Structural investigations of CLIC proteins and importin-α recognition of nuclear localisation signals

Author:
Mynott, Andrew Vincent

Publication Date:
2009

DOI:
https://doi.org/10.26190/unsworks/22982

Structural Investigations of CLIC Proteins and Importin-α Recognition of Nuclear Localisation Signals

Andrew Vincent Mynott

A thesis submitted in fulfilment of the requirement for
the degree of Doctor of Philosophy

University of New South Wales
Sydney, Australia

September, 2009

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed …….....

Date …….....

Papers resulting from this thesis

Part of the work presented in Chapter 2 of this thesis has been previously published.

Littler, D.R., Harrop, S.J., Brown, L.J., Pankhurst, G.J., Mynott, A.V., Luciani, P., Mandyam, R.A., Mazzanti, M., Tanda, S., Berryman, M.A., Breit, S.N., Curmi, P.M. *Comparison of vertebrate and invertebrate CLIC proteins: the crystal structures of Caenorhabditis elegans EXC-4 and Drosophila melanogaster DmCLIC.*
Proteins, 2008. **71**(1): p. 364-378.

ABSTRACT

The chloride intracellular channel (CLIC) family of proteins are an unusual class of chloride channels that possess the ability to auto-insert into cellular membranes. The CLICs exhibit ubiquitous tissue and cellular distributions and adopt a glutathione S-transferase fold in the soluble form that is highly conserved in vertebrates. CLIC homologues have been identified in the model organisms *C. elegans* and *D. melanogaster*, and in the former case have been extensively characterised in regards to function. In this thesis, we present the crystal structure of the *D. melanogaster* CLIC, revealing several unique features in the conserved invertebrate CLIC fold including an elongated C-terminal extension and metal binding site. The bound metal is identified as the potassium cation, resolving concerns regarding previously published work that assign the metal as the isoelectronic calcium cation.

It has been reported that a human CLIC protein, CLIC4, translocates to the nucleus in response to cellular stress, facilitated by a putative CLIC4 nuclear localisation signal (NLS). The CLIC4 NLS adopts α-helical structure in the native CLIC4 structural fold. It is proposed that CLIC4 is transported to the nucleus via the classical nuclear import pathway after binding the import receptor, importin-α. We have determined the X-ray crystal structure of a truncated form of importin-α bound to a CLIC4 NLS bearing peptide. The NLS peptide binds the major binding site in an extended conformation similar to that observed for the classical SV40 large T-antigen NLS. A tyrosine residue within the CLIC4 NLS makes surprisingly favourable interactions by forming side chain hydrogen bonds to the importin-α backbone. This structural evidence supports the hypothesis that CLIC4 translocation to the nucleus is governed by the importin-α nuclear import pathway, providing it can undergo a conformational rearrangement that exposes the NLS in an extended conformation.

We further analyse importin-α:NLS binding interactions by solving high resolution structures of truncated importin-α containing an empty binding site and bound to the SV40 NLS. A surprising interaction is discovered between importin-α and an NLS-like motif in the endogenous *E. coli* 30S ribosomal subunit S21, revealing new insight into importin-α recognition of full length cargo.

ACKNOWLEDGMENT

I would like to acknowledge all those who helped in the creation of this thesis, whether through direct assistance or indirect support. Firstly, thank you to my supervisors Paul Curmi and Louise Brown for their insightful scientific guidance, the time they have devoted to my work and for providing me the opportunity to do this research project. I would also thank our collaborator Bostjan Kobe for providing us with the original truncated importin-α clone and Sam Breit and members of his group at the Centre for Immunology at St. Vincent's hospital.

I extend my gratitude to lab members past and present: Stephen Harrop for his expert teachings on all things crystallography; Krystyna Wilk for many formal and informal discussions on science and life; Juanita Phang as a fellow PhD student; Mika Jormakka for his help in the lab and on the golf course; Anthony Duff for his support with experiments outside UNSW; Kim Vincent and Roberta Donadini for their support in the lab; Diana Edler for her support in the teaching and experimental labs; Gary Keenan for his efforts to keep rusty old equipment humming in the lab; David Jonas for his help with computer breakdowns and enjoyable games of soccer played on the village green and Dene Littler for invaluable help around the lab when I was just beginning this journey.

I would lastly like to thank my family and friends for their patience and understanding over the last few years. To Mimi, David, Cherrie, and Mia, thank you for supporting me in my endeavours. Most of all I would like to thank my parents, Maureen and John, for their love, support and patience over these last few years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In this thesis we perform structural investigations that focus on two separate homologous chloride intracellular channel (CLIC) family members from *Drosophila melanogaster* (*Dm*CLIC) and *Homo sapiens* (CLIC4). Both studies use X-ray crystallography techniques to examine the structural characteristics of these proteins. In the case of *Dm*CLIC, we analyse the overall structure including an interaction with a metal ligand that is identified as the potassium cation. In the case of CLIC4, we specifically analyse an interaction with the nuclear import receptor, importin-α, which is hypothesised to traffic CLIC4 to the nucleus in response to cell stress. In addition to this, we investigate importin-α binding interactions with nuclear localisation signals (NLSs) from the SV40 large T-antigen and *E. coli* 30S ribosomal subunit S21.

In this introductory chapter we present a summary of the current literature on the CLIC family of proteins as well as the nuclear import pathway. We begin by reviewing current understandings of the nucleocytoplasmic transport cycle and particularly the classical nuclear import cycle, which involves the importin-α recognition of NLSs (Section 1.1). We then review the literature on CLIC proteins, with a focus on human CLIC1 and CLIC4, as well as the invertebrate CLIC homologues (Section 1.2). Finally, we review a key literature article that has investigated a CLIC4 associated nuclear translocation event and an interaction between CLIC4 and importin-α (Section 1.3). A summary and description of research aims are then presented in Section 1.4.

1.1 Nuclear import and the karyopherins

All eukaryotic cells contain a nucleus which separates the cells genetic material from the cytoplasm through a double membrane known as the nuclear envelope. Trafficking of molecules across this membrane is critical for the overall function of the cell including gene expression and apoptosis [1]. The movement of molecules between the cytoplasm and nucleus is facilitated by large protein complexes known as nuclear pore complexes (NPC). The NPC perforates the nuclear membrane creating a channel that allows the transport of soluble molecules via either passive diffusion or a process mediated by transport factors. NPCs are enormous protein structures (~50 MDa in yeast, [2]) that are made up of multiple copies of nucleoporins, a large proportion of which contain characteristic phenylalanine-glycine repeats (FG repeats) [3]. These FG-nucleoporins contain tandem FG motifs that can be broken into two main groups: GLFG sequences and FxFG sequences, which are normally interspersed with charged or polar amino acid spacer sequences [3]. Nucleoporins are arranged in the NPC so that long stretches of FG repeats line the inner side of the pore creating a fine mesh that could provide both an energetic and physical permeability barrier [4]. While smaller molecules can passively diffuse through the NPC, larger transporters overcome the permeability barrier by recognising and transiently binding to FG-nucleoporins which act like a ladder for the transport proteins to move across the pore. Analysis of the yeast NPC shows there are approximately 30 individual nucleoporins comprising the complex [2, 5]. Mass spectrometry on isolated NPCs derived from rat liver show that the composition of the mammalian NPC is similar to that of yeast, consisting of 29 unique nucleoporins and a total mass of ~125 MDa [6, 7].

The nucleocytoplasmic transport pathway can be divided into an import cycle and export cycle. The import cycle is facilitated by nuclear import receptors that identify cytoplasmic cargo proteins to be transferred to the nucleus by recognising a specific amino acid sequence within the cargo known as a nuclear localisation signal (NLS). Conversely, nuclear export receptors discriminate between nuclear cargo to be transported to the cytoplasm and other nuclear proteins by recognising a nuclear export signal (NES). While the import and export signals form the basis for the regulation of

nucleocytoplasmic transport, there are a host of broader regulating processes including posttranscriptional and posttranslational modifications, as well as protein expression and stability [8].

While there are a number of transport pathways utilised by the cell to transverse the nuclear envelope, such as the nuclear transport factor - 2 (NTF2) importer [9] and mRNA exporter pathways [10], the largest and most well studied system involves the importin-β superfamily. Proteins in the importin-β family are also known as β-karyopherins and include both soluble import and export receptors (suitably named importins and exportins). Although a number of importin-β-like proteins can interact directly with cargo proteins, the adaptor molecule importin-α is the best characterised receptor and is more commonly involved in cargo recognition.

1.1.1 The classical nuclear import pathway

Of the different pathways the cell employs to traffic proteins across the NPC, the best understood and most extensively studied is what is now known as the classical nuclear import pathway. This pathway involves the recognition of NLS sequences by the import receptor, importin-α, described in Section 1.1.4. Other key proteins are NTF2 and Ran (Ras-related nuclear protein) (Section 1.1.2), as well as importin-β (Section 1.1.3). A description of significant processes in the import and export cycle now follows, with each key event represented in the schematic shown in Figure 1.1.

1. Free cytoplasmic importin-α is in the auto-inhibited state with a low affinity for NLSs. The importin-α IBB domain (Section 1.1.4.2) binds importin-β and switches importin-α into a high affinity form [11].
2. In the importin-α:β heterodimer state, importin-α has a free NLS recognition site which can bind NLS bearing cargo.
3. The importin-α:β:NLS trimeric complex crosses the NPC mediated primarily by interactions between importin-β and FG-nucleoporins [12].
4. Once in the nucleus, dissociation of the importin-α:β:NLS complex begins with the binding of Ran-guanosine triphosphate (RanGTP) to importin-β. This interaction effectively disables importin-β interaction with importin-α by

inducing a conformational change in importin-β [13]. The RanGTP binding site in importin-β partly overlaps with the binding site of importin-α and the resulting steric interference may also contribute to complex dissociation.

5. The importin-β:RanGTP complex is recycled to the cytoplasm where RanGTP is hydrolysed to Ran-guanosine diphosphate (RanGDP). The hydrolysis reaction is mediated by RanGAP (RanGTPase-activating protein) and promoted by RanBP1 (Ran-binding protein 1) [14].
6. After importin-β is removed from the import complex, importin-α is still bound to its NLS cargo within the nucleus. There are a number of factors known to contribute to importin-α:NLS dissociation. The removal of importin-β allows the IBB domain from importin-α to again compete for the binding site. This auto-inhibitory mechanism is likely to have the largest impact on complex dissociation. There could also be accelerated dissociation via an interaction with the nucleoporin nup2p (nup50 in vertebrates) which binds to importin-α in a configuration that overlaps the NLS minor binding site [15].
7. The importin-α:nup2p complex is dissociated by the export receptor CAS (or Cse1p in yeast) bound to RanGTP [16], which forms the export complex CAS:RanGTP:importin-α that recycles importin-α back to the cytoplasm. The interaction between CAS:RanGTP and importin-α may also have a role in importin-α:NLS dissociation. The CAS:RanGTP:importin-α complex is formed with importin-α in the auto-inhibited low affinity state, thus ensuring only cargo free importin-α is recycled to the cytoplasm [17].
8. After navigating through the NPC, importin-α is released from the export complex in the cytoplasm where it is free to participate in another import cycle. Dissociation of the CAS:RanGTP:importin-α complex occurs via the GTPase-activating protein RanGAP, which hydrolyses RanGTP to RanGDP [18].

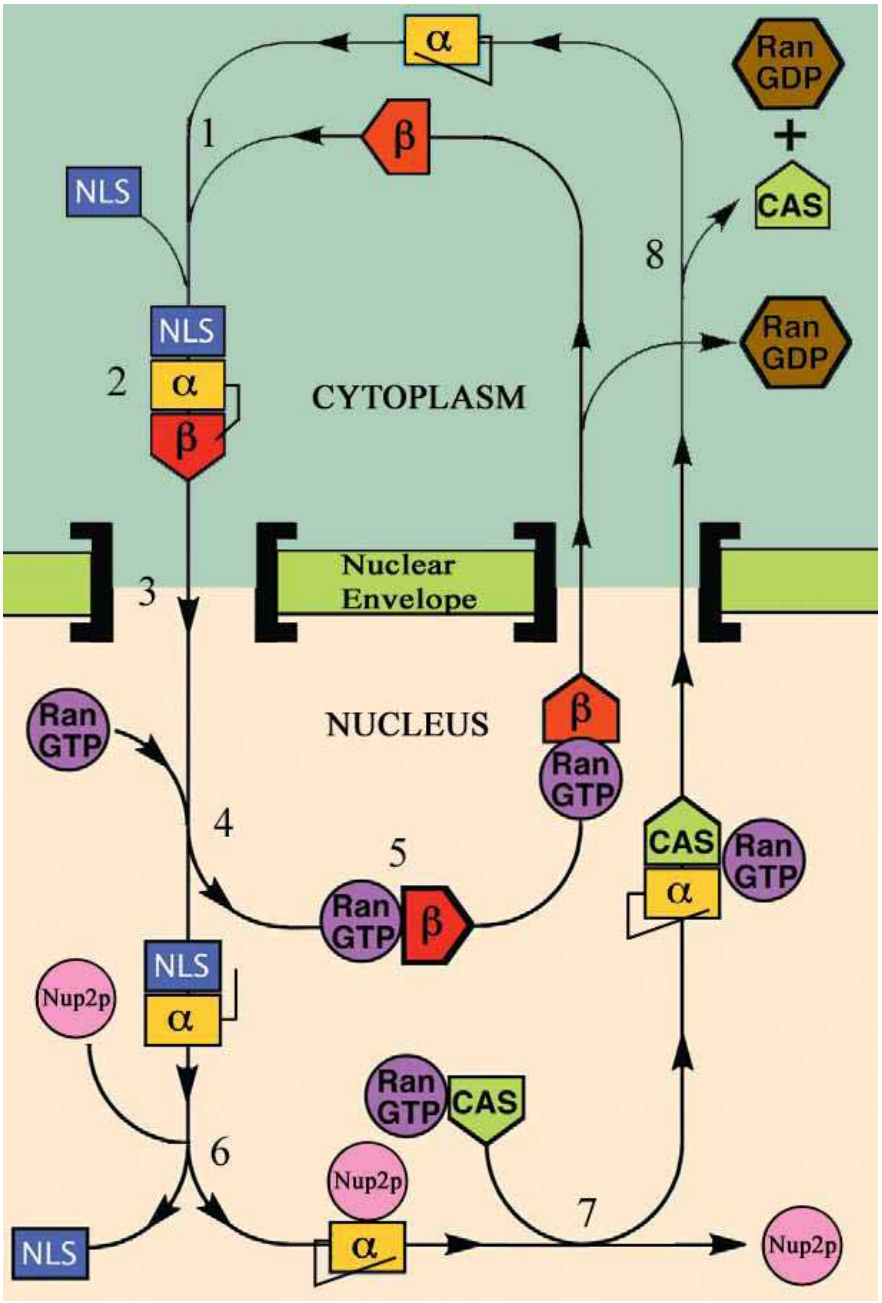

Figure 1.1: A schematic representation of the classical nuclear import cycle.

The numbering of each process in the import and export pathway corresponds to the description given in Section 1.1.1. Importin-α ('α' in picture) is shown with the autoinhibitory tail region. The figure has been adapted from Conti *et al.* 2006 [19].

1.1.2 The nuclear transport factor - 2 and Ran

The nucleocytoplasmic transport process is powered by a steep gradient of nuclear RanGTP and cytoplasmic RanGDP. The RanGTP molecules depleted from the nucleus after translocating to the cytoplasm in complex with transport molecules need to be

constantly replenished. After RanGTP is transported to the cytoplasm in complex with importin-β or the CAS export receptor, it is acted on by the RanGAP which stimulates GTP hydrolysis in conjunction with RanBP1, which dissociates RanGTP from importin-β [14]. This process generates cytoplasmic RanGDP which has a lower affinity for transport molecules than RanGTP.

Cytoplasmic RanGDP is transported back to the nucleus by the nuclear transport factor-2 (NTF2) where it is converted to RanGTP by the chromatin-bound guanine exchange factor (RanGEF), RCC1 [9]. The compartmentalisation of RCC1 in the nucleus and RanGAP in the cytoplasm ensures the steep RanGTP/RanGDP gradient is maintained.

In contrast to the classical import pathway, the Ran transport pathway relies on the binding specificity of NTF2 for RanGDP as opposed to RanGTP. The X-ray crystal structure of NTF2 revealed a dimer in which each subunit contains a hydrophobic pocket that interacts with the switch I and switch II regions of RanGDP (see Figure 1.2) [20]. A third RanGDP molecule is observed in the crystal structure; however, this is likely an artefact from crystal packing interactions. The switch regions in RanGDP have a different conformation in RanGTP and are thought to hinder the formation of an NTF2:RanGTP complex [21]. However, it has recently been shown that the switch loops of RanGDP are capable of adopting similar conformations as observed in RanGTP when bound to importin-β [22]. This perhaps suggests that there are additional mechanisms that govern NTF2 specificity for RanGDP.

The X-ray crystal structure of yeast NTF2 bound to peptides possessing an FxFG motif (FSFG) has previously been solved [23]. The peptides mimic FxFG repeats commonly found in nucleoporins that line the NPC. The NTF2 dimer presents two binding sites for the FxFG peptides that are located in hydrophobic cavities between the NTF2 subunits (see Figure 1.2). These binding sites should not interfere with the NTF2:RanGTP complex, allowing simultaneous interaction between NTF2 and its cargo (RanGDP) as well as FG-nucleoporins, as expected when NTF2 transverses the NPC.

It has recently been suggested that in addition to RanGDP and FG-nucleoporins, NTF2 may have other cargo protein binding partners. Using immunoprecipitation assays, pull down assays and fluorescence binding assays, NTF2 was shown to directly interact with an actin-binding protein, CapG, which is known to localise to the cytoplasm and nucleus [24]. Also, nuclear import assays showed that NTF2 promotes the nuclear import of CapG. This finding could expand the role of NTF2 in the nuclear import cycle to include cargo transport along with the importins and exportins, meaning it is not only confined to maintaining the RanGTP/RanGDP gradient as is commonly thought.

