339–345. ISSN: 00368075. DOI: 10.1126/science.1064535. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1064535.

Yano, Tomoya et al. (2016). "AGIA tag system based on a high affinity rabbit monoclonal antibody against human dopamine receptor D1 for protein analysis". In: *PLoS ONE* 11.6, pp. 1–20. ISSN: 19326203. DOI: 10.1371/journal.pone.0156716.

Yip, Calvin K, Julia Berscheminski, and Thomas Walz (2010). "Molecular architecture of the TRAPPII complex and implications for vesicle tethering". In: *Nature Structural & Molecular Biology* 17.11, pp. 1298–1304. ISSN: 1545-9993. DOI: 10.1038/nsmb.1914. URL: http://www.nature.com/doifinder/10.1038/nsmb.1914.

Yu, Guimei, Kunpeng Li, and Wen Jiang (2016). "Antibody-based affinity cryo-EM grid". In: *Methods* 100, pp. 16–24. ISSN: 10959130. DOI: 10.1016/j.ymeth.2016.01.010. URL: http://linkinghub.elsevier.com/retrieve/pii/S104620231630010X.

Zhang, Kai (2016). "Gctf: Real-time CTF determination and correction". In: *Journal of Structural Biology* 193.1, pp. 1–12. ISSN: 10958657. DOI: 10.1016/j.jsb.2015.11.003.

Zheng, Shawn Q et al. (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy". In: *Nature Methods* 14.4, pp. 331–332. ISSN: 1548-7091e. DOI: 10.1038/nmeth.4193. URL: http://www.nature.com/doifinder/10.1038/nmeth.4193.

Zhu, Jianghai et al. (2008). "Structure of a Complete Integrin Ectodomain in a Physiologic Resting State and Activation and Deactivation by Applied Forces". In: *Molecular Cell* 32.6, pp. 849–861. ISSN: 10972765. DOI: 10.1016/j.molcel.2008.11.018. URL: http://dx.doi.org/10.1016/j.molcel.2008.11.018.

Zhu, Jun et al. (1997). "Three-Dimensional Reconstruction with Contrast Transfer Function Correction from Energy-Filtered Cryoelectron Micrographs: Procedure and Application to the 70SEscherichia coliRibosome". In: *Journal of Structural Biology* 118.3, pp. 197–219. ISSN: 10478477. DOI: 10.1006/jsbi.1997.3845.

Zingsheim, H. P. et al. (1982). "Direct structural localization of two toxin-recognition sites on an ACh receptor protein". In: *Nature* 299.5878, pp. 81–84. ISSN: 0028-0836. DOI: 10.1038/299081a0. arXiv: arXiv:1011.1669v3. URL: http://www.nature.com/doifinder/10.1038/299081a0.