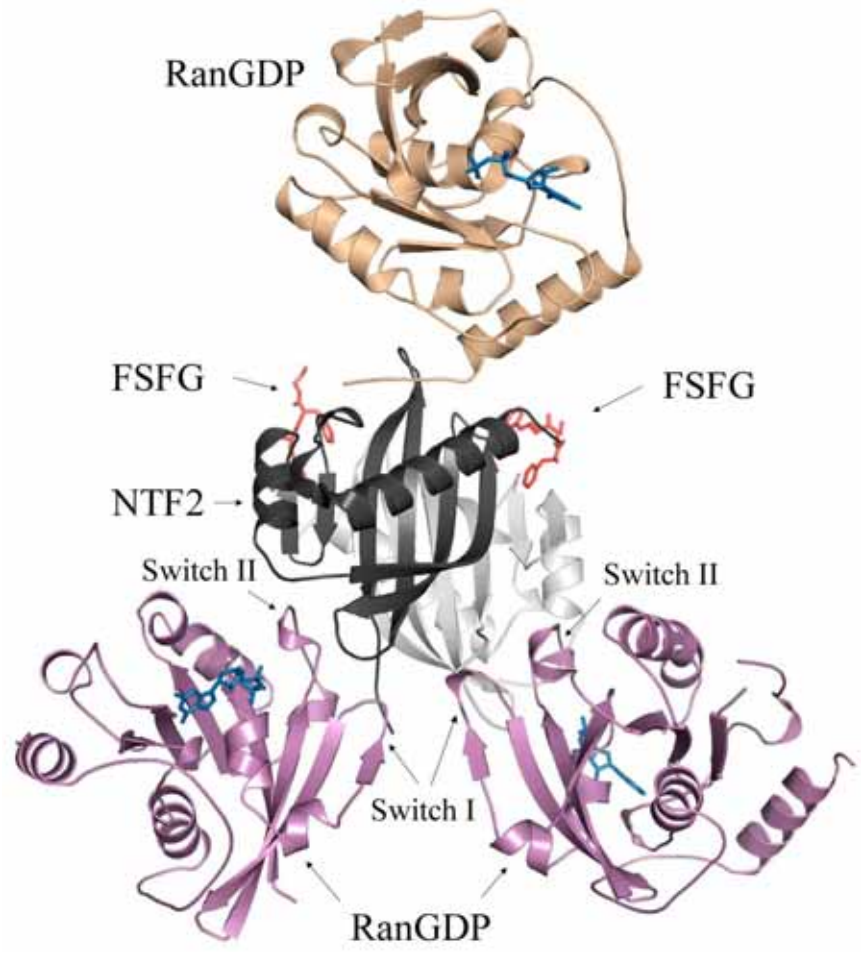

Figure 1.2: The superimposed structures of homodimeric NTF2 bound to RanGDP and FSFG peptides.

The NTF2:RanGDP complex (PDB: 1A2K) is shown, superimposed with bound FSFG peptides from the NTF2:FSFG complex (PDB: 1GYB). The two Ran subunits that form the biologically relevant complex with NTF2 are coloured purple, while the pale brown Ran subunit is likely the result of crystal packing. The GDP molecules bound to Ran (blue) and FSFG peptides (red) are shown in stick representation. The two NTF2 subunits forming the homodimer are coloured black and grey.

1.1.3 Importin-β

The largest family of receptors capable of ferrying substrates across the NPC is the importin-β superfamily, commonly referred to as the β-karyopherins. Proteins in this family have a molecular mass of ~100 kDa. The cell has an extensive representation of β-karyopherins, with 14 members identified in yeast (*Saccharomyces cerevisiae*), and at least 22 members in humans [25]. Of this group of proteins, importin-β (or Kap95 in yeast) has been the most extensively studied member. It has been shown that a majority of these karyopherin proteins including importin-β, CAS, Msn5p, transportin and exportin-1, have a role in the nuclear transport process serving as either importers or exporters, and sometimes both [25]. Their ability to interact with an array of nucleoporins in the NPC has been extensively studied. It has been found that importin-β contains two distinct FG-repeat core binding sites: one at the N-terminus [12, 26] and the other at the C-terminus [27]. These are located on the outer convex surface of the HEAT module (see Figure 1.3).

It has been found that importin-β can act as a direct carrier of cargo proteins by interacting with NLSs at the inner surface [28, 29], combining its function as both a transporter and receptor. With FG-nucleoporins binding in hydrophobic pockets on the outer surface, the inner concave surface of importin-β is free to participate in binding to cargo proteins in a manner that does not disrupt the translocation mechanism.

A unique feature of importin-β among import carriers is that it is capable of interacting with many more NLS bearing proteins indirectly by way of an adaptor receptor known as importin-α. Together these proteins can form the importin-α:β heterodimer [11]. In this configuration, importin-α is capable of recognising a vast array of NLSs that can dock to the heterodimeric complex and be shuttled across the NPC. Structural studies have shown that the mechanism of interaction between importin-β and a substrate is different to the well defined interaction between classical NLSs and importin-α. There are also a greater number of cargo residues participating in complex formation in the importin-β:cargo complexes (~30 residues in importin-β:PTHrP, [29]) compared to importin-α:NLS complexes (~6 residues in

importin-α:SV40 NLS [30], and ~18 residues in the importin-α:nucleoplasmin NLS [31]).

1.1.3.1 Importin-β structures

Importin-β is an all α-helical protein that consists of a repeating structural motif of approximately 50 residues known as the HEAT repeat motif. The HEAT motif derives its name from a series of proteins in which it was first identified (Huntington, elongation factor 3, PR65/A subunit of protein phosphatase and the TOR lipid kinase) [32]. Each HEAT repeat forms two α-helices that stack upon adjoining helices to produce a large superhelical protein. Structures of importin-β binding a functionally diverse range of proteins have been determined, including cargo proteins [11, 28, 29, 33], nucleoporins [12, 34], RanGTP [13, 35] and RanGDP [22]. Although there is currently no structure of the unbound state of importin-β, by analysing importin-β:cargo complexes in addition to small-angle X-ray scattering (SAXS) experiments, it has been suggested that a major factor in the importin-β recognition of cargo is an intrinsic flexibility in the helicoidal pitch that allows an induced fit binding mechanism [19]. This hypothesis that structural flexibility has a role in binding specificity is supported by other results that show the importin-β like protein transportin and the exportin Xpo-t also undergo conformational changes when bound to RanGTP [36].

In the structures of importin-β:cargo complexes, the mode of binding as well as the location of the importin-β binding site is quite variable. The importin-β-binding (IBB) domain of importin-α binds as an approximately 30 residue long structured α-helix in the importin-β inner concave surface near the C-terminus [11] (refer to Figure 1.3). Importin-β coils around the extended IBB domain forming an extensive network of contacts including electrostatic interactions that are made between a cluster of basic residues in the IBB domain and a region of conserved acidic residues on HEAT repeat 8. The cluster of negatively charged residues in the IBB domain corresponds to the binding interface seen in the auto-inhibited importin-α structure [37]. The equivalent IBB domain in snurportin1, an importin-α-like protein, binds in almost identical manner to importin-β [33].

Importin-β binds the cargo protein, sterol response binding protein-2 (SREBP2), with a hydrophobic set of interactions in the inner concave surface of importin-β. These interactions are concentrated at HEAT repeat 17 in importin-β, where a conserved phenylalanine (Phe752) and valine residue (Val755) form an important hydrophobic interface with SREBP2 tyrosine residues. Importin-β effectively holds two helices of the SREPB2 dimer in a configuration like molecular chopsticks [28].

Importin-β also binds the cargo substrate parathyroid hormone-related protein, or PTHrP [29]. In this case, the binding interaction occurs via a network of hydrogen bonds involving the PTHrP main chain and importin-β glutamine and asparagine arrays, not dissimilar to NLS interaction at the binding sites of importin-α. Approximately 30 residues of PTHrP bind in an extended conformation within the importin-β inner concave surface over HEAT repeats 2-11. This binding site significantly overlaps with the binding site of RanGTP.

The interaction between importin-β and nucleoporins occurs over the outer convex surface, where the nucleoporin FG-repeats bind in hydrophobic pockets. This is observed in the structure of importin-β complexed with the yeast nucleoporin nup1p [34], shown in Figure 1.3. In contrast, the importin-β interaction with Ran takes place over the inner concave surface of the importin-β superhelix towards the N-terminus. This means the RanGTP binding site overlaps with certain cargo proteins and that it only partially overlaps the importin-α IBB domain binding site. The interaction is observed in the structures of importin-β:RanGTP [13] and importin-β:RanGDP [22].

The structures of importin-β bound to its substrates complement the current understanding of the nuclear import model as shown in Figure 1.1. When importin-β transverses the NPC bound to cargo it must simultaneously interact with FG-nucleoporins, consistent with the observation that the binding sites for these proteins are on the inner and outer surface respectively. RanGTP facilitates the release of importin-β cargo in the nucleus, consistent with the overlap of importin-β binding sites for RanGTP and cargo substrates such as importin-α, SREBP2 and PTHrP.

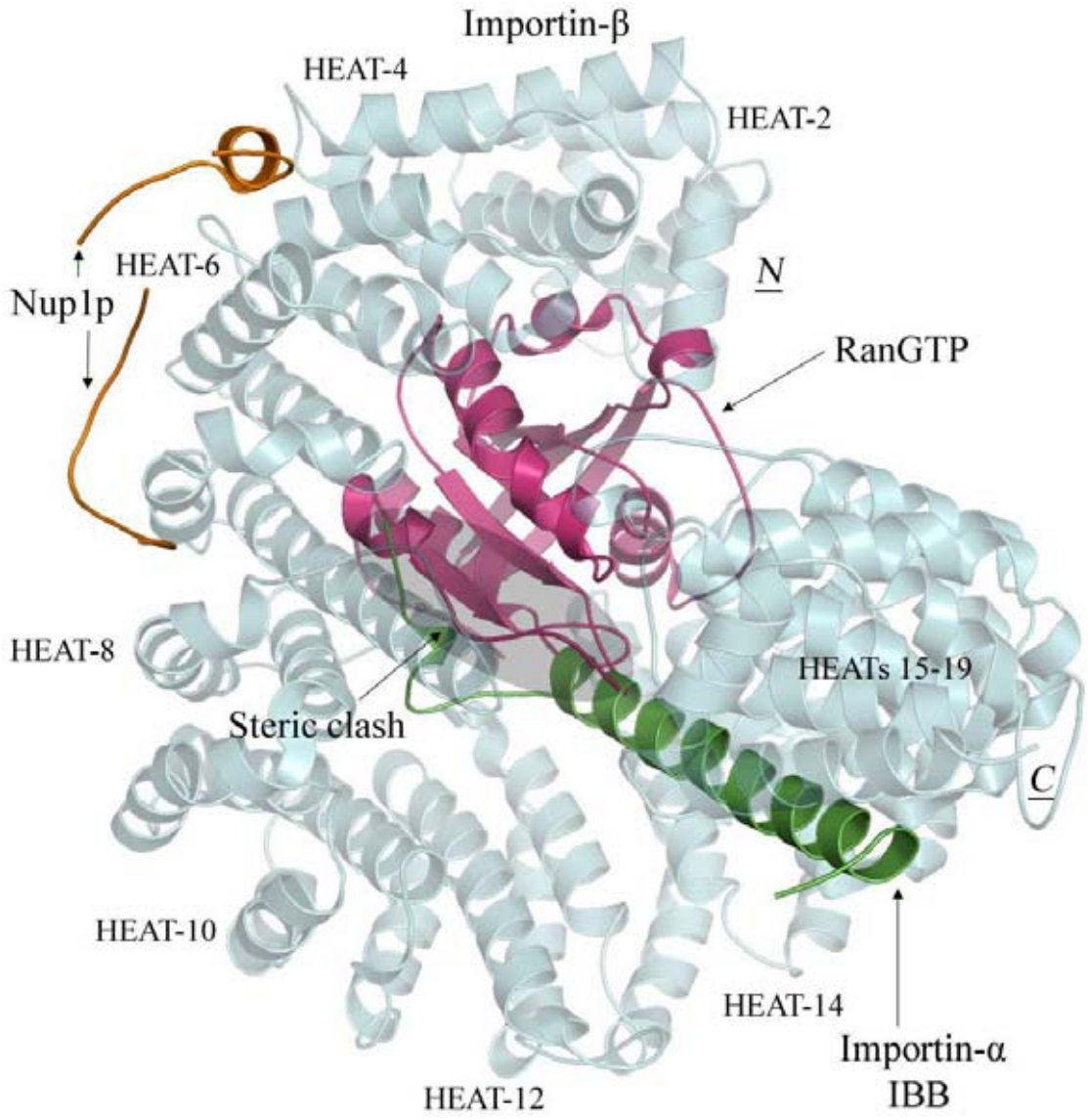

Figure 1.3: The superimposed structures of importin-β bound to the importin-α IBB domain, FG-nucleoporins and RanGTP.

The RanGTP (pink, PDB:2BKU) and Nup1p (orange, PDB:2BPT) molecules from importin-β bound structures are superimposed onto the importin-β:importin-α IBB complex (importin-α IBB domain green, importin-β cyan, PDB:1QGK). Importin-β HEAT repeats are labelled where appropriate. The N-terminus and C-terminus are also labelled. A hypothetical steric clash between the importin-α IBB domain and RanGTP is highlighted in grey.

1.1.4 Importin-α

Importin-α (Kap60 or Srp1p in yeast) is a ~60 kDa protein that has been extensively studied in the nucleocytoplasmic pathway where it functions as an adaptor molecule that recognises and binds to numerous NLS bearing proteins (*reviewed in* [38, 39]). It appears to be incapable of directly navigating across the NPC by

interacting with FG-nucleoporins like the β-karyopherins. To accomplish the task of transporting cargo to the nucleus, importin-α enlists the aid of importin-β by directly binding to the transporter to form a stable heterodimer, importin-α:β. The ternary complex consisting of importin-α:β-cargo, where the cargo is bound to the importin-α subunit, is then able to translocate to the nucleus through the NPC, mediated by importin-β interactions with FG-nucleoporins [11].

Importin-α contains two well defined structural domains: an unstructured 70 residue long N-terminal domain and a larger 460 residue long all α-helical C-terminal domain. It is the flexible N-terminal domain of importin-α that interacts with importin-β and is therefore denoted as the importin-β-binding domain (IBB) [11, 40]. In contrast, the C-terminal domain of importin-α is structured into repeating sets of three α-helices known as armadillo (ARM) repeats which tandemly stack to create a superhelical tertiary structure. The ARM structural motif nomenclature derives its name from the founding member of the structural class, the segment polarity gene *armadillo* which was initially discovered in *Drosophila melanogaster* [41, 42]. Importin-α has ten ARM repeats with each ARM domain containing approximately 40 residues which fold into three helices (H1, H2 and H3) that form a roughly triangular cross-section. They are structurally similar to the HEAT motif found in importin-β, having diverged from a shared evolutionary origin [32]. The repeating ARM domain creates an inner concave surface lined by the H3 helices which present a favourable docking platform ready for interactions with binding partners.

The role of importin-α as a nuclear import receptor is not unique to the ARM structural family of which it is a member. One other well studied member of the group is the superhelical protein β-catenin, which is involved in the Wnt growth factor signalling pathway [43]. β-catenin forms complexes that present similar binding characteristics as seen in importin-α:NLS structures [44], and it also capable of nuclear import and export independent of other transport factors like the β-karyopherins [45, 46]. While β-catenin does not interact with importin-α or importin-β, its nuclear accumulation is inhibited by importin-β suggesting there is an overlapping interaction in the import pathway, possibly at the NPC when interacting with FG-nucleoporins [47].

1.1.4.1 The importin-α C-terminal domain

The large superhelical C-terminal domain of importin-α (residues 70-529 in *Mus musculus*, isoform α2) is highly conserved across species, where mouse importin-α2 shares a sequence identity of 46% with yeast importin-α (Kap60) and 94% with human importin-α2 (Figure 1.4). Accordingly, the structure of importin-α is also conserved across species and allows for functional and structural comparisons: mouse (α2) and yeast (Kap60) structures have an RMSD of 1.72 Å across 379 C_α atoms; mouse (α2) and human (α2) have an RMSD of 1.47 Å across 373 C_α atoms.

The C-terminal domain consists of ten ARM structural repeats that form two well characterised cargo binding sites referred to as the major or minor binding sites. These sites are both located in slight depressions on the concave face of the protein near regions of invariant tryptophan and asparagine arrays (Figure 1.5 and Figure 1.6). The major binding site is situated towards the N-terminus and spans H3 helices in ARM repeats 2, 3 and 4. It is denoted the major binding site due to a higher affinity for NLSs and its ability to bind longer stretches of substrate residues commensurate with a larger number of bonding interactions. The minor binding site spans H3 helices in ARM repeats 7 and 8 and has a lower affinity for NLSs. The minor binding site interacts with shorter stretches of residues compared to the major binding site. Both sites participate in interaction with bipartite NLS sequences, while monopartite NLSs predominantly interact with the major binding site [31].

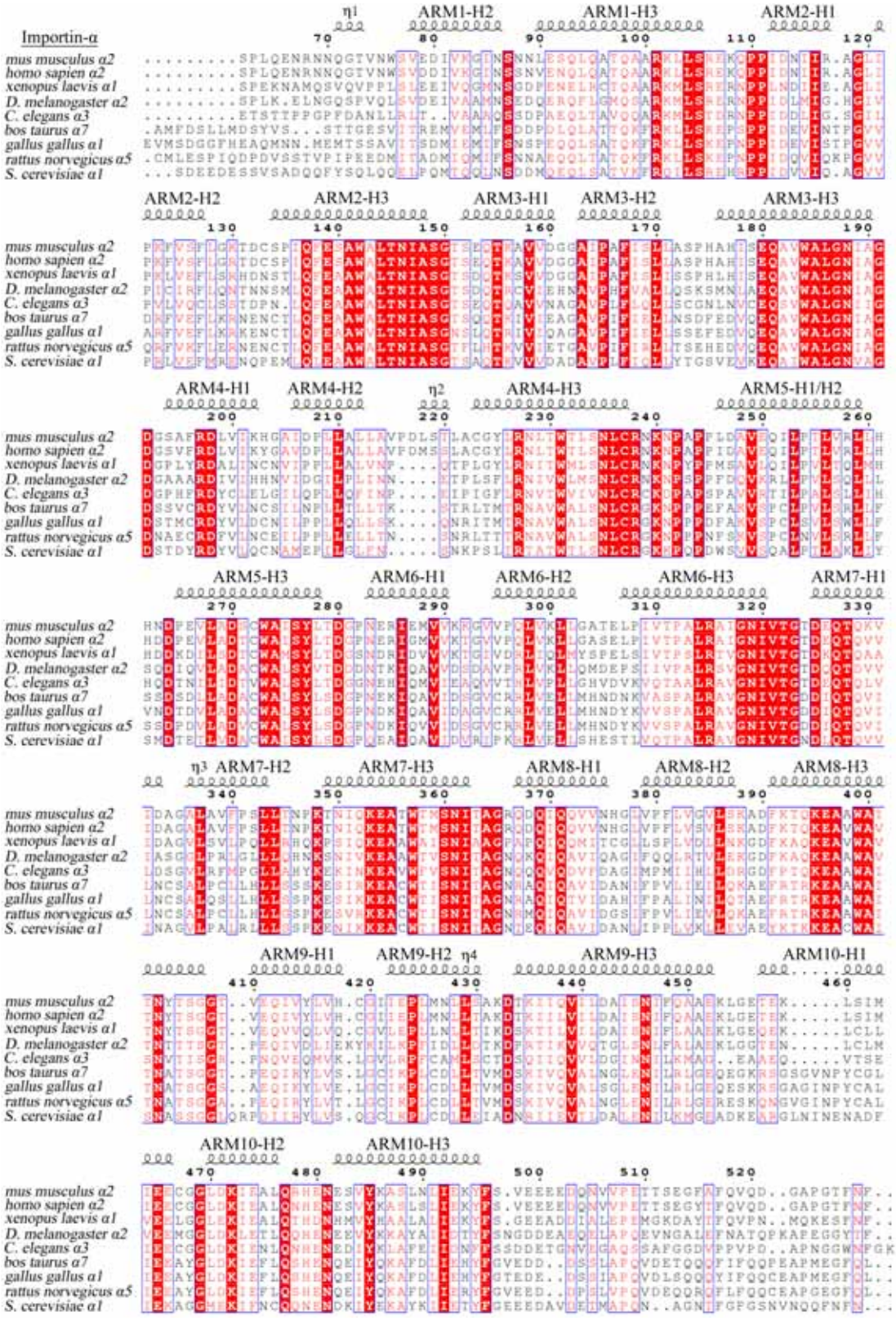

Figure 1.4: A sequence alignment of importin-α homologues over the C-terminal domain.

Conserved residues are red and invariant residues are white with a red background. Secondary structural elements and sequence numbering correspond to *Mus musculus* importin-α2. There are four 3_{10} helices denoted by η1-4, α-helices are denoted by ARM repeat and H1-H3. Multiple sequence alignment performed using CLUSTALW [48] and presented using ESPript [49].

A notable feature of importin-α is the conservation of surface exposed residues in the C-terminal domain (Figure 1.5). The highly conserved surface residues are concentrated on the importin-α inner concave face, which corresponds to the ARM H3 helices (see Figure 1.4 and Figure 1.5B). This includes the importin-α major binding site which is shown in Figure 1.5A bound to the founding member of the classical NLSs, the SV40 NLS. By comparison, the outer convex surface which corresponds to the H1 and H2 helices does not contain a significant number of conserved residues (see Figure 1.5C). Interestingly the conserved surface includes a region between the major and minor binding sites that does not interact with monopartite NLSs and only weakly interacts with bipartite NLSs [50]. However, this central region may be significant in governing interactions with nucleoporins which bind importin-α over the conserved patch of residues in a mostly non-NLS binding mode [15, 51].

Considering it is more common to find conserved residues in the buried and hydrophobic regions of a protein, the prevalence of conserved residues that are solvent exposed in importin-α clearly defines a functionally essential ligand binding site. The recognition of NLSs in this site is fundamentally governed by a conserved and repeating motif of tryptophan and asparagine residues of the form WxxxN, where x is any residue (Figure 1.4). This exact motif is present in the H3 helix of ARMs 2-4 and ARMs 7-8, which essentially divides the importin-α binding region into a major binding site and minor binding site, respectively. The central H3 helices break the repeating motif, where ARM 5 replaces the asparagine with a tyrosine (WxxxY) and ARM 6 replaces the tryptophan with an arginine (RxxxN).

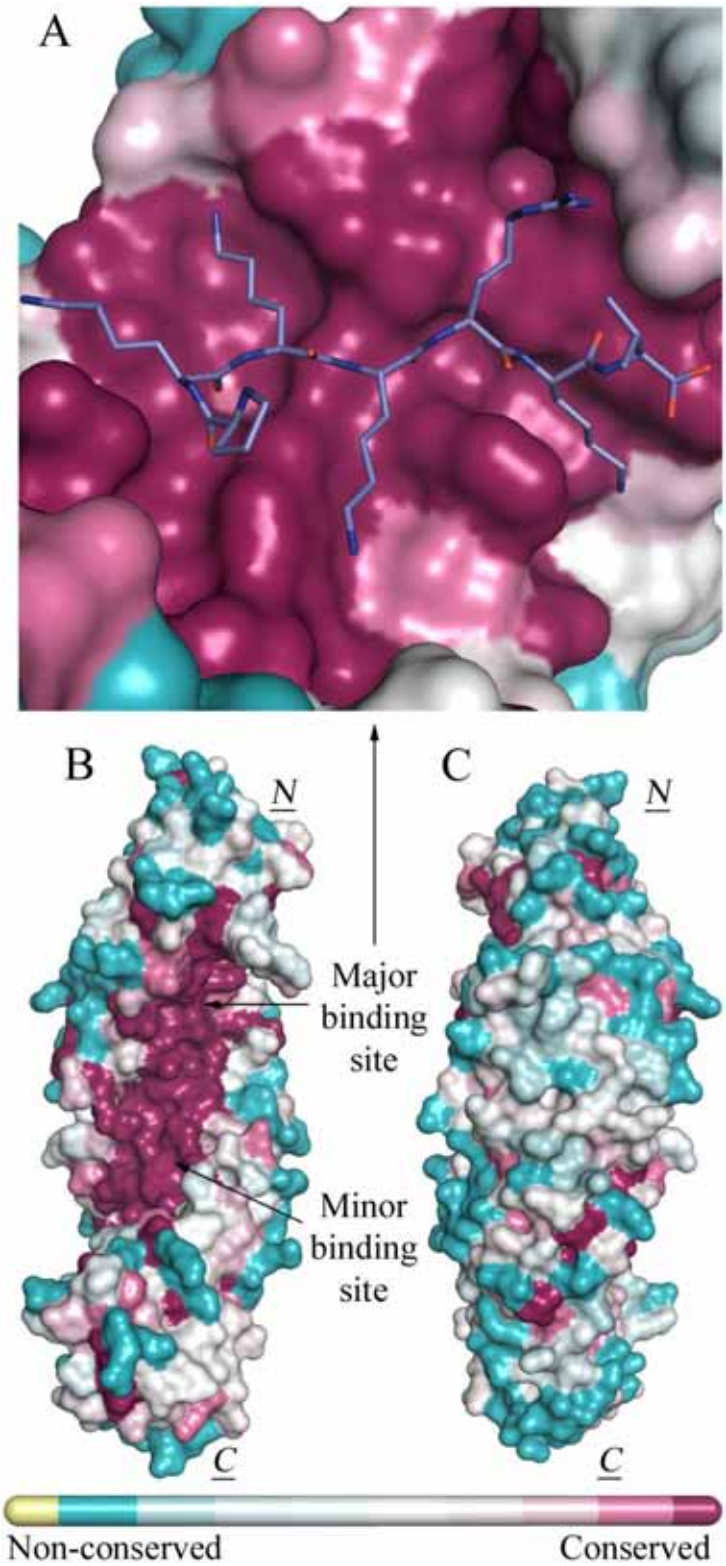

Figure 1.5: The molecular surface of importin-α coloured by conservation score.

The colour bar is shown at the bottom, varying from non-conserved (yellow) to conserved (purple). Residue alignment was performed using CLUSTALW [48] over 30 importin-α sequences in 15 species. Conservation scores were calculated using Bayesian methods on the ConSurf server [52].

A) The bound SV40 NLS (blue, stick representation) in the importin-α major binding site. The structure shown is our importin-α:SV40 NLS complex, described in chapter 4. **B)** The inner concave surface of importin-α. Major and minor binding sites are labelled. **C)** The outer convex surface of importin-α.

The repeating WxxxN motif is often referred to as the tryptophan and asparagine array. The invariant tryptophan array includes residues W142, W184, W231, W273 in the major binding site and residues W357 and W399 in the minor binding site (Figure 1.6). These tryptophans present solvent exposed indole side chain groups that form well defined hydrophobic pockets on the protein surface. The invariant asparagine array includes residues N146, N188 and N235 in the major binding site and residues N361 and N403 in the minor binding site. The asparagines main role in cargo binding is to form a network of critical hydrogen bonds to the main chain of an NLS.

In order to bind in a favourable manner with the asparagine and tryptophan arrays, an NLS must present itself in an extended conformation. This therefore applies a layer of regulation in determining which NLS sequences can interact with importin-α. It is difficult to envisage how a structured NLS, made up of secondary structural elements (α-helix, β-strand), could interact with the importin-α binding sites, and indeed there are currently no structures of such a complex. Further substrate specificity is conferred by a number of invariant acidic residues near the major binding site (E180, D192, E266, D270) and minor binding site (D325 and E396) as shown in Figure 1.6. These negatively charged residues confer favourability for an NLS sequence containing a cluster of basic residues due to possible long range electrostatic interactions.

Since the modes of interaction between an NLS and importin-α binding sites are so well defined for both the main chain and side chains, a naming scheme was developed based on the positions of bound NLS residues [30]. The naming scheme is applicable to many, if not all, importin-α:NLS complexes and is now routinely used in the literature. In the major binding site the positions of six NLS residues are labelled (P1 - P6), and in the minor binding site the positions of two NLS residues are labelled (P1′ - P2′) (see Figure 1.6). These binding slots follow the directionality of the bound NLS from N-terminus to C-terminus which runs anti-parallel to importin-α. A slight variation of this nomenclature has sometimes been used in the literature, which shifts the major binding site position P1 one residue towards the C-terminus [39, 53].

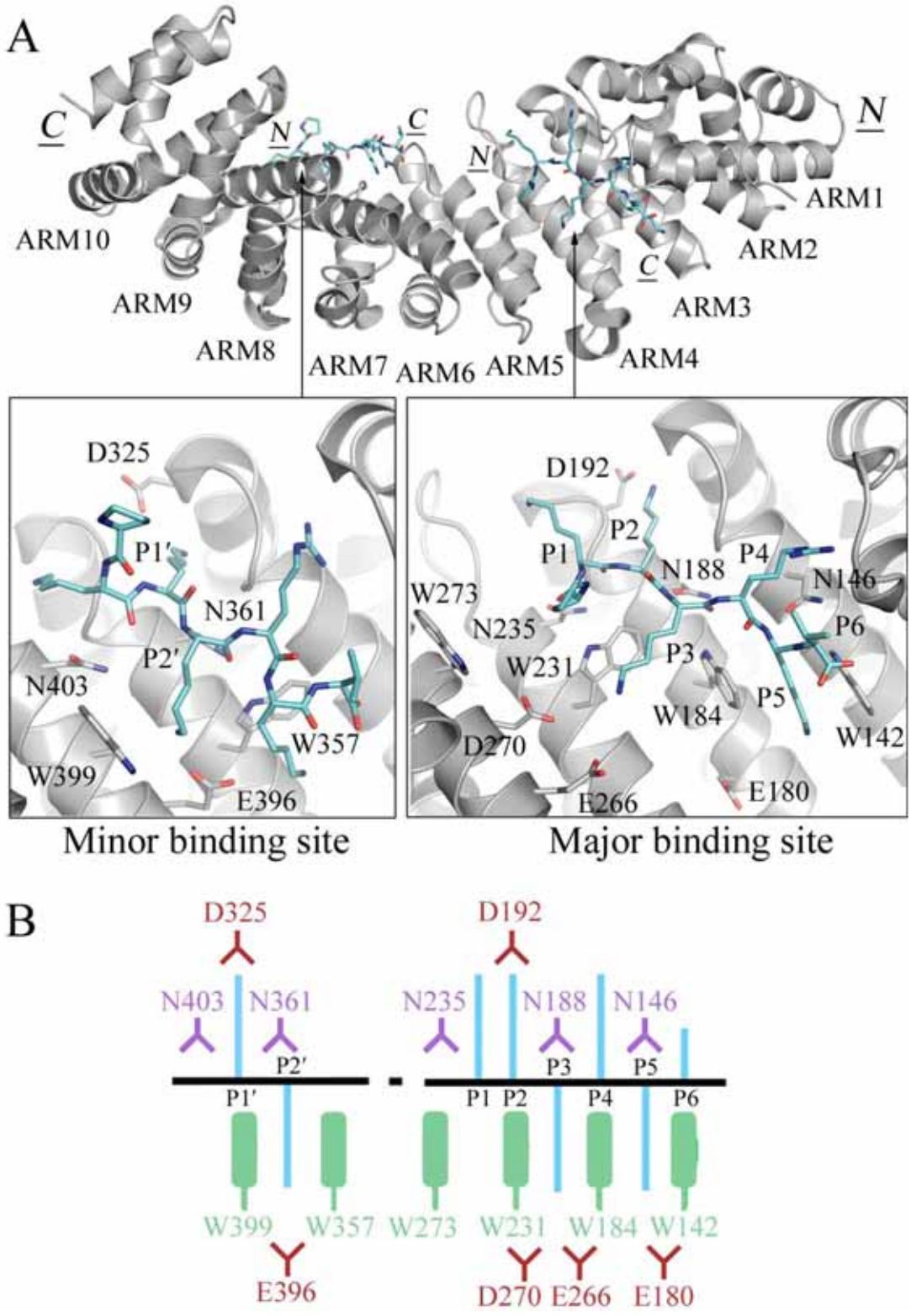

Figure 1.6: The major and minor binding sites in importin-α.

A) A structure of the importin-α:SV40 NLS peptide complex is shown (PDB:1EJL). Importin-α is shown in cartoon representation (grey) and the SV40 NLS (PKKKRKV) is shown in stick representation (cyan). The termini of importin-α and both SV40 NLS peptides are labelled. *Inset*: The major and minor binding sites are magnified. Key importin-α residues and binding slots are labelled.

B) Schematic of importin-α binding sites corresponding to panel A. The minor (P1′ and P2′) and major binding site (P1 - P6) are shown.

1.1.4.2 The importin-α N-terminal domain

The unstructured and flexible N-terminal domain of importin-α (residues 1-69 in *Mus musculus*, isoform α2) mediates the interaction with importin-β when forming the importin-α:β heterodimer, and has thus been designated the importin-β-binding (IBB) domain. The ability to directly bind to importin-β is necessary for the involvement of importin-α in nucleocytoplasmic transport, as importin-α is unable to independently navigate the NPC like β-karyopherins. A cluster of basic residues in the importin-α IBB domain resembles the classical NLS moiety recognised by the importin-α C-terminal domain substrate binding sites, and it has been shown that the IBB domain is necessary for the normal transport function of importin-α [40, 54]. These findings reveal importin-α is a unique receptor which uses an intramolecular NLS containing domain to regulate transport. The IBB domain of mouse importin-α2 shares a sequence identity of 52% with yeast importin-α (Kap60) and 95% with human importin-α2 (see Figure 1.7).

Gorlich *et al*. studied the IBB domain by generating a series of peptides that correspond to the N-terminal amino acid sequence of *Xenopus laevis* importin-α [55, 56]. They found that a highly conserved stretch of 41 residues (residues 10-50 in *X. laevis*) make up the IBB domain that specifically binds to importin-β (refer to Figure 1.7). Experiments measuring import activity showed that this was the shortest stretch of residues necessary to ensure nuclear import of IBB fusion proteins.

The molecular basis for direct binding of the IBB for importin-β was confirmed when the X-ray crystallographic structure of the human importin-α:β complex was solved [11]. The structure, seen in Figure 1.3, consists of a highly basic 44 residue long peptide corresponding to amino acids 11-54 in importin-α, complementing the *in vitro* import assay results obtained by Gorlich *et al*. [55]. The majority of charged residues in the IBB domain form salt bridges with importin-β, stabilising the interaction. A subset of these are highly conserved and critical to binding importin-β, particularly Arg13 and Lys19 at the N-terminus of the IBB domain, and Arg51 at the C-terminus (Figure 1.7). These electrostatic interactions to conserved aspartate residues in importin-β are facilitated by inserting the aliphatic portion of the negatively charged

side chain into a hydrophobic pocket on the importin-β surface, formed by conserved tryptophan or phenylalanine residues. This side chain interaction is also characteristic of the interaction between importin-α and NLSs.

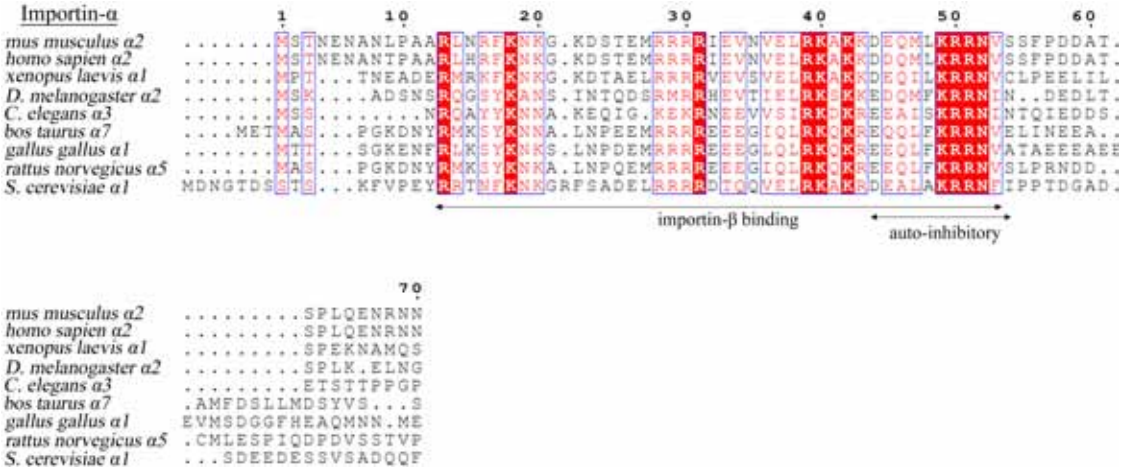

Figure 1.7: A sequence alignment of the N-terminal domain in importin-α homologues.

Conserved residues are red and invariant residues are white with a red background. The region of the IBB domain that directly binds to importin-β is marked (residues 13-53), as well as the region that participates in auto-inhibitory function (residues 44-54). Multiple sequence alignment performed using CLUSTALW [48] and presented using ESPript [49].

In addition to the role of the importin-α N-terminal domain in mediating interaction with importin-β, it is also known to regulate NLS binding to the importin-α C-terminal domain. This secondary role was first suggested after liquid phase binding assays showed an interaction between the importin-α substrate recognition sites and a peptide corresponding to human importin-α1, residues 23-49 (residues 26-52 in mouse importin-α2) [54]. Furthermore, this peptide which includes the conserved ^{49}KRRN^{51} NLS-like motif was also shown to compete with SV40 NLS binding, suggesting the sequence could participate in auto-inhibition.

A full length mouse importin-α (isoform 2) X-ray crystal structure solved in the absence of NLS cargo revealed the nature of the intrasteric inhibition mechanism of the IBB [37]. Importin-α residues 44-54 (see Figure 1.7) were found to bind to the major binding site in the importin-α C-terminal domain, making interactions consistent with those seen in the previously solved cargo bound structure of the yeast importin-α:SV40 NLS complex [30] and importin-α:β heterodimer. In particular, the conserved residues ^{49}KRRN^{51} make the most significant main chain and side chain

interactions, positioned in the critical binding slots P2 - P5 (see Figure 1.6). Electron density in the importin-α minor binding site could not be modelled definitively and therefore it was not confirmed whether or not the downstream conserved motifs, such as ^{39}RKxK^{42} or ^{28}RRRR^{31}, participate in binding to form a bipartite auto-inhibition NLS [37].

Of importance is an earlier structure of yeast importin-α (residues 88-530) which was solved in a cargo free form (or *apo* form) as a homodimer with the importin-α construct used for structure determination lacking the N-terminal IBB domain [30]. Therefore, there was no chance of observing the IBB domain auto-inhibitory mechanism as seen in mouse importin-α. However, an alternative auto-inhibition mechanism was hypothesised. Since each subunit of the importin-α dimer was observed to obscure the minor binding site and partially obscure the major binding site of the other subunit, the process of homodimerisation could also interfere with importin-α:cargo interactions. With the major binding site still partially available to bind NLSs, such an auto-inhibition mechanism could complement auto-inhibition from the IBB domain rather than replace it. Dimerisation has not been observed in mouse importin-α or human importin-α structures, meaning any functional significance of importin-α dimerisation may be specific only to yeast.