Appendix A

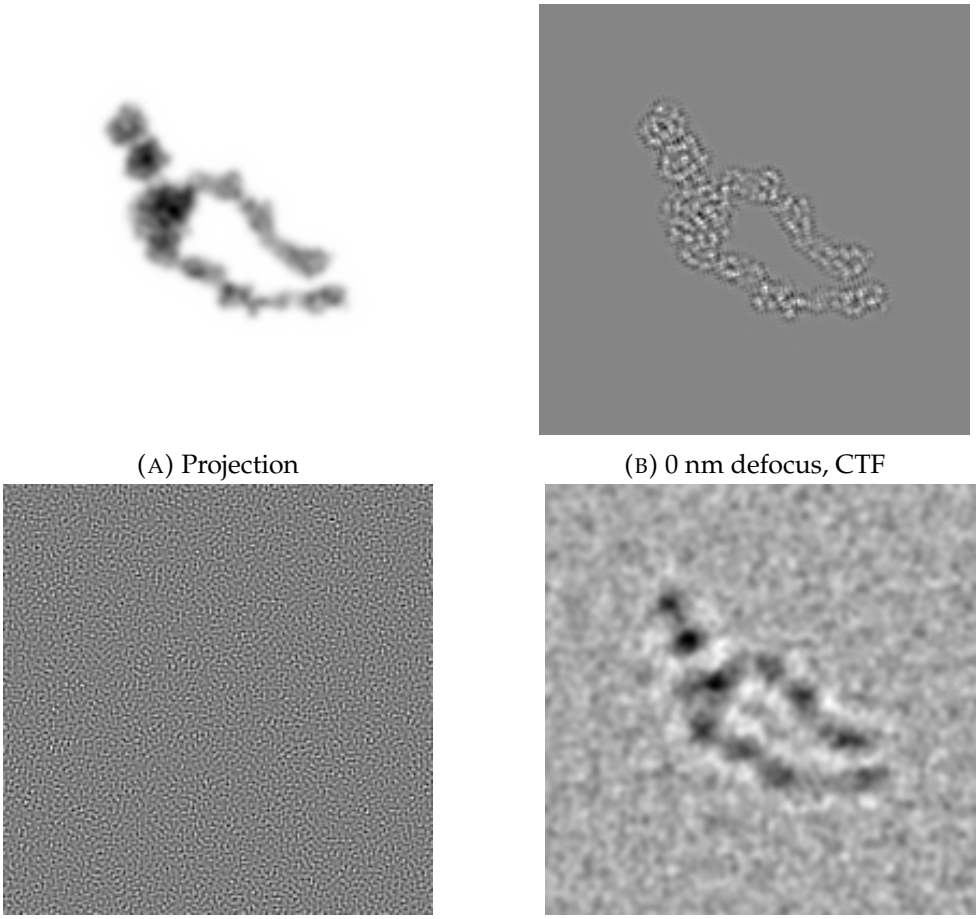
The Contrast Transfer Function

A.1 Introduction

When an object is imaged in the electron microscope it is distorted by the spherical aberration of the microscope so that it does not display a faithful projection but gives a modified projection that is modulated by the microscopes spherical aberration and defocus (Erickson and Klug, 1971; Wade, 1992). The Fourier transform of an object is another representation of this object, and given the many image analysis techniques that can be applied in Fourier space it is a useful representation. Physically, the Fourier transform of an object is present in the back focal plane of the microscope where the higher spatial frequencies relate to the higher scattering angles (i.e., a high scattering angle contains high frequency information). The CTF modulates the contrast that is present in these frequencies so that contrast is not uniform, but will flip to white and back again as the spatial frequency increases.

A.2 The perfect CTF

Figure A.1a shows an example of an atomic projection that is faithful to the atomic organisation (**A.1a**), and how in focus images that are modified by the EM still contain information (**A.1b**), but this is lost when more realistic imaging conditions are used (**A.1c**). Defocusing the microscope causes a phase shift that generates contrast and allows the object to be recognised (**A.1a**). The CTF is a description of the changes to contrast that occur at the various spatial frequencies in the back focal plane of the microscope. Although many different formulations of the CTF equations are in use (see **section A.2.1**), I prefer Kai Zhang's equations from his paper (Zhang, 2016) as they have an explicit inclusion for astigmatism of the image. They are reproduced below:



(A) Projection

(B) 0 nm defocus, CTF

(C) 0 nm defocus, CTF, noise

(D) 2500 nm defocus, CTF, noise

FIGURE A.1: **Projections from an atomic model and applied CTF under perfect and simulated imaging conditions**. All projections are of the deformed atomic model of $\alpha_{IIb}\beta_3$ integrin with NZ-1 Fab bound to the W2 loop (residues 136-137) modified using the IMAGIC (Heel et al., 1996) CTF2D-OPERATIONS command. A.1a shows a perfect atomic projection with inverted contrast (as would be expected from imaging with EM. A.1b shows the same projection but with a CTF applied using 0 nm defocus (i.e., in focus). The focus projection becomes indistinguishable from the background when noise is added and the same CTF operation is performed, A.1c. Increasing the defocus improves the contrast in A.1d and some features are observed. The parameters used were sample width: 500 Å^2, spherical aberration (C_s): 2.0 mm, focal length: 2.8 mm, objective aperture: 50 μ m, acceleration voltage: 200 kV, chromatic aberration: 0.5 eV, source size: 1.0 μ m, generic envelope factor: 0.4, and for A.1c and A.1d had random noise applied prior to applying the CTF using the IMAGE-MENU command, μ=0, σ=10.

TABLE A.1: Parameters used in the CTF equation

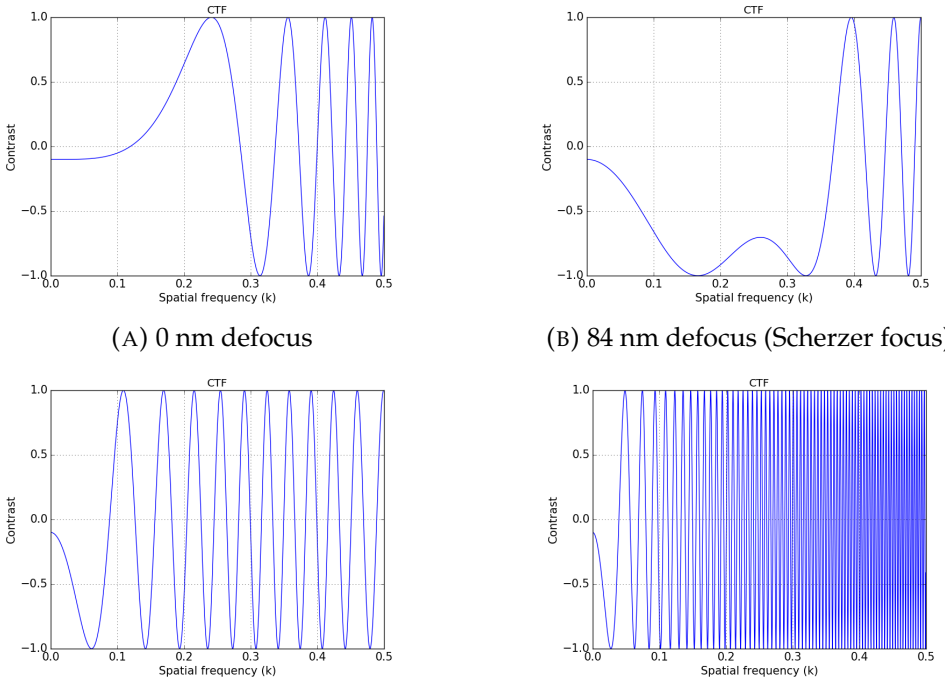
Parameter	equivalent term	unit	Explanation
k	s, α		spatial frequency, or scattering angle
Δz	Δf, z(θ)	μm, nm, Å	defocus, underfocus is positive
A	W, $f\alpha$		Amplitude contrast
C_s		mm	Spherical aberration
λ		Å	Wavelength of electron

$$\begin{aligned} CTF(\vec{s}) &= -\sqrt{1-A^2}\cdot\sin(\gamma(\vec{s})) - A\cdot\cos(\gamma(\vec{s})) \\ &= -\sin(\Delta\phi + \gamma(\vec{s})) \end{aligned} \tag{A.1}$$

$$\gamma(\vec{s}) = \gamma(s,\theta) = -\frac{\pi}{2}C_s\lambda^3 s^4 + \pi\lambda z(\theta)s^2 \tag{A.2}$$

$$z(\theta) = z_u\cos^2(\theta - \theta_{ast} + z_v\sin^2(\theta - \theta_{ast}) \tag{A.3}$$

The various terms are outlined in **table A.1**. From the formulation in **equation A.1** it is clear that the maximum and minimum values are between -1 and +1, which relates to total black or total white contrast. Although the majority of contrast is generated by defocusing the microscope, there is still a contribution from scattering, or amplitude, contrast that is modelled by the cos term (that can be expressed as a phase shift, $\Delta\phi$). I go into some more detail regarding this in **Appendix ??**). From the equation underfocus is a positive value as it creates the correct contrast (i.e., negative) at low spatial frequencies which is what we observe. The behaviour of the CTF can be quite simple to model as I have done in **Figure A.2** that shows the relative contrast at each spatial of the various spatial frequencies. Initially (**A.2a**) there is very little contrast at low spatial frequencies apart from a contribution by amplitude contrast, and so although the image contains high frequencies information there is not enough contrast to correctly align the image. Increasing the defocus increases the contrast at the low frequencies and allows the object to become visible. Since the contrast oscillates between black and white unless this is computationally correct then analysis is limited to the resolution where the contrast flip occurs first. This has a special term, the *Scherzer focus* (Scherzer, 1949), the defocus that has the largest area under the first peak *before it crosses the zero axis and transfers no contrast*. This is important as the data contained up to that first cross over is correct. Although this gives use the best estimate of the real structure without applying computational corrections to the contrast it has severe limitations for biological specimens as their scattering contrast is so low there is still essentially not useful information for alignment at such a close defocus. Rather, biological specimens are images at a much