There has currently been no further supporting evidence for a dimerisation induced auto-inhibition mechanism for importin-α. The structure of the yeast importin-α:Cse1p:RanGTP export complex does show an IBB domain auto-inhibition mechanism for yeast importin-α similar to that seen in mouse importin-α [17]. In this yeast importin-α complex, basic motifs in the IBB domain interact with both the major and minor binding sites. At the major binding site, the conserved motif ^{54}KRRN^{57} is located in binding slots P2 - P5 (identical to mouse importin-α) while at the minor binding site the motif ^{36}RRRR^{39}, conserved in vertebrates and yeast, forms critical interactions that were not observed in the auto-inhibited mouse importin-α structure.

In terms of the nuclear import cycle, the importin-α auto-inhibitory region acts like a regulatory switch for NLS binding between a high affinity form (empty binding site) in the cytoplasm and a low affinity form (occupied binding site) in the nucleus.

This is supported by evidence that shows importin-α has an increased affinity for NLSs after it binds importin-β through the IBB domain, which contains the auto-inhibitory sequence [54, 57]. This presumably frees up the NLS binding site and switches importin-α to the high affinity form. Further experiments have shown that the removal of the N-terminal domain of yeast importin-α (producing a truncated protein containing residues 88-530) resulted in stronger binding to the SV40 NLS compared to the wild type, full length importin-α [58].

Further *in vitro* binding assays measured that importin-β increases the affinity of full length importin-α for NLSs by several orders of magnitude: from K_d >10 μM to K_d = 33 nM [58]. The binding affinities of monopartite (SV40) and bipartite (nucleoplasmin) NLSs are similar for a truncated importin-α which lacks the IBB domain (K_d = 17 nM and K_d = 14 nM, respectively), and the importin-α:β complex (K_d = 35 nM and K_d = 48 nM) [59]. Control experiments show that the SV40 NLS binds specifically to importin-α and that there is no measurable interaction with importin-β on its own [58].

1.1.4.3 Non-structural analysis of importin-α:NLS interactions

As previously described, the nuclear localisation signal (NLS) refers to a group of well characterised amino acid sequence motifs that are recognised by importin-α. An NLS present in the cargo protein generally consists of either one or two highly basic residue clusters containing lysine and arginine residues. Where there is just one basic motif consisting of ~6 residues, the NLS is referred to as monopartite. Where there are two basic motifs that bind importin-α, the NLS is referred to as bipartite and normally consists of a cluster of ~6 basic residues in the major binding site and two basic residues in the minor binding site separated by between 9-11 residues.

The prototypical NLS was discovered in the Simian virus 40 large T-antigen [60], and is simply referred to as the SV40 NLS. This monopartite NLS sequence contains residues 126-132 from the SV40 large T-antigen, ^{126}P**KKKRK**V^{132}, with basic residues occupying major binding site positions P1 - P5 (*bold*). The bipartite NLS sequence in nucleoplasmin derived from *Xenopus laevis* serves as the founding

member of bipartite classical NLSs [61]. It consists of residues 155-170, ^{155}**KR**PAATKKAGQA**KKKK**^{170} with basic residues occupying the major binding site positions P2 - P5, and minor binding site positions P1′ - P2′.

In addition to the prototypical NLSs, there have been a number of other proteins shown to contain functional NLSs with a range of diverse sequences. A comprehensive *in vitro* sequence analysis was performed by Hodel *et al*. in 2001 [62] which used an alanine scanning technique to probe the relative significance of the amino acids within the SV40 NLS and the c-*myc* proto-oncogene NLS (PA-^{P1}A**KR**V**K**L^{P6}-D). To perform these studies, a truncated importin-α construct was used that lacked the N-terminal domain auto-inhibitory sequence.

The results from the alanine scanning studies of the NLS sequences showed that the SV40 lysine corresponding to the P2 major binding site (K128) was found to contribute the most to binding affinity (SV40 NLS, K_d = 9 nM compared to SV40 K128A, K_d = 3,000 nM, [62]). This is consistent with previous *in vivo* experiments that showed that the NLS function is lost when this residue is mutated, even if an arginine substitutes for the lysine [60]. In contrast, the residues at P3 and P5 favour long basic side chains but are not specific for either arginine or lysine. Their relative energy contribution to NLS binding is approximately equal and ⅔ that of the lysine at P2 [62]. Binding slot P4 which contains an arginine in SV40 and a valine in c-*myc* has a lower relative contribution to binding energy, approximately ½ that of P3 in SV40 and ⅓ that of P3 in c-*myc*. This means the binding slot at P4 is the most tolerant out of the core binding slots P2 - P5, capable of accepting a more diverse range of residues.

The quantification of a bipartite sequence or N-terminal KR basic motif was also demonstrated in the Hodel *et al*. alanine scanning study [62]. It was found that this motif contributed a modest increase to binding energy (~3 kcal/mol) approximately equal to that of the binding energy for a single residue at P3 or P5 in the major binding site. Furthermore this relative binding energy is also approximately equal to the energy of inhibition set by the IBB domain [62]. The smaller energy contributed by the minor binding site is possibly one reason the site is not as well defined when compared with the major binding site. In an important follow-up study, the NLS binding affinity

values obtained with the alanine scanning technique were later shown to correlate well with the *in vivo* import of NLS cargo [63].

A number of studies now support an NLS consensus sequence in the importin-α major binding site to take the form, ^{P2}K(K/R)x(K/R)^{P5} [50, 62, 64-66], where x is any amino acid. The consensus for the minor binding site is $^{P1'}$KR$^{P2'}$. These sequences are short basic motifs reflecting the small contribution to binding energy provided by flanking regions, even in the major binding site P1 and P6 positions. Nevertheless, it has been suggested that the non-basic flanking regions of an NLS may still have an important role in governing a functional NLS [62, 67].

1.1.4.4 Structural analysis of importin-α:NLS interactions

Due to a concerted effort in the nuclear transport field to identify the specific cargo for various nuclear transporters and to elucidate the structural mechanism behind their interaction, many structures of nuclear transport complexes have now been determined by X-ray crystallography. The most common structure of a transport complex seen in the literature is that of an NLS peptide bound to importin-α as either a monopartite NLS (such as the SV40 NLS) or bipartite NLS (such as the nucleoplasmin NLS) (shown in Figure 1.8). While it is rare to find significant differences in NLS binding to importin-α, the importin-α:androgen receptor NLS complex has a novel interaction configuration in that it binds to the major binding site analogous to the SV40 NLS, but then sharply turns away from the minor binding site (see Figure 1.8).

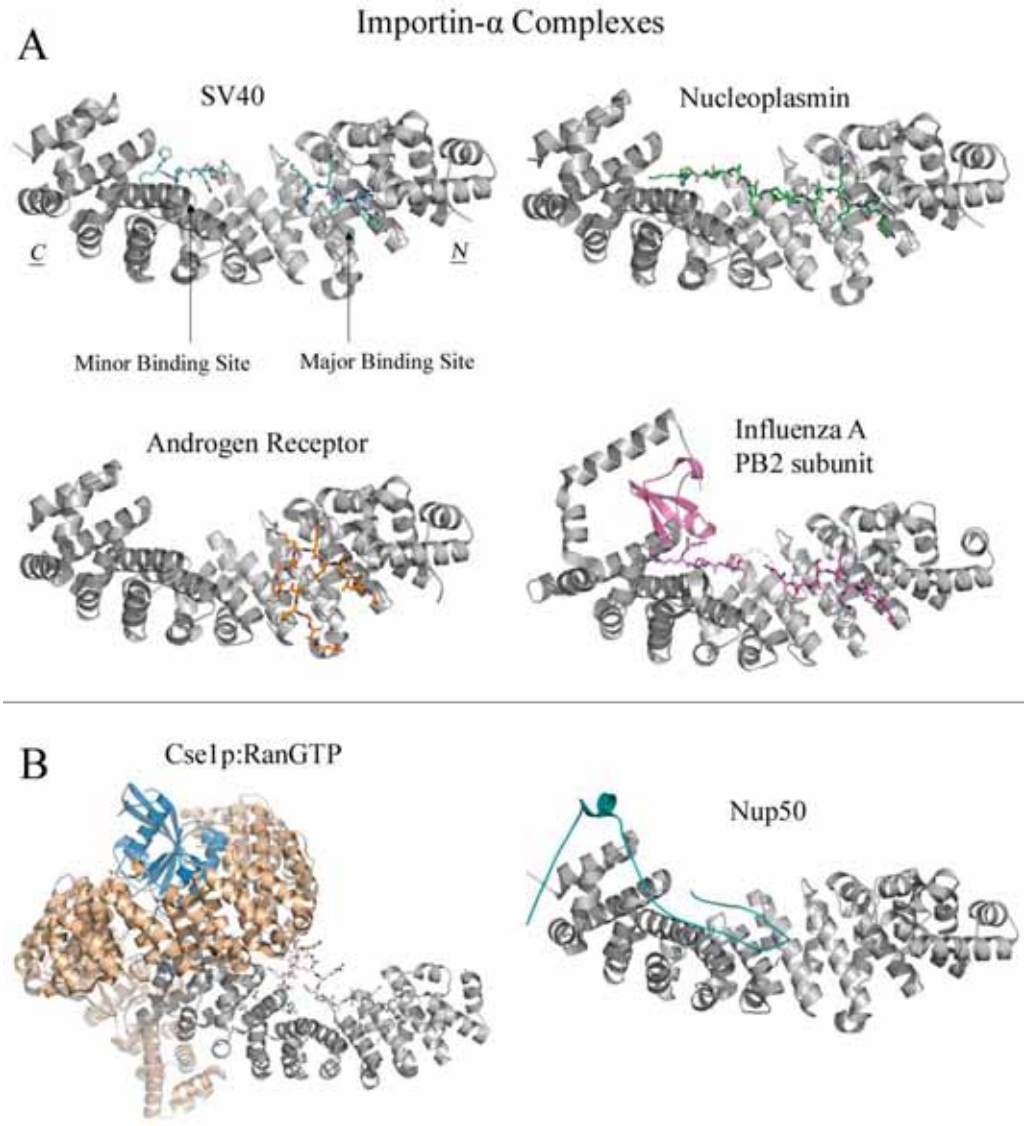

Figure 1.8: Structures of importin-α complexes.

Importin-α is shown in the same orientation in all structures (grey, cartoon representation).

A) Structures of importin-α bound to NLS cargo (stick representation). Includes the SV40 NLS (cyan, PDB:1EJL), nucleoplasmin NLS (green, PDB:1EJY), androgen receptor NLS (orange, PDB:3BTR) and the influenza A PB2 subunit (magenta, PDB: 2JDQ). The N-terminus and C-terminus of importin-α have been labelled in the SV40 NLS complex, as well as the major and minor binding sites.

B) Structures of importin-α bound to non-cargo proteins. The Cse1P:RanGTP export complex (beige and blue respectively, PDB:1WA5) is shown with the importin-α auto-inhibitory region in stick representation. The nup50 complex (teal, PDB:2C1M) is also shown.

Synthetic peptides are often used in structural studies to represent the NLS in the cargo proteins which means the rest of the cargo protein is not present so as to hamper crystallization efforts. The few exceptions to this for importin-α cargo proteins (i.e. excluding nucleoporins and exportins), are the structure of human importin-α5 bound to an 82 residue long PB2 C-terminal domain fragment (residues 678-759) of the

influenza A virus polymerase [68] (see Figure 1.8) and human importin-α1 bound to CBP80 [69]. In the influenza A PB2 - importin-α structure, a bipartite NLS located towards the C-terminus in the PB2 fragment (^{738}**KR**x$_{12}$**KK**I**K**^{735}) with a 12 residue linker binds the minor and major binding sites in a typical unfolded and extended conformation analogous to the nucleoplasmin NLS. However, in addition to this, a folded domain located towards the N-terminus of the PB2 fragment is observed in the structure, mainly interfacing with the importin-α C-terminal domain near ARM repeats 8-10.

The folded portion of the PB2 bound cargo consists of residues 688-735 in a compact α/β structural domain. One interesting aspect of the folded domain is the well coordinated hydrogen bonding between residue Lys718 in the PB2 fragment and residues Gly284, Asn286 and Thr325 in importin-α at the C-terminus of H3 helices in ARMs 5 and 6. This places Lys218 in a position that fits close to the minor binding site asparagine array, in a binding slot similar to P1′. Also of note is a close electrostatic interaction between Glu509 near the importin-α C-terminus and Arg703 in the PB2 binding fragment, which appears to stabilise a conformational change in the ARM 10 H2 and H3 helices in importin-α. The importance of ARM 10 interaction with the PB2 cargo is unclear.

Structural studies of importin-α:NLS complexes have shown them to form in a manner consistent with the defined consensus sequence, ^{P2}K(K/R)x(K/R)^{P5}. A structural interpretation of this consensus correlates well with *in vitro* mutagenesis studies, which in turn correlate well with *in vivo* nuclear import [63]. Inspecting the binding sites of importin-α, there is specificity for positively charged side chains due to negative electrostatic potential from both H3 helix dipoles and more importantly, strategically placed acidic residues around the major binding site. The C-terminus helix dipoles from H3 helices in ARMs 1-4 should favour a basic residue in P1, P2 and P4, although the weak electrostatic interaction does not appear to be a stringent factor in determining residue specificity at these locations. The presence of Asp192, Glu180, Glu266 and Asp270 around the binding site means that a basic residue is favoured at P2, P3 and P5, offering insight into the preference for lysines and arginines in these binding pockets. The only preference for a basic residue at P4 comes from the H3 helix

dipoles of ARMs 1 and 2, and since this electrostatic interaction is expected to be weak it perhaps explains the tolerance of non-basic residues at this location.

The importance of a lysine residue at P2 has been demonstrated by mutagenesis studies that have shown the residue is necessary for a high NLS binding affinity to importin-α and also for an NLS to function in nuclear import [60, 62, 70]. P2 is also unique among the binding pockets in that a lysine and arginine are not interchangeable at this particular location. Not surprisingly an NLS lysine occupies the P2 major binding slot in all determined importin-α:NLS structures solved to date. A structural interpretation of the P2 lysine interaction with importin-α provides an explanation as to why the residue is so significant. At P2, the terminal nitrogen atom of the lysine side chain (N^{ζ}) forms three well coordinated hydrogen bonds to importin-α: to the main chain carbonyl group of Gly150, to the hydroxyl side chain of Thr155, and to the negatively charged side chain of Asp192. This is the only salt bridge formed in NLS interaction with importin-α. The longer side chain of an arginine residue at P2, which cannot substitute for the lysine, would possibly be unable to maintain the favourable side chain hydrogen bonding arrangement and may also force the NLS main chain into an unfavourable position.

A structural interpretation of the binding slots P1 and P6 at the ends of the major binding site suggest there is little preference there for a basic residue. There are few side chain interactions at these locations that could provide residue specificity, consistent with the defined consensus sequence limited to P2 - P5.

Table 1.1 presents a list of the currently known structures of the β-karyopherins (importin-β and transportin) and importin-α proteins. A summary of the major binding site configuration as well as binding affinity (K_d) for importin-α:NLS complexes is listed where both data is available.

Table 1.1: A list of β-karyopherin and importin-α structures

Organism	Karyopherin	PDB	Ligand	Major Binding Site	K_d *(nM)*	Resolution (Å)	References	
				1 2 3 4 5 6			Structure	K_d
Saccharomyces cerevisiae	Importin-α	1BK5	-	-	-	2.2	Conti, *et al*. **1998** [30]	-
	Importin-α	1BK6	SV40	P **K K K R K** V	9-12	2.8	Conti, *et al*. **1998** [30]	[51, 58, 62]
	Importin-α	1EE4	c-*myc*	A A **K R** V **K** L	6	2.1	Conti, *et al*. **2000** [31]	[62]
	Importin-α	1EE5	Nucleoplasmin	Q A **K K K K** L	9	2.4	Conti, *et al*. **2000** [31]	[51]
	Importin-α	1UN0	Nup2p	-	2.1	2.6	Matsuura, *et al*. **2003** [51]	
	Importin-α	1WA5	CSE1p:RanGTP	L A **K R R** N F[a]	-	2.0	Matsuura, *et al*. **2004** [17]	-
	Importin-α	2C1T	Nup2p	-	2	2.6	Matsuura, *et al*. **2005** [15]	
	Importin-β	2BKU	RanGTP		0.3[b]	2.7	Lee, *et al*. **2005** [13]	[71]
	Importin-β	2BPT	Nup1p		8	2.0	Liu, *et al*. **2005** [34]	[72]
	Importin-β	3EA5	RanGDP		1500	2.5	Forwood, *et al*. **2008** [22]	
Mus musculus	Importin-α	1IAL	Auto-inhibited	M L **K R R** N V[a]	-	2.5	Kobe, **1999** [37]	-
	Importin-α	1EJY	Nucleoplasmin	Q A **K K K** - -[c]	1-2	2.9	Fontes, *et al*. **2000**	[15, 59]
	Importin-α	1EJL	SV40	P **K K K R K** V	5-8	2.8		
	Importin-α	1IQ1	Importin-α (44-54)	M L **K R R** N V	4000	2.8	Catimel, *et al*. **2001** [59]	
	Importin-α	1PJM	Retinoblastoma	P L **K K** L **R** G	68[d]	2.5	Fontes, *et al*. **2003** [50]	[73]
	Importin-α	1PJN	N1N2	**K** A **K K** S **K** G	5[d]	2.5	Fontes, *et al*. **2003** [50]	[74]
	Importin-α	1Q1S	SV40-phosphorylated	P **K K K R K** V	24	2.3	Fontes, *et al*. **2003** [67]	
	Importin-α	1Q1T	SV40-non phosphorylated	P **K K K R K** V	10	2.5		
	Importin-α	1Y2A	PLSCR1	- G **K** I S **K** H	45	2.2	Chen, *et al*. **2005** [75]	
	Importin-α	2C1M	Nup50	-	1	2.2	Matsuura, *et al*. **2005** [15]	
	Importin-α	3BTR	Androgen receptor	A **R K** L **K K** L	5000	2.6	Cutress, *et al*. **2008** [53]	
	Importin-α	-	CLIC4	V A **K K** Y **R** N	-	2.0	See chapter 4.	-
	Importin-β	1UKL	SREBP-2		-	3.0	Lee, *et al*. **2003** [28]	-

Organism	Karyopherin	PDB	Ligand	Major Binding Site 1 2 3 4 5 6	K_d *(nM)*	Resolution (Å)	References Structure	 K_d
Homo sapiens	Importin-α	2JDQ	Influenza A PB2 subunit	A T **K R** I **R** M	-	2.2	Tarendeau, *et al.* **2007** [68]	-
	Importin-β	1IBR	Ran GppNHp		-	2.3	Vetter, *et al.* **1999** [21]	-
	Importin-β	1QGK	Importin-α IBB		-	2.5	Cingolani, *et al.* **1999** [11]	-
	Importin-β	1M5N	PTHrP		1-4[b]	2.9	Cingolani, *et al.* **2002** [29]	[29, 76]
	Importin-β	1O6O	NSP1P		100	2.8	Bayliss, R., *et al.* **2002** [12]	
	Importin-β	1O6P	GLFG peptide		200	2.8		
	Importin-β	2P8Q	Snurportin IBB		2	2.4	Mitrousis *et al.* **2008** [33]	
	Transportin	1QBK	Ran GppNHp		2	3.0	Chook, et al. **1999** [77]	[78]
	Transportin	2H4M	M9		2-42	3.1	Lee, et al. **2006** [79]	[78, 79]
	Transportin	2OT8	hnRNPM		10	3.1	Cansizoglu, et al. **2007** [80]	
	Transportin	2Z5J	-		-	3.4	Imasaki, et al. **2007**	-
	Transportin	2Z5K	TAP		17-19	2.6	Imasaki, et al. **2007**	[78, 81]
	Transportin	2Z5N	hnRNP		3	3.2	Imasaki, et al. **2007** [81]	
	Transportin	2Z5O	JKTBP		1000	3.2	Imasaki, et al. **2007** [81]	

[a] Sequences represent the importin-α auto-inhibitory region.