(A) 0 nm defocus

(B) 84 nm defocus (Scherzer focus)

(C) 500 nm defocus

(D) 2500 nm defocus

FIGURE A.2: **Plots of a CTF with changing defocus**. This figure shows the changing CTF as defocus is increased. Defocus ranges are A.2a: 0 nm, A.2b: 84 nm defocus (Scherzer focus), A.2c: 500 nm, A.2d: 2500 nm. A.2b is important as the so called Scherzer focus, the point where the largest area of the first peak before it crosses the zero contrast. As defocus is increased (**A.2b-A.2d**) the CTF rapidly oscillates between black and white which need to computationally corrected. Only the defocus was changed in **equations A.1** and **A.2**, all other terms were kept constant with spherical abberation: 2.0 mm, acceleration voltage: 200 kV and no astigmatism. These curves were generated with a custom script written in Python 2.7.12.

higher defocus and then the CTF parameters are estimated and the phases are corrected (i.e., all flipped to white contrast) as outlined in **section A.5**. As can be seen by comparing **A.2c** and **A.2d** an increase in deocus causes the rapid oscillation of the CTF across the zero axis which means that the spatial frequencies where no f is transferred increases. Generally for biological samples the defocus is selected so that the first zero is at a resolution that contains the largest dimension of the target protein.

A.2.1 Other formulations of the CTF

The literature contains several different formulations of the CTF that are essentially mathematically equivalent as long as one is careful about the units and various conventions. Zhu et al., (1997) have an equivalent wave-function equation (A.5) but

their formulation of the CTF includes an integer of size 2 that increases the oscillation. Functionally it is equivalent but with amplitudes reaching 2 rather than 1 as with **equation A.1**.

$$H(k) = 2[\sin\gamma(k) - W\cos\gamma(k)] \tag{A.4}$$

$$\gamma(k) = 2\pi(-0.5\Delta z\lambda k^2 + 0.25C_s\lambda^3 k^4) \tag{A.5}$$

Another formulation that includes some more terms and makes it explicit that the spatial frequency relates to scattering angle in the back focal plane is from Erickson and Klug, (1971). Although all of these equations are functionally identical.

$$T_{ph}(\alpha,\phi) = -T^0(\alpha,\phi)A(\alpha)f(\alpha)\sin\chi(\alpha) \tag{A.6}$$

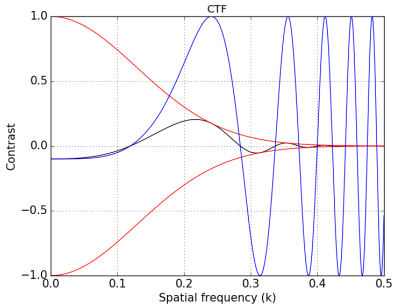
$$\chi(\alpha) = \frac{2\pi}{\lambda}\left[-\frac{1}{4}C_s + \alpha^4 + \frac{1}{2}\Delta f\alpha^2\right] \tag{A.7}$$

A.3 Envelope Functions

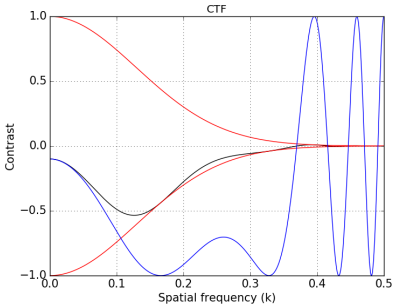
As outlined in **Appendix B** a real EM suffers from a number of issues that limit the amplitude of the various spatial frequencies that can be realistically obtained. This severely attenuates the signal that can be meaningfully collected from an EM image. **Figure A.3** shows the same simulated CTF from **figure A.2** with a generic envelope function (**equation B.5**) with a very optimistic of 15 Å^2, which is much lower than real experimentally obtained values. It should be clear why in focus and Scherzer focus do not provide any meaningful contrast, since almost all of the amplitudes are outside the envelope function (blue curves in **A.3**) and the remaining contrast producing curves are very small (black lines in **A.3**). The higher the B-factor the quicker the signal is attenuated.

A.4 Acceleration voltage and the CTF

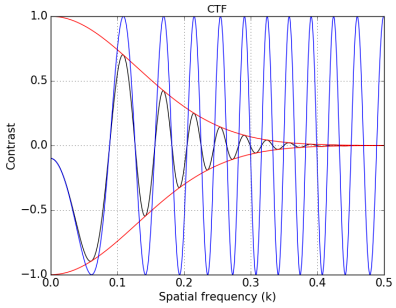
Equation A.2 depends on the wavelength of the electron that is used and so the acceleration of the microscope is important. **Figure A.4** shows the change in the CTF as the acceleration voltage of the microscope is increased. Importantly the higher the voltage of the EM that is used the fewer times the phases will flip for a given



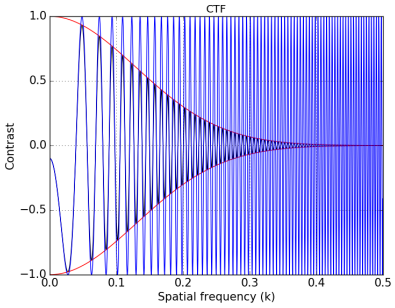
(A) 0 nm defocus



(B) 84 defocus (Scherzer focus)

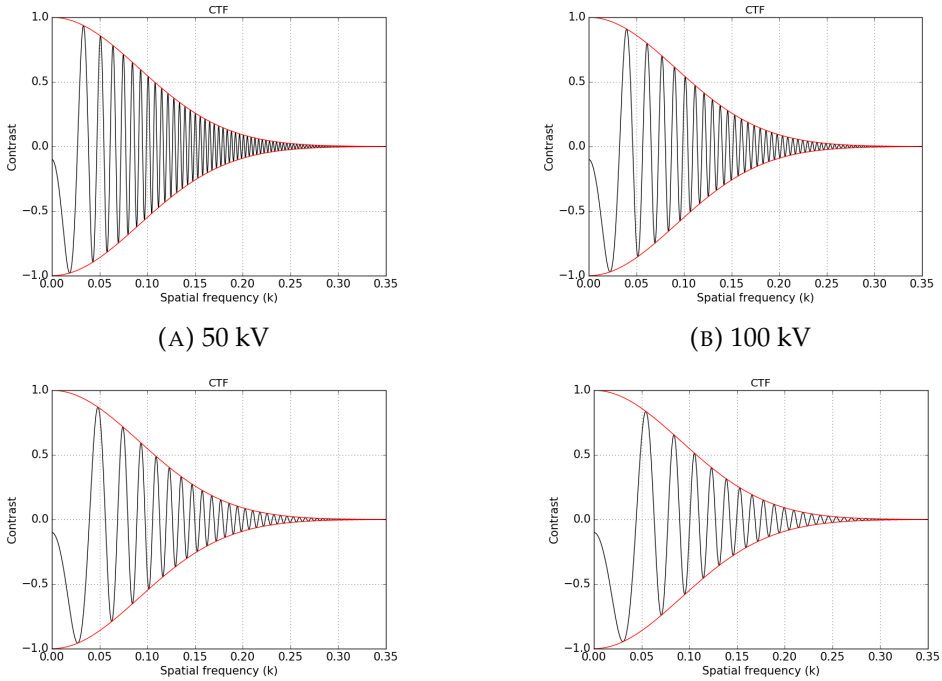


(C) 500 nm defocus



(D) 2500 nm defocus

FIGURE A.3: **Plots of a perfect CTF against an envelope function limited CTF as defocus is increased**. This figure compares a perfect CTf (blue curves) to a realistic CTF (black curves) limited by a generic envelope function (red curves) with B-factor of 15 Å^2. At the in focus condition, **A.3a**, there is very little amplitude contained within the envelope function and so contrast is very low. As defocus is increased (**A.3b-A.3d**) the CTF rapidly oscillates between black and white which need to computationally corrected. Only the defocus was changed in **equations A.1** and **A.2**, all other terms were kept constant with spherical aberration: 2.0 mm, acceleration voltage: 200 kV and no astigmatism. These curves were generated with a custom script written in Python 2.7.12.



(A) 50 kV (B) 100 kV

(C) 200 kV (D) 300 kV

FIGURE A.4: **Plots of a CTF with changing acceleration voltage**. Changed to the acceleration voltage, A.4a 50 kV, A.4b 100 kV, A.4c 200 kV, A.4d 300 kV. Other parameters are kept constant. Spherical abberation: 2.0 mm, B-factor: 30 $Å^2$, defocus: 2500 nm.

defocus values which means that it is easier to fit a CTF curve to the data, although the decreased contrast that results is an issue that can limit analysis.