[b] These values do not correspond to the importin-α species used in the structure.

[c] The nucleoplasmin Lys170 side chain does not fit in the P5 binding pocket.

[d] These values are for NLS peptides binding the importin-α:β complex.

1.2 The CLIC protein family

1.2.1 CLIC proteins and chloride channels

Regulation of cellular properties such as ion homeostasis and cell volume is mediated by proteins that allow the free passage of ions across cellular membranes. These ion channel proteins normally provide an avenue for ion transport by creating pores in phospholipid bilayers that allow the passive diffusion of ions to flow across an electro-chemical gradient. Different channels can selectively allow passage for particular ions across the membrane, at the basic level differentiating between negatively charged anions and positively charged cations.

Of the various anions found in organisms, perhaps the most common is the chloride anion, Cl^-. There are a number of well studied and distinct classes of chloride ion channels currently known, including the chloride channel family (CLC), the cystic fibrosis transmembrane conductance regulator (CFTR) and the GABA (γ-aminobutyric acid) and glycine ligand gated receptors (see [82] for a review).

In addition to these groups of chloride ion channels, is the more recently identified chloride intracellular channel (CLIC) protein family. The CLICs have been hypothesised to be able to form functionally significant ion channels after the founding member of the CLIC family, p64 (later known as CLIC5B), was first identified due to its affinity for the chloride ion channel blocker indanyloxyacetic acid 94 (IAA-94) [83, 84]. CLIC proteins appear to be found in all animal species, and are widely distributed across tissue and cell types. All CLICs contain a conserved structural module that consists of approximately 230 residues (see Figure 1.9), although p64 and CLIC6 also contain long N-terminal extension sequences. In the case of p64, this additional N-terminal domain arises due to alternative splicing of the CLIC5 gene [85]. The CLIC module is homologous to the glutathione-S-transferase (GST) superfamily, and structural studies have confirmed CLICs adopt the canonical GST fold [86, 87].

While other chloride ion channel families such as the CLCs have a well studied and typical integral membrane bound state and associated topology models within the

membrane [88], the CLICs are unusual as they exist primarily in a soluble state localised to the cytoplasm or nucleoplasm. The soluble structure of CLICs does not resemble the typical structure of other membrane proteins [87]. Despite this they have an apparent ability to insert directly into membranes, although the process that governs this insertion is still unknown and will likely require oligomerisation and a structural rearrangement in the monomeric subunit. By comparison, the structurally similar members of the GST protein family do not normally possess the ability to auto-insert into the membrane [89].

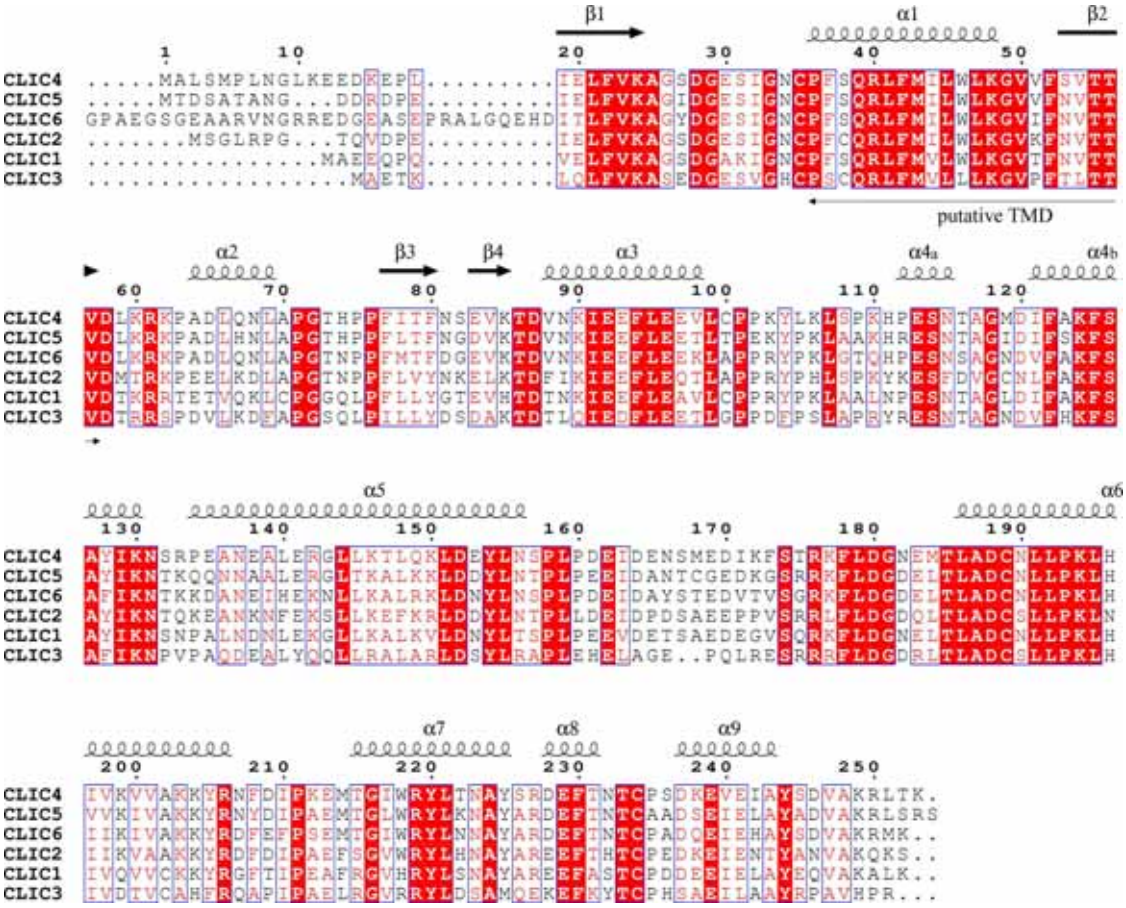

Figure 1.9: Multiple sequence alignment of the six human CLIC proteins.

Sequence numbers and secondary structural elements correspond to CLIC4. The putative transmembrane domain (TMD) is marked from residue 36-57. Multiple sequence alignment performed using CLUSTALW [48] and presented using ESPript [49].

1.2.2 CLIC1

CLIC1 is perhaps the most extensively studied member of the CLIC proteins, and was the first member to be discovered in humans [90]. Consistent with other CLICs, CLIC1 is widely expressed within the body and the CLIC1 transcript has been identified in nearly all human tissues studied [90-92], with the highest levels of expression reported in muscle, liver and kidney. CLIC1 appears to be important within the cell and was identified by mass spectrometry in a study identifying differentially regulated proteins in colorectal cancer cells [93]. In this study CLIC1 was found to be upregulated across a variety of tumour tissue regardless of tumour grade, pathological tumour-node-metastasis stage or localisation.

The subcellular localisation of CLIC1 varies across different tissue types as reported in the literature. CLIC1 transfected CHO-K1 cells show primary localisation to the nuclear membrane and nucleoplasm, and hence was first named Nuclear Chloride Channel 27 kDa (NCC27), where 27 kDa refers to the apparent molecular weight of CLIC1 [90]. Other studies have supported the identification of prominent CLIC1 nuclear localisation [91]. CLIC1 has also been consistently observed to localise on the apical membrane [91, 94]. In sperm cells, CLIC1 localises to the acrosomal region of the sperm head, and it has been suggested CLIC proteins may play an important role in spermatozoa function [95].

A recent and comprehensive study of CLIC1 distribution in mice found CLIC1 is expressed in a number of unilayered epithelial cells such as the stomach and small intestine, as well as a diffuse distribution in skeletal muscle [94]. The same study examined the distribution of CLIC1 in different cell lines, finding CLIC1 is present on the plasma membrane in Panc1 cells and in an intracellular compartment below the apical plasma membrane in T84 cells. The high variability of CLIC1 localisation could suggest it has different cellular roles within different cell types.

1.2.2.1 CLIC1 physiological function

The biological function of CLIC1 is still unknown and its apparent capacity to fulfil various cellular roles makes it harder to definitively establish. It is now accepted that CLIC1 has the ability to form functioning anion channels despite contradictory results that show the soluble monomeric structural form does not resemble a typical membrane protein and the majority of CLIC1 exists primarily in the soluble form. There is a single hydrophobic stretch in the N-terminal domain of CLIC1 which has been designated the putative transmembrane domain (TMD), which encompasses α-helix 1 and β-strand 2 in the region Pro25 to Val46. Evidence from patch clamp single-channel recordings on the plasma membrane of CHO-K1 cells using N- and C-terminus directed CLIC1 antibodies to inhibit channel activity, suggests that the integral membrane bound form of CLIC1 has an external (*trans*) N-terminus and an internal (*cis*) C-terminus [96]. Furthermore, experiments show that CLIC1 channels are redox sensitive from the *trans* side of the membrane, which is likely due to Cys24 [97]. This result places a lower limit on the putative TMD whereby at least the first 24 residues are externally located in the CLIC1 integral membrane form. While the soluble structure of CLIC1 has been determined [87], the structure of CLIC1 in an integral membrane form remains elusive.

During initial characterisation of CLIC1, electrophysiology experiments showed it has the ability to form anion channels in CHO-K1 cells with a single channel conductance of 22 ± 5 pS [90]. These results were confirmed by experiments measuring anion channel activity on both the plasma membrane (8.5 ± 0.6 pS) and the nuclear membrane (16.2 ± 1.4 pS) [96]. Furthermore, the addition of an epitope tag (FLAG) to CLIC1 in antibody inhibition experiments suggest CLIC1 is directly involved in the anion channel activity [96]. Further electrophysiology studies in CHO-K1 cells suggested CLIC1 channels vary according to different stages of the cell cycle [98].

It became increasingly important to identify whether or not CLIC1 is capable of solely creating these anion channels or if it required partners meaning it could be a channel regulator. This was elucidated when CLIC1 channels in planar lipid bilayers

were measured using tip-dip electrophysiology, which measured a conductance of 31 ± 1.8 pS [99], comparable to tip-dip single channel recordings in transfected CHO-K1 cells of 31.2 ± 1.5 pS [87]. Both *in vivo* and *in vitro* channels were blocked by the chloride channel blocker IAA-94 [99]. Experiments by other groups support these results that suggest purified CLIC1 is capable of creating ion channels on its own, and that both *in vivo* and *in vitro* CLIC1 channels can be measured with similar properties [100].

It was suggested by Ashley and co-workers in 2003 [101], that in order to categorise a given protein as an ion channel it must satisfy a number of conditions including the ability to form ion channels *in vivo* and identical channels *in vitro*. In addition, the protein must conform to a structure sufficient to cross membranes and it must show altered function when critical pore forming regions are modified. The first two tests have now been well studied for CLIC channels, while importantly the structure of the integral membrane bound form is unknown. There has been some progress however in showing altered channel function. Littler *et al.* 2004 [102] described the structure of the redox controlled non-covalent CLIC1 dimer, and it was shown that critical cysteine residues, Cys24 and Cys59, are necessary for channel activity in artificial lipid bilayers. Not only was the oxidised CLIC1 dimer capable of forming ion channels, it was found to have increased activity compared to reduced monomeric CLIC1. These findings support a redox controlled CLIC1 channel mechanism. Although, a recent conflicting report as to the exact redox regulation mechanism shows that a reduced environment can actually increase channel activity [97].

Recent work has shown that CLIC1 channels may be concurrently regulated by cAMP and CFTRs, resulting in increased channel activity and membrane association [103]. CLIC1 (and CLIC5) channels also appear to be regulated by the presence of F-actin on the cytosolic (*cis*) side of the plasma membrane in planar lipid bilayers, with cytochalasin B disassembly of actin reversing the inhibition [104].

CLIC1 has also been linked to β-amyloid induced microglial activation. Stimulation of neonatal rat microglia leads to an increase in total CLIC1 and

furthermore to measurable chloride conductance in resting cells [105]. It has been specifically shown that CLIC1 translocates from the cytoplasm to the plasma membrane in response to β-amyloid treatment in microglia cells [106]. By blocking these channels with the IAA-94 inhibitor or suppressing CLIC1 expression by RNA interference, events involved in β-amyloid induced microglial activation (such as neuronal apoptosis or the release of TNF-α) are affected [105].

Further experiments have additionally shown that blocking the β-amyloid induced CLIC1 channels affects the production of neurotoxic reactive oxygen species [106]. Since CLIC1 activation is also dependent on the creation of reactive oxygen species, a forward feeding mechanism of CLIC1 induced reactive oxygen species has been proposed [106]. CLIC1 has also been implicated in the cells response to β-amyloid stimulated mononuclear phagocytosis. Using the CLIC1 channel blocker IAA-94, an increase in phagocytosis was noticed [107]. It was also observed that the β-amyloid stimulation of inducible nitric oxide synthase is prevented by channel blocking.

1.2.2.2 CLIC1 structure

The first crystal structure of a CLIC family member was solved within our lab in 2001 and confirmed the suggestion [89] that there is a structural relationship between the glutathione-S-transferase (GST) superfamily and CLICs [87]. A 1.4 Å model of human CLIC1 was shown to adopt the canonical GST fold, with close structural similarity to the omega class GSTs (16% identity, C_α RMSD 2.08 Å) [108]. The CLIC1 construct used during crystallization contains two point mutations: E151G that was inadvertently generated through PCR infidelity and Q63E that is the result of a sequencing error in the discovery of CLIC1 (refer to NCBI reference sequence NP001279.1 and the updated version NP001279.2 [109]). In the structure, CLIC1 consists of two distinct domains joined by a short interdomain loop which are arranged into a flat shape with approximate dimensions of 50 Å x 50 Å x 20 Å (see Figure 1.10).

The N-terminal domain has a thioredoxin-like fold that is made up of a mixed β-sheet of four strands, and includes three α-helices. Near the N-terminus of β-strand 3

is an invariant *cis*-proline (*cis*-Pro65 in CLIC1) that is conserved in all thioredoxin folds and therefore is likely to be structurally critical. The C-terminal domain is all α-helical and contains an acidic loop region known as the "foot" loop. This is a mostly unconserved region of the molecule that is not present in the related GSTs. The loop is highly mobile and would likely appear disordered in the crystal structure if not for packing interactions, as is the case for CLIC4 (see Figure 1.10).

There is evidence from the structure that the proline rich interdomain loop and the N-terminal helices 1 and 3 exist as two different conformers. This results in an angular shift of helices 1 and 3 of 4.4° and 1.9°, respectively, between what are termed the major and minor conformers [87]. Since these helices are at the interface between the N-terminal and C-terminal domains, this suggests CLIC1 possesses considerable plasticity, possibly facilitating the necessary structural rearrangement of the soluble monomeric structure into an integral membrane form.

A bioinformatics study searching for mutations in the CLIC gene within tumour cells showed that there is occasionally a mutation corresponding to Leu168 → Ser168 in CLIC1 [110]. Leu168 is part of a short motif, ^{167}FLDG^{170} (178-181 in CLIC4), conserved across CLICs 1-6, which is located in the extended loop region between helix 5 and helix 6 near the acidic foot loop. This region also corresponds to the beginning of what is sometimes referred to as the GST-motif II, a sequence that encompasses helix 6 and the preceding loop [111]. Leu168 is completely buried in a hydrophobic pocket formed by helix 4a, 5, 6 and 7. The mutation to serine, a polar residue, could possibly have a global effect on the CLIC1 C-terminal domain structure by destabilising a critical hydrophobic core.

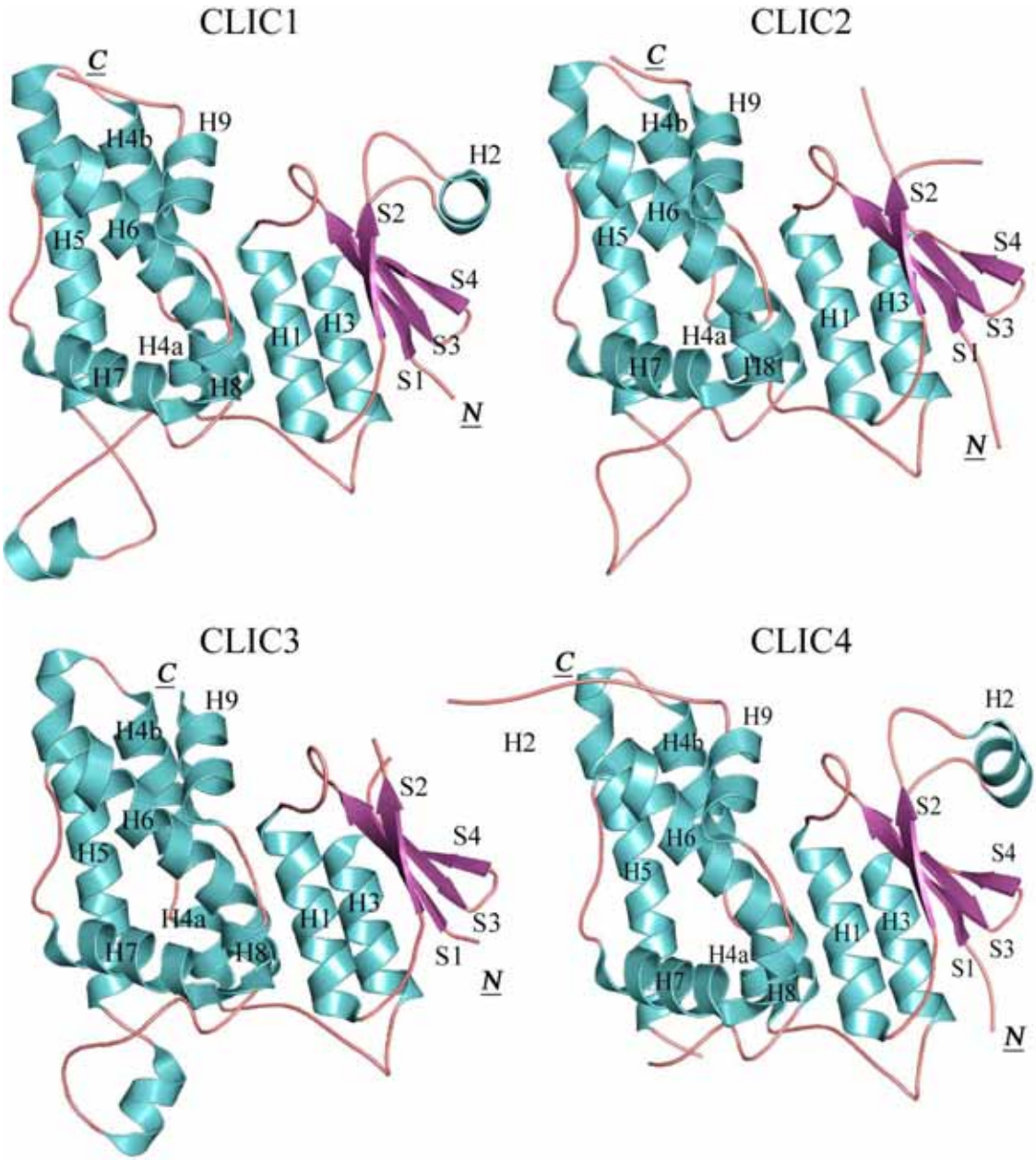

Figure 1.10: Crystal structures of human CLIC family members.

Secondary structural elements are coloured and labelled according to type: α-helix (cyan, H1-H9), β-strand (purple, S1-S4) and random coil (pink). The N-terminus and C-terminus are also labelled. Structures shown include CLIC1 (PDB:1K0M), CLIC2 (PDB:2R4V), CLIC3 (PDB:3FY7) and CLIC4 (PDB:2AHE).

A further 1.8 Å resolution structure of CLIC1 (PDB: 1K0N) was also determined and contains a single covalently bound GSH molecule [87]. In CLIC1, the N-terminus of helix 1 contains a conserved glutaredoxin-like motif (CxxC), ^{24}CPFS^{27}, which forms the covalent interaction with GSH through a disulfide bond between Cys24 and the GSH thiol group. The backbone of the GSH tripeptide lies in a shallow groove formed by the loop connecting helix 2 and β-strand 3, as well as the side chain of Phe26. The

GSH binding site in CLIC1 is partially conserved across mammalian CLICs (see Figure 1.11B) and it overlaps with the G-site in GST, which also binds GSH.

The CLIC2 structure has been solved at 1.85 Å in complex with two GSH molecules. However, GSH binds in a completely independent manner to that found in CLIC1 or the GSTs [112] (see Figure 1.11). In the case of CLIC2, the GSH molecules do not covalently bind to CLIC2, but form an interaction in hydrophobic grooves on the protein surface. The first GSH molecule mostly interacts with the CLIC2 N-terminal domain near helix 1 and β-strand 2 where the amino group of the γ-glutamyl moiety forms a hydrogen bond with Glu25 (Figure 1.11A). The second GSH molecule is wedged between two CLIC2 crystallographic symmetry mates in a hydrophobic pocket (Figure 1.11B). While the CLIC2 binding surface is partially conserved in the first GSH molecule interaction, the second GSH molecule binds in a non-conserved region. It was suggested that the failure of GSH to bind CLIC2 in a manner similar to CLIC1 could be due to local interference in the binding site posed by Phe83 [112]. It is noted that in both CLIC1 and CLIC2, the GSH molecule is modelled with significantly higher B-factors than CLIC residues in its local environment.

Currently there has been no further work on the significance of CLIC:GSH interaction and no further insight as to how this might relate to function. It could be that GSH binding is not functionally significant for CLIC proteins as it is for GSTs. The binding of glutathione to GST proteins is well studied and supports its functional role where it is known to catalyse the conjugation of electrophilic substrates to glutathione [113]. However, the important contacts between most GSTs and GSH come from the partner monomer in a GST dimer. The CLICs do not share this GST specific dimerisation and are therefore more closely related to the monomeric omega class of GSTs, which also lack a high affinity for GSH and other substrates [108].

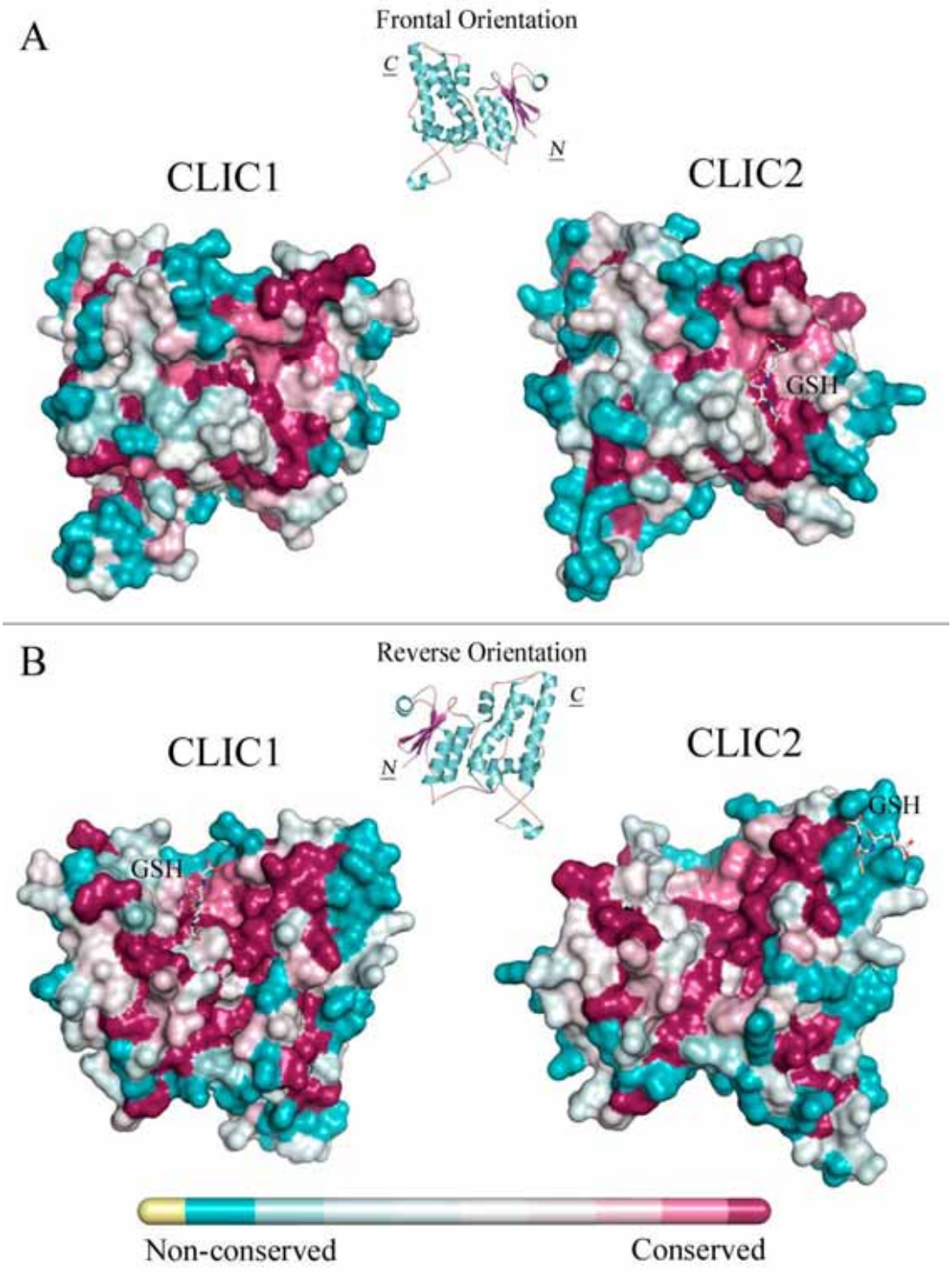

Figure 1.11: The molecular surface of CLIC1 and CLIC2 coloured by conservation score.

The colour bar is shown at the bottom varying from non-conserved (light yellow) to conserved (purple). Sequence alignment was performed using CLUSTALW [48] over 23 CLIC sequences in 8 species. Conservation scores were calculated using Bayesian methods in Consurf [52].

A) The structures of CLIC1 (PDB: 1K0M) and CLIC2 (PDB:2R4V) are shown in surface representation. A GSH molecule (stick representation) bound to CLIC2 is visible. The cartoon orientation matches the surface orientations of the proteins.
B) CLIC1 and CLIC2 are shown from the reverse angle, with one bound GSH molecule in CLIC1 and one in CLIC2 visible.

Plasticity in the CLIC1 module was further explored by Littler *et al*. through the X-ray crystallographic structure of the non-covalent CLIC1 dimer [102]. It was found that under oxidation, CLIC1 underwent a reversible structural transition that involved a major rearrangement of its N-terminal domain (see Figure 1.12). This structural change forms a large hydrophobic N-terminal domain surface that facilitates the interface of the dimerism state. Importantly, previously separated cysteine residues 24 and 59 are brought close enough to form an intramolecular disulfide bond.

The regions of the protein that enabled the restructuring of the protein were analysed by Ramachandran plots which show the most important residues involved in the transition are located near the GSH binding site. Although there is minor movement in the C-terminal domain, there are no significant changes observed near a putative CLIC NLS site at the C-terminus of helix 6.

An unfolding equilibrium experiment performed in a reduced environment and subjecting CLIC1 to acidic conditions, found there is a highly populated intermediate structural state [114]. The sub population contains a modified N-terminal domain which may constitute a newly discovered structural transition to further support the notion of plasticity in CLIC1. However although the conditions were redox controlled, it cannot be conclusively ruled out that the structural change is similar to that seen in the CLIC1 dimer transition.

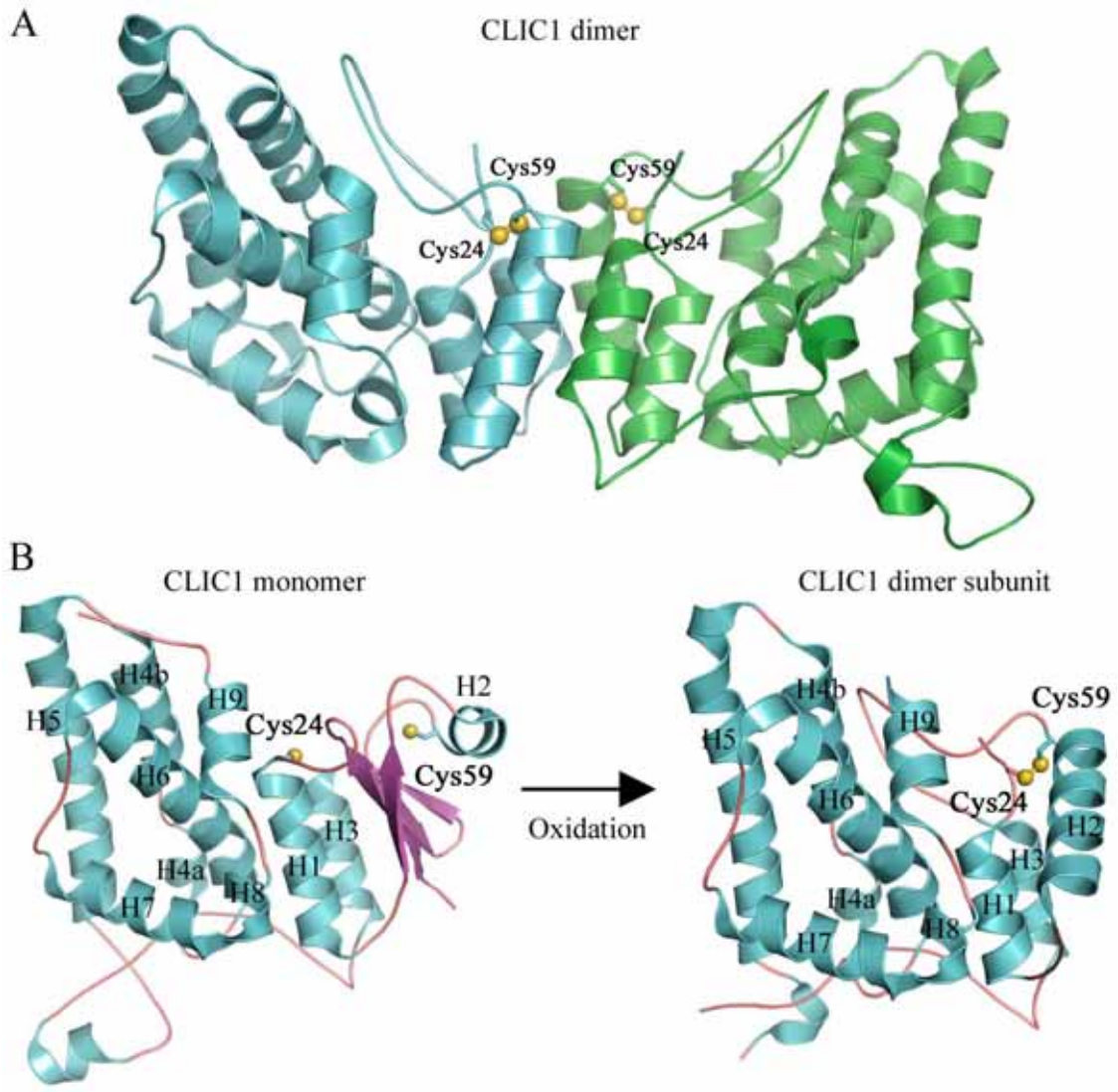

Figure 1.12: The oxidised CLIC1 dimer.

A) The all α-helical CLIC1 dimer showing the locations of critical cysteine residues (Cys24 and Cys59) that form intramolecular bonds (PDB:1RK4). The two subunits are shown in cyan and green.

B) The CLIC1 monomer is shown coloured by secondary structure elements, with the location of Cys24 and Cys59 labelled. CLIC1 undergoes a redox controlled structural rearrangement which facilitates the formation of a disulfide bond between Cys24 and Cys59, which is shown in the CLIC1 dimer subunit.

1.2.3 CLIC4

CLIC4 is another member of the CLIC family that is ubiquitously expressed in different cell and tissue types. It was the first homologue of p64 to be identified [115], consisting of a 253 amino acid sequence, with a predicted molecular weight of 28.6 kDa [116]. Northern blot analysis showed the human CLIC4 transcript is present across most if not all tissue types examined with high levels in the heart, placenta and

skeletal muscle [117]. In mouse tissues the highest levels of CLIC4 mRNA was found in heart, lung, kidney, liver and skin [118]. Immunoblotting experiments showed CLIC4 is present in rat brain, kidney, liver and lung [115].

During identification, the CLIC4 gene was seen to localise to the cytoplasm and mitochondria of both mouse keratinocytes and rat hepatocytes [118]. Because of this, it is sometimes referred to as mitochondrial CLIC or mtCLIC, however the intracellular distribution varies. In HEK-293 and HT-4 cells, CLIC4 was immunolocalised to the endoplasmic reticulum [116]. In rat hippocampal neuron cells, CLIC4 associates with large dense core vesicles [119]. In Panc1 cells, it was found that human CLIC4 is expressed in an intracellular vesicular pattern, while in human kidney cells it expresses in a diffuse pattern near the apical domain of proximal tubule cells [117]. CLIC4 has also been found to localise to the plasma membrane in HEK-293 cells [120]. The finding that CLIC4 localises to the midbody and centrosome of dividing cells has suggested it may have a role in the cytoskeletal processes [121]. It is also seen that CLIC4 localises primarily to the nucleus in human and mouse epidermal skin cells [122].

Recent studies have shown that the subcellular localisation of CLIC4 shifts to the nucleus in response to cell stress from a predominantly cytoplasmic distribution in quiescent mouse keratinocytes [123].

1.2.3.1 CLIC4 physiological function

There has been little support for a well defined CLIC4 anion channel, with independently run experiments reporting variations in conductance and sensitivity measurements. However like CLIC1, CLIC4 is becoming increasingly accepted as an ion channel protein.

An initial experiment incorporated HEK-293 ER vesicles, containing CLIC4, into planar lipid bilayers which measured poorly selective channels between 10-50 pS, with no apparent IAA-94 sensitivity [116]. Channel activity on the plasma membrane has been measured using the patch clamp technique, which found CLIC4 channel

conductance to be lower than that measured in the earlier experiment, in the order of 1 pS [120]. This study found CLIC4 channels are sensitive to IAA-94 at a relatively high (100 μM) concentration and that cytoplasmic antibodies raised against the C-terminal domain of CLIC4 are able to block channel activity. Although CLIC4 may be involved in cytoskeletal processes, CLIC4 channels are not affected by *trans* F-actin in contrast to the inhibitory effect it has on CLIC1 and CLIC5 channel activity [104].

Investigations using tip-dip physiology suggest a single channel conductance of 30 pS [86]. The same study found CLIC4 channels are sensitive to *trans* redox conditions with no channel activity observed under reduced conditions (5 mM DTT), as well as enhanced membrane association under oxidative conditions (0.4 mM H_2O_2) as measured by surface plasmon resonance. In an attempt to clarify the location of the putative TMD, protease digestion, using proteinase K on microsomes containing CLIC4, showed there is a large C-terminal cytoplasmic (*cis*) domain (~23 kDa) and a smaller N-terminal luminal (*trans*) domain (~6 kDa, including the integral membrane domain). This is consistent with the integral membrane topology for CLIC1 which defines a putative TMD of ~20 residues, corresponding to Pro36-Val57 in CLIC4 [116].

In a recent paper by Singh and Ashley, 2007 [124], the authors have sought to resolve discrepancies in previous channel conductance measurements with a more systematic electrophysiology study. CLIC4 was incorporated into planar lipid bilayers and single channel recordings were measured using the patch clamp technique. Their findings support earlier results that show CLIC4 ion channels are poorly selective for anions [116], a non-specificity that is also observed in CLIC5 [104] and which casts doubt over the suitability of the CLIC nomenclature. Contrary to previous findings, it was also shown that CLIC4 channel activity can be measured when both the *trans* and *cis* chambers are in a reduced environment (1 mM DTT, 10.3 ± 1.0 pS). In addition, channel activity was measured when both chambers are in an oxidised environment maintained artificially (0.1 mM H_2O_2, 8.9 ± 1.1 pS) or close to physiological conditions (5 mM GSH, 14.8 ± 1.6 pS). It was reported that CLIC4 channel activity is still observed under reduced conditions, with a channel conductance similar to that obtained under oxidised conditions (~15 pS) [124]. Consistent with CLIC1 redox

sensitivity [97], the *trans*, not *cis*, redox potential was found to modify the behaviour of CLIC4 channels by decreasing channel activity after the addition of an oxidising agent (GSSG) to the *trans* chamber. However these results appear to further contradict previous redox sensitivity measurements that suggest an oxidised environment increases activity, while a reduced environment decreases activity [86]. A CLIC4 C-terminal deletion construct, consisting of the first 61 residues containing the putative TMD, is also capable of forming channels albeit at lower conductance compared to full length CLIC4 [124]. It was found that the deletion construct created channels that maintain redox sensitivity and are blocked by a *trans* oxidative environment controlled by 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB), similar to full length CLIC4 [124].

The often contradictory anion channel results obtained thus far lead to a somewhat confusing and vague understanding of the physiological role of CLIC4. Other possible roles for CLIC4 have been studied including an involvement in cellular apoptosis. CLIC4 mRNA is upregulated in p53 +/+ mouse keratinocytes, as well as keratinocytes treated with the exogenous tumour necrosis factor-α (TNF-α) [118]. Lowering CLIC4 levels with antisense CLIC4 resulted in suppression of p53 mediated apoptosis while the apoptic pathway governed by Bax was not affected [125]. In addition to p53 induced upregulation, CLIC4 is also upregulated in response to c-Myc mediated apoptosis and suppressing the upregulation using RNA interference inhibits Myc induced apoptosis [126]. CLIC4 is upregulated during the conversion of fibroblasts to myofibroblasts by the transforming growth factor-β1 (TGF-β1), a protein that is part of a major growth regulatory pathway [127]. This upregulation in response to TGF-β1 was found to be specific to CLIC4 of the CLIC family members.