A.5 CTF Estimation

As can be seen from **figure A.1** the collected image loses some resemblance to the true particle due to the CTF of the microscope. To determine the true structure we need to estimate the defocal values the wave equation (**equation A.2**) so that we can solve the CTF and correct all the phases (i.e., flip them all to positive or negative density). To this end there are a number of CTF estimation programs (Mallick et al., 2005; Sheth et al., 2015; Vulović et al., 2012; Velázquez-Muriel et al., 2003) but the two that I have familiarity are gCTF (Zhang, 2016) and CTFfind4 (Rohou and Grigorieff, 2015). They output a lot of useful files that contain a lot of interesting information that can be used to assess the quality of the of the micrographs that I used to get an overview of my micrographs (see Results 5.2.2).

After CTF estimation there are a few questions that we need to think about when looking at our data:

1. What is the defocal range ouf our data?

2. How astigmatic is the data? i.e, is there a large variation between $Defocus_u$ and $Defocus_v$ values.

3. How close are we to Nyquist? Is it obvious that we are collecting bad data in some way?

A.6 Summary

The CTF can seem a very complicated term, but in describes the change in contrast caused by the *imperfectability* of our electron microscopes. To recapture the correct image (or the true image) we need to do *CTF correction*, to reverse some of the problems that the EM introduces.

With some simple code is possible to explore who changing the various parameters change the behaviour of the CTF and gain a better understanding of this equation that is central to the operation of the EM and analysis of protein structure.

Appendix B

Envelope Functions

Envelope functions truncate the signal from the microscope.

B.1 The different envelope functions

There are many different envelope functions that have been determined in the literature. For equations B.1, B.2, B.3, and B.4 they are from Zhu et al., (1997), and the other equations are from Frank, (2006).

B.1.1 Source size

$$E_{pk}(k) = exp[-\pi^2 q^2 (k^3 C_s \lambda^3 - \Delta z k \lambda)^2] \tag{B.1}$$

q is the effective source size, see **Appendix C** for a discussion on what the different source sizes of some of the guns are.

B.1.2 Energy spread

$$E_{es}(k) = exp[-\frac{1}{16ln2}\pi^2 \delta z^2 k^4 \lambda^2] \tag{B.2}$$

δz is the defocus variation caused by the energy spread.

B.1.3 MTF envelope function

$$E_f(k) = 1/[1 + (\frac{k}{k_f})^2] \tag{B.3}$$

k_f is a term that is determined for each material.

B.1.4 Drift, speciment charging and scattering

$$E_g(k) = exp[-(\frac{k}{k_g})^2] \tag{B.4}$$

k_g is the half width in Fourier Space, but what does that mean, I am not sure. This is mentioned on p.93 of (Frank, 2006) and maybe discussed in a bit more detail in **Appendix D**.

B.1.5 Generalised envelope function

Although precise envelope functions can be derived for each of the various sources of envelope (e.g., source size) a generalised envelope function can also be determined.

$$E = exp[-Bs^2] \tag{B.5}$$

Some of the problems associated with determined the B-factor in this way are outlined in Wlodawer et al., (2017).

B.2 Simulated envelope function

I made a simple simulation of an envelope function to show how quickly the signal can be attenuated with only 'small' B-factors.

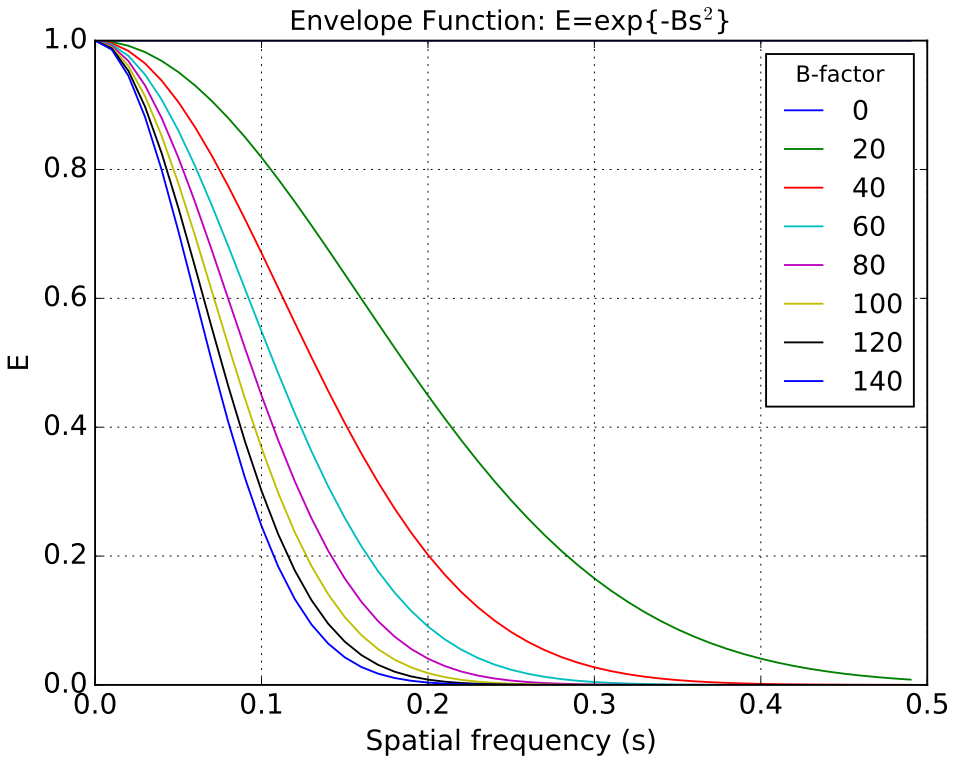


FIGURE B.1: This is a generic envelope function showing how it changes with increasing B value (i.e., increased atomic movement). Written in Python.

Appendix C

EM guns

Electron microscopy depends on a reliable source of electrons to illuminate some sample (citation needed). To this end a number of electron sources, commonly referred to as electron guns, have been developed. Electron emission from a cathode can be modelled by the following Schottkey equation (Orloff, 2008), with emission J_s, electric field strength F, temperature T, and the materials work function ϕ. Other constants are defined in the abbreviations.

$$J_s = \frac{4\pi m_e(kT^2)}{h^3} exp(\frac{e^{3/2}F^{1/2}}{(4\pi\epsilon_0)^{1/2}kT} - \frac{\phi}{kT}) \tag{C.1}$$

C.1 Thermionic guns

When the electric field is very low (i.e., F $\approx$ 0) then the Schottkey equation can be simplified to Richardson's Law (Williams and Carter, 2009).

$$J = AT^2 exp(-\frac{\phi}{kT}) \tag{C.2}$$

In this case the obvious way to increase the current is to increase the temperature, which is done with tungsten filament (which operate at a very high temperature) or to decrease the work function of the material, as with LaB_6 guns.

C.2 Field Emission guns

As can be seen from the full Schottkey equation increasing the electric field strength can also increase the electron beam current. This is utilised in field emission guns. There are two major types. The cold field gun, and the Schottkey gun. Both create a fine tip that can experience a very high electric field and allow electrons to tunnel

past the work function. Field emission guns have reduced *chomatic abberation* and higher *spatial coherency* and so are much better sources of electrons that thermionic guns[1]

[1]With everything in science it always *depends*. If the microscope will only be used for screening or at very low resolutions then perhaps it is better to by a cheaper gun and spend that money elsewhere.

Appendix D

Aligning Images

Using the notation that was set out by Frank, (2006). An image is made up of L rows and M columns of pixels, with total pixels J=L $\times$ M. This lets us use an index for the *j*th pixel using the following equation:

$$j = (m-1) \times L + l \tag{D.1}$$

If we have N images the each image (p_i, i=1 ... N), then we can reference the equivalent pixel from an image with the p_{ij} notation. For these pixels to be equivalent then the images need to be aligned so that they reference the same location (of a protein for example) and then an analysis will have some meaning. If we can convert each image, p_{ij}, with some transformation $\mathbf{T}_i$ such that all p_j are in the equivalent location then we can take a *meaningful* average of each pixel, with $p'_{ij} = \mathbf{T}_i p_{ij}$. If there is a *true* alignment (by some measure) then we take a sum of all those pixels and generate an average with the following equation:

$$\bar{p}_j = \frac{1}{N} \sum_{i=1}^{N} p'_{ij} \tag{D.2}$$

The *deviation* from a true and perfect alignment can be modelled in Fourier space, as a Guassian attenuation of the signal. Which is clearly the same as equation B.4 as discussed in Appendix B.1.4:

$$\mathfrak{F}(\hat{p}) = \mathfrak{F}(\bar{p}) exp[-(\frac{k}{k_g})^2] \tag{D.3}$$

For these averages to be analysed we need to align the collection of noisy 2D images which can be done by cross-correlation (**section D.1**) or Bayesian methods (Scheres, 2012).