In an important experiment highlighting the role of CLIC4 in nuclear translocation, a focus of this thesis, it was shown that CLIC4 translocates to the nucleus in human osteosarcoma cells as well as mouse S1 keratinocytes, in response to endogenous and exogenous stress inducers [123]. After translocation, CLIC4 was found near the nuclear envelope and in the nucleoplasm. Directing CLIC4 to organelle compartments using an adenoviral vector was found to accelerate apoptosis. The CLIC4 mediated apoptosis pathway is independent of the apoptotic protease activating factor (Apaf) apoptosis pathway [123]. CLIC4 was also found to translocate to the

nucleus in keratinocytes undergoing growth arrest by differentiation, senescence or TGF-β [122]. Additionally, CLIC4 expression increases in human and mouse keratinocyte differentiating cells induced by Ca^{2+}, serum and 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA) [122].

The regulation of CLIC4 by factors involved in cancer pathogenesis suggests that it may be involved in tumour suppression and presents itself as a possible target for cancer therapy [128]. Using cDNA arrays it was found CLIC4 expression is reduced in renal, ovarian and breast cancers and across matched tissue types CLIC4 tends to localise to the nucleus in normal human epithelial cells but not in cancerous cells [110]. In contrast CLIC4 is up-regulated in tumour stromal cells. With an increasing tumour grade, total CLIC4 is reduced for most tissue types studied while the opposite effect is noticed in tumour stroma where total CLIC4 increases with tumour grade. This compares with an observed upregulation of CLIC1 in all colon tumour tissue types studied [93].

Interestingly the restoration of CLIC4 levels in established xenograft tumours in mice was able to retard tumour growth, with nuclear targeted CLIC4 being more effective than cytoplasmic CLIC4 [110]. In contrast the overexpression of CLIC4 in tumour stroma exacerbated tumour growth.

CLIC4 also has also been implicated in angiogenesis where it is required for tube formation. An antisense CLIC4 construct transfected into human telomerase-immortalized dermal microvascular endothelial cells (TIME), as well as RNA interference were used to suppress CLIC4 expression which resulted in severe disruption of tubular morphogenesis [129]. Although a role in tube formation for the invertebrate CLIC homologue EXC-4 had already been established [130], this was the first time a human CLIC member had also been shown to function in this role.

Further studies have supported a significant role for CLIC4 in angiogenesis. An *in vitro* experiment utilising human endothelial cells showed that reduced CLIC4 expression by RNA interference inhibits cell proliferation and tubular network

formation [131]. Increased CLIC4 expression was found to promote proliferation and network formation.

A recent study has made use of a CLIC4 knockout mouse and observed *in vivo* angiogenesis defects [132]. It was also found that the CLIC4 (-/-) mice had impaired vacuolar acidification along the endothelial tubular morphogenesis pathway. It is worth noting that the CLIC4 knockout mouse is otherwise healthy.

1.2.3.2 CLIC4 structure

The X-ray crystallographic structure of monomeric CLIC4 was solved by our group at a resolution of 1.8 Å [86] (Figure 1.10). The clone used for crystallization trials had a 14 residue C-terminal extension due to an incorrect antisense primer, and this construct which proved suitable for structure determination, was called CLIC4(extension) or simply CLIC4(ext). Subsequently, a structure of wild type CLIC4 has been solved at 2.2 Å by an independent group [133]. While the overall structure is virtually identical to that of CLIC4(ext), the wild type asymmetric unit contains three copies of the CLIC4 monomer (Figure 1.13). Although the trimeric complex was reported as a unique oligomeric species, it is likely the complex is formed mainly due to crystal packing interactions and therefore has no relevant biological significance. There is also no evidence for a trimeric CLIC4 species reported elsewhere in the literature and on size exclusion chromatography it runs as a monomeric species even after treatment with hydrogen peroxide, in contrast to CLIC1 [86].

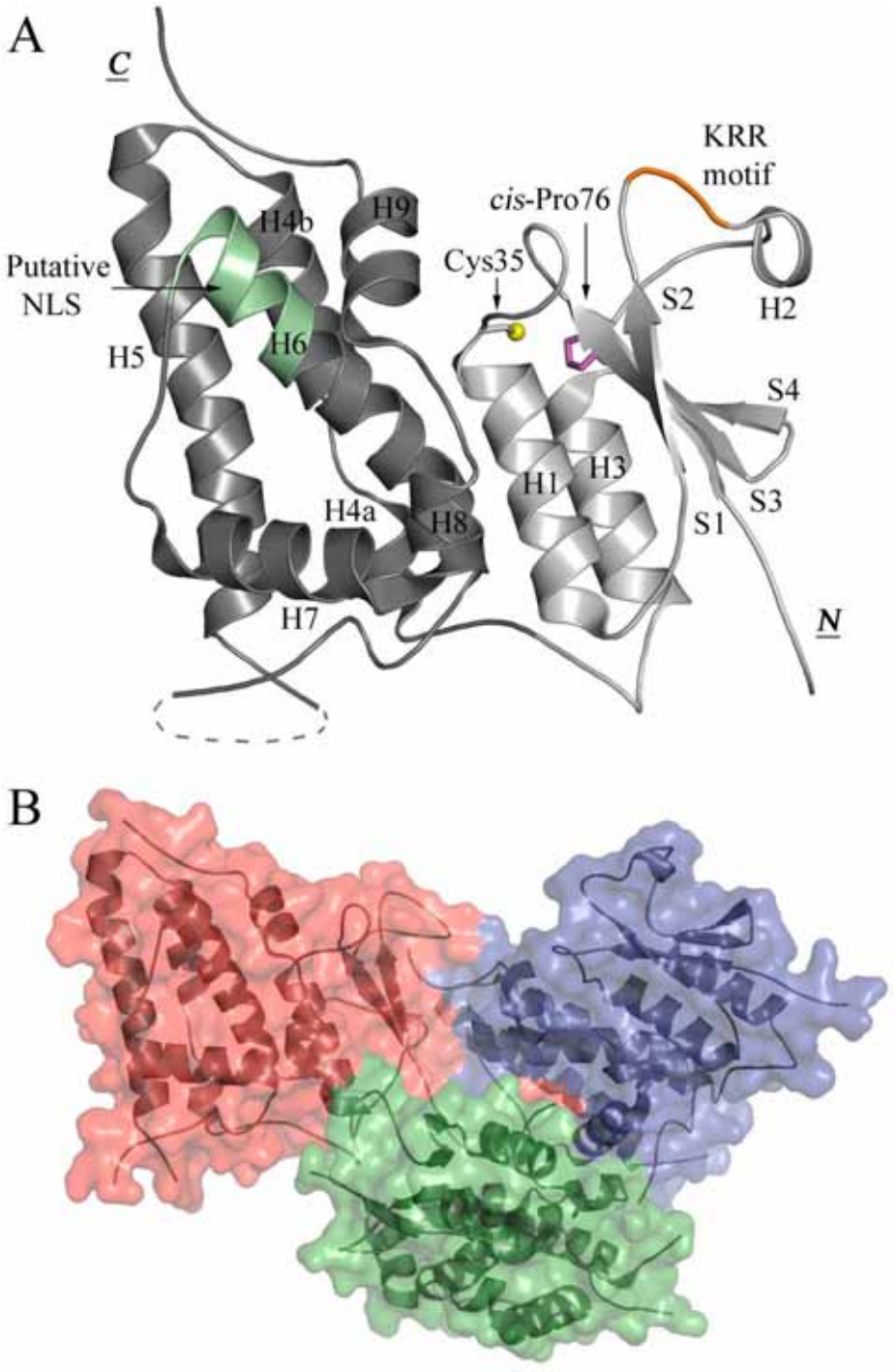

Figure 1.13: The CLIC4 crystal structure presented with three molecules in the asymmetric unit.

A) A cartoon representation of subunit C in the CLIC4 asymmetric unit (PDB:2D2Z). The N-terminal thioredoxin-like domain is shown in light grey, the all α-helical C-terminal domain is shown in dark grey. The disordered foot loop is depicted by a dashed line. The putative NLS (KVVAKKYR) and KRR motif are shown in green and orange, respectively. The active cysteine, Cys35, is shown in ball and stick (sulfur, yellow sphere) and the conserved *cis*-Pro76 is shown in magenta.

B) The transparent surface of the three CLIC4 monomers (PDB:2D2Z) in the asymmetric unit are coloured in green (subunit A), red (subunit B) and blue (subunit C), with a cartoon representation of each subunit shown in grey.

The structure of monomeric CLIC4 closely resembles that of monomeric CLIC1 [87], displaying the canonical GST fold consisting of two well defined and separate structural domains: the N-terminal thioredoxin fold and the C-terminal all α-helical domain (Figure 1.13). The invariant *cis*-proline at the N-terminus of β-strand 3 corresponds to *cis*-Pro76 in CLIC4 and the glutaredoxin-like motif (CxxC) corresponds to ^{35}CPFS^{38}, where Cys35 is an active cysteine [124]. The flexible foot loop region in the C-terminal domain is disordered in both of the published CLIC4 crystal structures and thus residues ~160-173 are not included in the final models. The main chain atoms of CLIC1 and CLIC4 are in very similar positions with an RMSD of 0.56 Å over 209 C_α atoms corresponding to the high sequence identity between CLIC1 and CLIC4 of 67%. Notable differences only occur near flexible regions or in the connecting loops between helix 2, β-strand 3 and β-strand 4 in the N-terminal domain [86]. CLIC4 does not contain a cysteine residue near helix 2 which is equivalent to Cys59 in CLIC1 and therefore does not form an oxidised dimer similar to that observed for CLIC1 [86].

There are two clusters of basic residues in CLIC4 corresponding to the short motif ^{60}KRR^{62} near helix 2 and the putative NLS sequence, ^{199}KVVAKKYR^{206} [90, 123], which is located at the C-terminus of helix 6 (Figure 1.13). Both of these regions are structurally identical in CLIC1. In the CLIC4 putative NLS, key basic residues Lys199, Lys203 and Lys204 line the solvent exposed face of helix 6. The last basic residue in the NLS motif, Arg206, makes contacts with the main chain of the C-terminal tail and is present at the start of the loop that connects helix 6 to helix 7. The hydrophobic residues, valine and alanine, are buried while Tyr205 makes a hydrogen bond with the carbonyl group of Glu241 in helix 9.

1.2.4 Invertebrate CLIC proteins

The CLIC family of proteins is highly conserved across vertebrate species, with at least six CLIC orthologues present in most species. However CLIC-like proteins are not restricted to the vertebrate chordate subphylum. Proteins that share a low sequence similarity with CLICs have been found in many invertebrate species, including nearly all arthropods and nematodes. Phylogenetic analysis indicates that GST proteins

diverged independently from CLIC proteins, and that the vertebrate and invertebrate subfamilies of CLIC proteins also diverged independently [134].

The nematode *Caenorhabditis elegans* is known to contain two CLIC homologues, EXC-4 and EXL-1 (EXC-4-like protein 1), while the arthropod *Drosophila melanogaster* contains just one (referred to as *Dm*CLIC). Human CLIC4 shares a sequence identity of 30% with *Dm*CLIC, 25% with EXC-4 and 24% with EXL-1, while *Dm*CLIC shares a sequence identity of 39% with EXC-4 and 35% with EXL-1. EXC-4 and EXL-1 share a sequence identity of 28% (see Figure 1.14). While *Dm*CLIC remains relatively unstudied, the two nematode CLICs have been functionally characterised. Berry *et al*. 2003 undertook an extensive study of EXC-4 and were the first to functionally characterise an invertebrate CLIC [130].

EXC-4 expression is seen in a variety of cells including the excretory system, tubular rectal epithelia and tubular neuronal support cells [130]. The subcellular distribution of EXC-4 is distinct, with localisation to the lumenal membrane of excretory, duct and pore cells, as well as other tubular membranes. It was found that only the first 55 residues of EXC-4, which corresponds to the putative TMD, are needed to maintain lumenal membrane localisation. A single point mutation to the conserved leucine residue within the putative TMD (L46P), located near the centre of helix 1, was found to be sufficient to disrupt the membrane localisation of EXC-4 [130]. In contrast, EXL-1 was found to localise mainly to lysosomal membranes and the Golgi apparatus [130], although it is also targeted to membranes with a minimal N-terminal domain consisting of residues 1-53 [134].

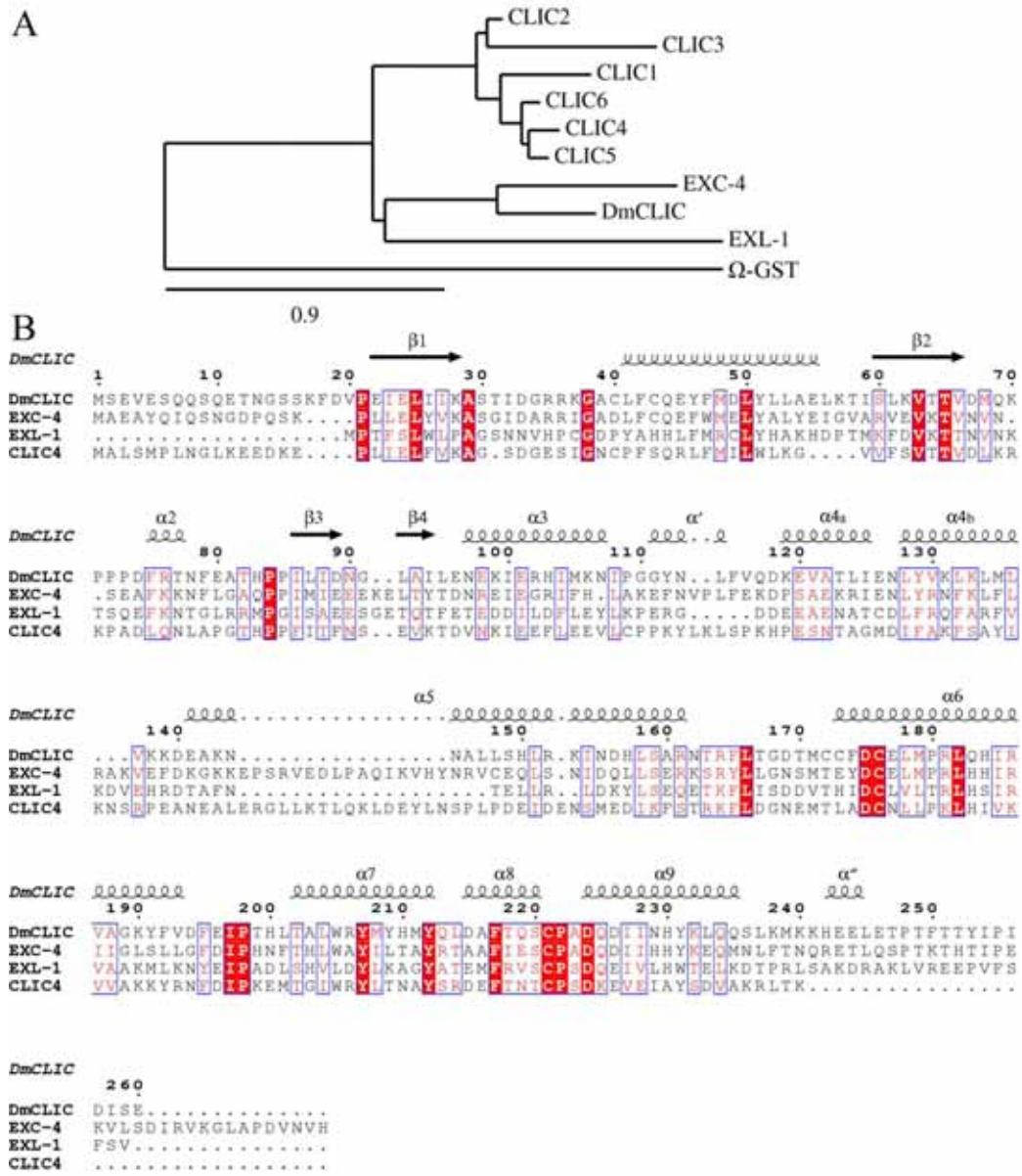

Figure 1.14: Phylogram and multiple sequence alignment of human CLIC family members and invertebrate CLIC homologues.

A) A phylogram of vertebrate CLICs (*h*CLICs 1-6), invertebrate CLICs (*Dm*CLIC, EXC-4 and EXL-1) and Ω-GST. Sequence alignment generated using CLUSTALW [48] and tree reconstruction performed using the maximum likelihood method implemented in PhyML [135].
B) Multiple sequence alignment performed using CLUSTALW [48] and presented using ESPript [49]. Sequence numbers and secondary structural elements correspond to *Dm*CLIC. Two α-helices that do not fit the standard labelling are labelled α′ and α″.

The first CLIC implicated in tubular morphogenesis was EXC-4 when mutations in the *exc-4* gene resulted in the formation of cysts in the *C. elegans* excretory canal [136]. The major tubular component of the *C. elegans* excretory canal system is a single excretory cell, and in the case of the null EXC-4 mutants the apical membrane of the canal no longer maintains a narrow tubule structure. Instead, what is seen is a

swelling of the excretory cell lumen which results in a phenotype typified by enlarged cysts in the *C. elegans* excretory canal (see Figure 1.15).

Subsequently the protein encoded by the *exc-4* locus was found to correspond to a CLIC orthologue, and it was confirmed that the phenotype mutant alleles that were previously studied contained molecular lesions [130]. It was found that the cystic phenotype was capable of being rescued with heat shock expression of wild type EXC-4. In addition, EXL-1 and *Dm*CLIC were capable of rescuing the cystic phenotype, suggesting a mostly conserved functional role between the invertebrate CLICs [130].

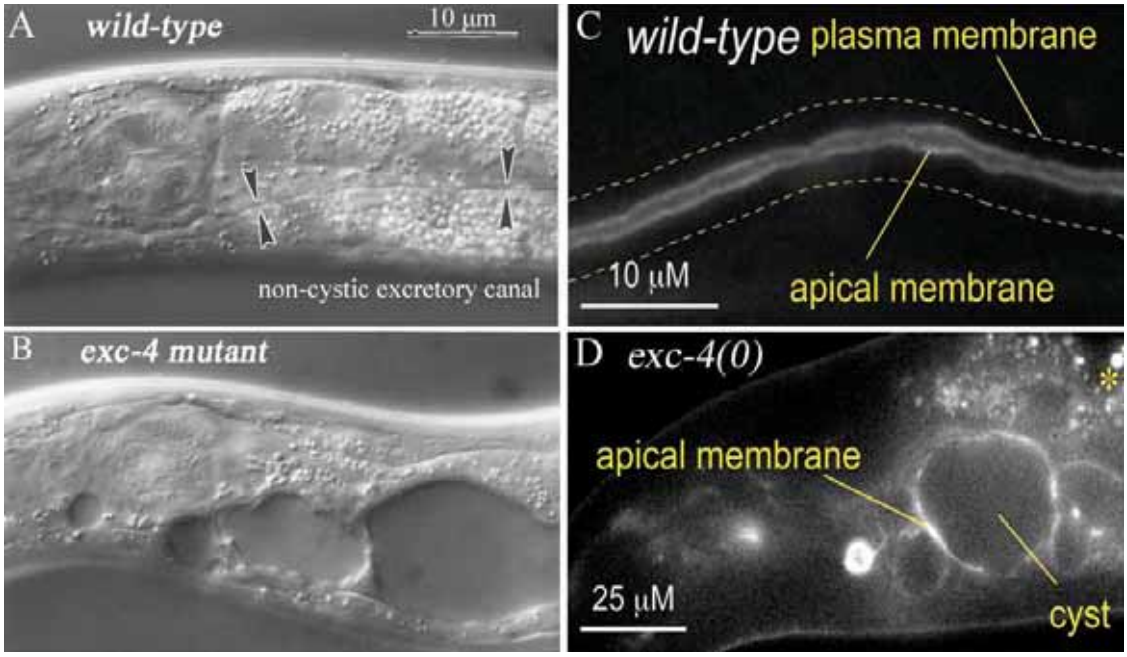

Figure 1.15: The *C. elegans exc-4* mutant phenotype.