D.1 Cross-correlation

The following is the formulation of the cross-correlation function (CCF) that was set out by Frank, (2006).

We can represent images as vectors in multi-dimensional space, and so a simple Euclidean distance between the end points of these two vectors can be used as a judge of similarity. If we aim to maximise similarity then we can perform a number of rotational, and translational operations and compare each of these new coordinates to find the most similar position.

$$f_1(\vec{r_j})j = 1, 2, \ldots, Jf_2(\vec{r_j})j = 1, 2, \ldots, J \tag{D.4}$$

Introduction of a rotation and translation vector $\vec{R_\alpha}, \vec{r\prime}$ produces the following expression:

$$E_{12}^2(\vec{R_\alpha}, \vec{r\prime}) = \sum_{j=1}^{J}[f_1(\vec{r_j}) - f_2(\vec{R_\alpha}\vec{r_j} + \vec{r\prime})]^2 \tag{D.5}$$

When this is expanded out there are three terms, the first two which are invariant and the third term which is the cross-correlation function. When the cross-correlation function is maximised the images are optimally aligned.

$$\Phi_{12}(\vec{R_\alpha}, \vec{r\prime}) = \sum_{j=1}^{J} f_1(\vec{r_j}) f_2(\vec{R_\alpha}\vec{r_j} + \vec{r\prime}) \tag{D.6}$$

Appendix E

Publications and presentations

E.1 Publications

1. Brown, Z. P., Arimori, T., Iwasaki, K. & Takagi, J. Development of a new protein labeling system to map subunits and domains of macromolecular complexes for electron microscopy. J. Struct. Biol. (2017). doi:10.1016/j.jsb.2017.11.006

2. Brown, Z. P., & Takagi, J. "The PA tag: a versatile peptide tagging system in the era of integrative structural biology" in Nakamura, H. Springer-Nature: *Integrative structural biology with hybrid methods*

3. Fislage, M., Zhang, J., Brown, Z. P., Mandava, C. S., Sanyal, S. & Frank, J. Cryo-EM of GTPase deficient 70S EF-TuH84A GTP complex reveals new initial codon selection state of aa-tRNA. Nucleic Acids Research (Submitted)

4. Shibai A., Arimoto T., Yoshinaga T., Tsuchizawa Y., Khureltulga D., Brown Z. P., Kakizuka T., & Hosoda K. Recognition of posture and locomotion elements of conspecific biological motion in medaka. Scientific Reports (Submitted)

E.2 Conference presentations

1. Brown Z. P., Arimori T., Iwasaki K., & Takagi J. Development of a new protein labeling system to map subunits & domains of macromolecular complexes for electron microscopy (EM). 第40回日本分子生物学会年会

2. Brown Z. P., Kitago Y., Matoba K., Iwasaki K., Hattori M., & Takagi J. Recombinant production of intact full-length Reelin protein allowed visualization of its multimer by negative-stain electron microscopy. 第15回日本蛋白質科学会年会

E.3 Poster presentations

1. Brown Z. P., Arimori T., Iwasaki K., & Takagi J. Development of a new protein labeling system to map subunits & domains of macromolecular complexes for electron microscopy (EM). クライオ電子顕微鏡によるタンパク質の高分解能単粒子構造解析

2. Brown Z. P., Arimori T., Iwasaki K., & Takagi J. Development of a new protein labeling system to map subunits & domains of macromolecular complexes for electron microscopy (EM). 第40回日本分子生物学会年会

3. Brown Z. P., Kitago Y., Matoba K., Iwasaki K., Hattori M., & Takagi J. Recombinant production of intact full-length Reelin protein allowed visualization of its multimer by negative-stain electron microscopy. 第15回日本蛋白質科学会年会