Panels on the left show differential interference contrast micrographs from Buechner *et al.* [136]. Panels on the right show fluorescence micrographs from Berry *et al.* [130].

A) Wild type *C. elegans* showing the location of the non-cystic excretory canal.
B) The *C. elegans* exc-4 mutant displays a characteristic cystic phenotype.
C) Wild type *C. elegans* showing localisation of a non-functional exc-4:GFP reporter construct.
D) The exc-4 null mutant showing the localisation of a non-functional exc-4:GFP reporter construct.

Probing further into tubular morphogenesis as a general functional role of CLIC family members, it was discovered that only the first 66 amino acids of EXC-4 were necessary to facilitate rescue of the cystic phenotype when fused to the C-terminal domain of CLIC homologs [134]. The L46P point mutant was not able to rescue the cystic phenotype while single point mutants at residue 36 (D36N) and residue 39

(C39A), which correspond to the EXC-4 glutaredoxin-like motif ^{36}DxxC^{39}, still localised to the luminal membrane and maintained rescue activity. Although not capable of rescue activity on its own, a series of chimera constructs consisting of the EXC-4 putative TMD fused to the C-terminal domain of human CLIC1, as well as to distantly related proteins from *C. elegans* (Ω-GST and Σ-GST), were all capable of rescuing the null EXC-4 phenotype [134]. The results suggest function is highly conserved in the C-terminal domain across both vertebrate and invertebrate CLICs as well as GSTs.

To extend this analysis, two more chimera constructs were developed: the first consists of the EXC-4 helix 1 substituted into human CLIC1 (CLIC1[EXC4-h1]), and the second consists of the human CLIC1 helix 1 substituted into EXC-4 (EXC-4[CLIC1-h1]) [134]. It was found that neither the CLIC1[EXC4-h1] construct nor the EXC-4[CLIC1-h1] construct were able to rescue the null EXC-4 phenotype. The localisation of CLIC1[EXC4-h1] in the cytoplasm was unchanged compared to localisation of the CLIC1 wild type. The EXC-4[CLIC1-h1] construct also localised in the cytoplasm, compared with wild type EXC-4 which localised on the membrane. The results reveal that the EXC-4 putative TMD helix 1 is critical for function and targeting of EXC-4, but cannot confer these properties to the heterologous CLIC1 protein [134].

The invertebrate CLICs have also been briefly examined as to their ability to function as chloride ion channels like their distantly related vertebrate cousins. Like CLIC1 and CLIC4, EXC-4 and *Dm*CLIC have an affinity for lipid bilayers which is pH dependent, and in the case of *Dm*CLIC, oxidation dependent as well [137]. Using tip-dip physiology and chloride efflux assays, it was found that EXC-4 forms channels at low pH (4-5), while *Dm*CLIC forms channels at pH 6 with a conductance of approximately 30 ± 0.27 pS, which is consistent with the conductance of vertebrate CLIC channels [86, 87].

1.2.5 CLIC interaction partners

There have been many CLIC interaction partners identified consistent with the hypothesis that the CLICs fulfil various roles in the cell and may be implicated in signal transduction pathways. Some of the interactions have not been thoroughly characterised and thus may only represent an indirect association or non-specific binding. A subsection of CLIC interaction partners have been well studied and are more likely to directly bind CLICs. A summary of CLIC1 and CLIC4 interaction partners is shown in Table 1.2.

One of the first CLIC1 interaction partners identified is the scaffolding protein known as kinase A-anchoring protein 350 (AKAP350) [85]. It was found that a variety of rabbit CLIC homologues were capable of interacting with AKAP350 using a yeast two-hybrid system. CLIC5, p64, CLIC1, CLIC4 and CLIC6 were all found to interact with a segment of AKAP350 known as the pericentrin homology region (PHR) [138]. A 133 amino acid binding region in the PHR was found to interact with CLIC1 [85]. Other experiments showed that a p64 (CLIC5B) deletion construct needed only the last 120 residues to bind to AKAP350, while any further deletions of the region eliminated binding. However, when considering the veracity of these findings, it needs to be considered whether or not these intradomain deletion constructs are likely to be properly folded. The 120 residue binding region of CLIC5B corresponds to the C-terminal domain of the CLIC module truncated at the N-terminus of helix 5, which indicates the C-terminal domain helices 4a and 4b are not necessary for AKAP350 interaction.

A number of cytoskeletal proteins isolated from rat brain extracts were found to interact with CLIC4 lacking the putative N-terminal TMD in GST pull down assays [139]. These included 14-3-3ζ and 14-3-3ε, β-actin, creatine kinase, dynamin I and α-tubulin. However only 14-3-3ζ and dynamin I were found to bind directly through gel overlay and reverse pull down assays [139]. The CLIC4 construct used in the direct binding assays corresponds to residues 60-253, a truncation in the N-terminal domain of the CLIC module that excludes β-strands 1 and 2 as well as α-helix 1. However this corresponds to another intradomain deletion construct that may not be properly folded.

Another CLIC interaction partner is the protein phosphatase 1 isoform, PP1γ2, which is found primarily in the testis with a role in regulating sperm motility [95]. Pull down assays using GST fused CLIC1, CLIC4 and CLIC5, found all bound to PP1γ2 while none bound to 14-3-3ζ [95], in contrast to previous work [139].

CLIC1 was identified as a Like-sm 1 (LSM1) binding partner in a large scale yeast 2-hybrid screening experiment [140]. LSM proteins contain the structurally characteristic SM domain and LSM1 is known to form a cytoplasmic heteromer that includes LSM 1-7 and functions in mRNA decapping and degradation [141].

It has also been found that CLICs bind to sedlin, a relatively unknown protein [142]. Mutations to the sedlin gene are known to cause a condition known as spondyloepiphyseal dysplasia tarda, though how the process occurs is yet to be determined. A yeast homologue of sedlin is a subunit of the transport protein particle (TRAPP) complex which is involved in vesicle transport processes. Using a yeast two-hybrid system, the C-terminal region of CLIC1 corresponding to residues 138 to 201 was seen to be sufficient for Sedlin binding. This segment is a relatively small truncated sequence that cuts helix-5 at the centre and includes all of helix-6, and therefore may not resemble the corresponding region in the full length folded protein.

CLIC4 has recently been found to interact with the cytoplasmic C-terminus of the histamine H3 receptor (H3R) [143] in multiple *in vitro* binding assays (refer to Table 1.2). H3R is a G-protein coupled receptor (GPCR) which along with other histamine receptors regulates various functions in the central and peripheral nervous systems. H3R has been specifically shown to reduce the synaptic release of neurotransmitters including histamine, dopamine, γ-aminobutyric acid (GABA) as well as others [144].

The CLIC4 region spanning residues 121-253 appears to mediate the majority of the H3R interaction, although there is still weak binding to short C-terminal segments (CLIC4 residues 156-253, 185-253 and 156-223). In the residue segment 121-156 are helix 4b and helix 5, while in the segment 223-253 are helix 8 and helix 9 as well as part of helix 7. This could suggest the interaction between CLIC4 and H3R benefits

from a properly folded C-terminal domain, where helices 4, 5, 7, 8 and 9 are not explicitly required for direct binding.

CLIC3 was discovered due to an interaction with a member from the mitogen activated kinase family, ERK-7 [145], through yeast-2-hybrid screening. CLIC3 specifically interacts with the C-terminal tail of ERK-7 which is known to be important for function, activity and the targeting of ERK-7 to the nucleus [146].

A search for signal transduction partners for the dopamine receptors including D_2R, D_3R and D_4R found an interaction with rat CLIC6 through yeast 2-hybrid screening [147]. The dopamine receptors are integral membrane proteins with the cytoplasmic C-terminal tail able to bind CLIC6. A direct association between CLIC6 and D_3R was confirmed by using a GST pull down assay, where the CLIC6 construct corresponded to a C-terminal CLIC module segment beginning at helix 3 (residue 92 in human CLIC4) through to the C-terminus. It was observed that a 40 residue construct corresponding to helix 4b and helix 5 is not sufficient for binding D_3R [147].

CLIC5 (CLIC5A) was initially identified in a pull down assay as an interaction partner of a cytoskeletal complex including ezrin, actin, α-actinin, gelsolin and IQGAP1 from placental microvilli [92]. It was subsequently shown using the same pull down assay that a GST-CLIC5 fusion construct acting as bait was able to bind actin, α-actinin, ezrin, IQGAP1 and an unknown protein at 70 kDa (p70) [148]. It is likely that CLIC5 binds directly to ezrin and p70 since a drug capable of disrupting actin polymerization (LatB) had no effect on their interaction. CLIC5 may bind other skeletal proteins through one of these intermediate subunits. The binding sites participating in interaction between CLIC5 and ezrin have yet to be identified or characterised.

There have also been a number of studies showing that CLICs are acted on by various kinases as a possible post-translational mechanism of functional regulation. CLIC4 contains at least four protein kinase C (PKC) consensus sites as well as a single protein kinase A (PKA) consensus site. It was shown *in vitro* that CLIC4 is acted on by PKC at multiple sites increasing the apparent molecular weight by 10 kDa [116],

which corresponds to the molecular weight in native rat brain [115] supporting the possibility that this is functional phosphorylation activity that occurs *in vivo*. CLIC4 has also been reported to be tyrosine phosphorylated *in vivo* and phosphorylated *in vitro* by PKA and casein kinase II [101].

Membrane vesicles containing p64 were shown to produce anion channels which have increased activity under treatment of alkaline phosphatase [149]. This is in contrast to earlier studies that showed p64 associated channels are activated by phosphorylation [150]. It has also been shown that bovine and avian p64 interacts with the Src family tyrosine kinase $p59^{fyn}$ via SH2 and SH3 binding [151, 152]. This $p59^{fyn}$ mediated phosphorylation also increased channel activity and furthermore the effect could be cancelled by treatment with alkaline phosphatase.

Through site directed mutagenesis the Src kinase was found to phosphorylate the p64 residue Tyr33, which is located in the unique N-terminal extension sequence [151]. Since other CLICs do not possess an equivalent Tyr33 residue this particular interaction does not appear to be significant across the entire CLIC family.

Table 1.2: Summary of CLIC1 and CLIC4 interaction partners

CLIC	*Interaction Partner*	*Function of Interaction Partner*	*Technique*	*CLIC binding domain*	*Ref.*
CLIC1	AKAP350	Scaffolding	Yeast 2-hybrid	123-241[1]	[85]
	PP1γ2	Sperm motility	Pull down	-	[95]
	Sedlin	Vesicle transport	Yeast 2-hybrid	138-201	[142]
	LSM1	mRNA decapping and degradation	Yeast 2-hybrid	-	[140]
CLIC4	AKAP350	Scaffolding	Yeast 2-hybrid	134-253[1]	[85]
	PP1γ2	Sperm motility	Pull down Immunoprecipitation	-	[95]
	14-3-3ζ	Scaffolding	Pull down Immunoprecipitation	60-253	[139]
	Dynamin I	Endocytosis	Pull down	60-253	[139]
	Importin-α	Nuclear transport receptor	Immunoprecipitation	199-206[2]	[123]
	Histamine H3R	Mediate neurotransmitter release	Pull down Immunoprecipitation Immunofluorescence	121-253	[143]

[1] These interaction domains are inferred from sequence comparison with an experimentally determined interaction domain.
[2] This region corresponds to the putative CLIC4 NLS.

1.3 A putative CLIC4 NLS interaction with importin-α

The presence of a CLIC NLS was first suggested after amino acid sequence analysis found two NLS-like motifs in CLIC1 corresponding to ^{49}KRR^{51} and ^{192}KKYR^{195} [90]. The first sequence encompasses residues 49-51 in the CLIC1 N-terminal thioredoxin-like domain, which is part of the loop region between β-strand 2 and α-helix H2. The second NLS-like sequence encompasses residues 192-195 in the all helical C-terminal domain, which is located near the C-terminus of α-helix H6. The presence of consensus NLS sequences was examined due to the predominantly nuclear localisation of CLIC1 in all cell lines studied including CHO-K1 cells [90]. However the mechanism of CLIC1 translocation to the nucleus was not immediately explored further.

In a paper by Suh *et al*. 2004 [123], the possibility of a functional CLIC NLS was looked at again in conjunction with its relation to cellular apoptosis. In their work, Suh *et al*. studied the effect of multiple stress inducing agents on endogenous CLIC4 intracellular localisation in mouse keratinocytes. The stress inducers used included camptothecin, tumour necrosis factor-α (TNF-α), DNA-damaging agents (etoposide, adriamycin, mitomycin) and metabolic inhibitors (cycloheximide and actinomycin D). Results showed that in all these cases CLIC4 went from a predominantly cytoplasmic distribution to a nuclear distribution before the appearance of apoptic bodies [123]. Exogenous CLIC4 fusion constructs containing either a HA-tag or GFP fused to the N-terminus were also found to translocate to the nucleus in response to TNF-α or etoposide treatment. Furthermore, CLIC4 translocates to the nucleus in human osteosarcoma cells experiencing p53 mediated apoptosis, showing that both endogenous and exogenous apoptic pathways are capable of directing CLIC4 to the nucleus [123].

In both untreated and TNF-α treated keratinocytes the majority of CLIC4 is localised in the nucleoplasm (~75%) with a smaller amount located on the nuclear membrane (~25%) [123]. An interesting feature of the nuclear distribution is that endogenous CLIC4 appears to associate with the nuclear pore. Immunoprecipitation

experiments showed that TNF-α or etoposide treatment of keratinocytes increased the constitutive interaction between CLIC4 and various members of the nuclear import machinery including Ran, nuclear transport factor-2 (NTF2) and importin-α [123].

A C-terminal CLIC4-GFP fusion construct prevents nuclear accumulation in contrast to the N-terminal GFP-CLIC4 construct which allowed nuclear accumulation, albeit at a reduced level. This may suggest the physical interaction of CLIC4 with the nuclear import machinery occurs towards the C-terminus, which is consistent with an interaction involving the KKYR motif as a functioning NLS, as opposed to the KRR N-terminal motif. A deletion CLIC4 construct lacking the putative NLS was not able to translocate to the nucleus, whereas constructs lacking the putative TMD or N-terminal domain were able to translocate to the nucleus in the absence of treatment by TNF-α (see Figure 1.16). Because of this, it has been hypothesised that there is intrasteric regulation of nuclear import by the CLIC4 N-terminal domain. However, there is the possibility that the N-terminal deletion mutants are not properly folded and have an exposed NLS, which could also account for the observed CLIC4 nuclear translocation in untreated cells.

In a crucial experiment, the CLIC4 putative NLS motif (^{199}KVVAKKYR^{206}) was eliminated by amino acid substitution of basic residues (^{199}TVVAITYG^{206}), residues that would be essential for binding importin-α [123]. In contrast to wild type CLIC4, this CLIC4 ΔNLS mutant is unable to translocate to the nucleus in response to TNF-α or etoposide treatment of S1 keratinocytes (see Figure 1.16), providing specific evidence that the CLIC4 NLS motif is functionally significant in enabling the translocation of CLIC4 to the nucleus.

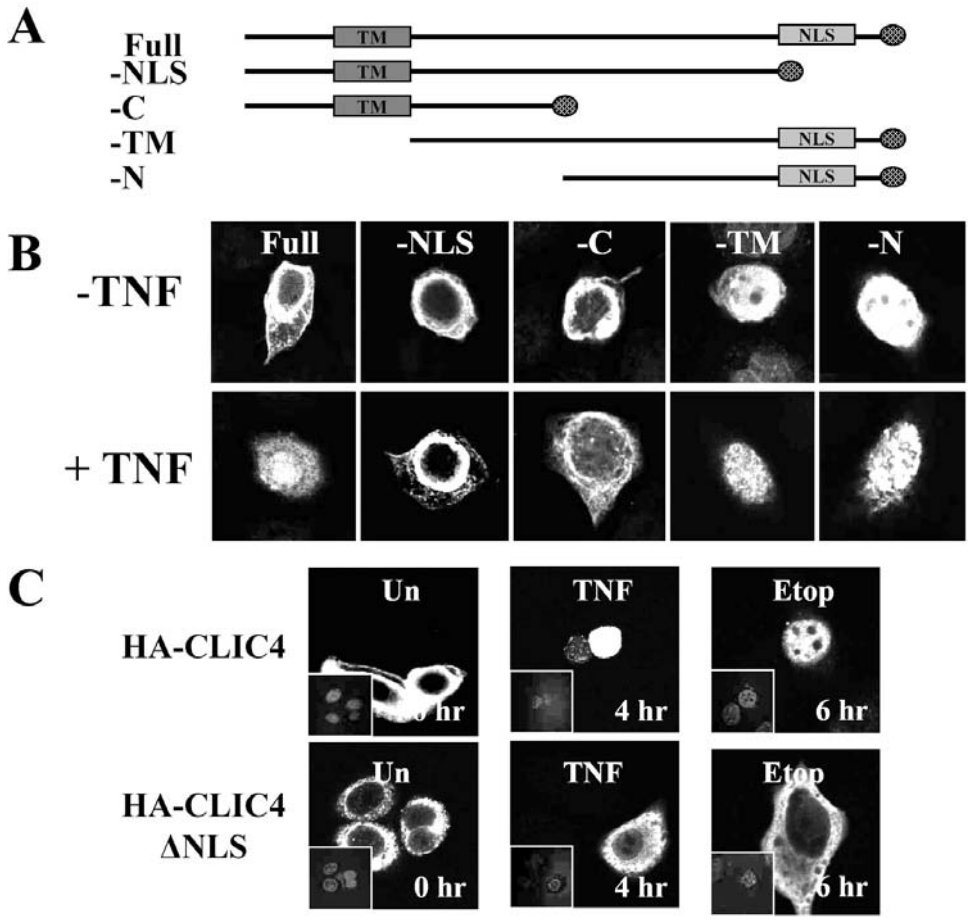

Figure 1.16: Nuclear localisation of CLIC4 WT, CLIC4 -NLS and deletion mutants before and after treatment with cellular stress inducers.

This figure shows confocal microscopy analysis of immunostained S1 keratinocytes from Suh *et al.* 2004 [123].

A) The CLIC4 deletion constructs used in the nuclear import assays, with His-V5 C-terminus tag. These include Full (WT), -NLS (no NLS domain), -C (no C-terminal domain), -N (no N-terminal domain) and -TM (no transmembrane domain).

B) S1 keratinocytes transfected with deletion constructs with (+TNF) and without (-TNF) TNF-α treatment. Images show subcellular localisation detected by immunostaining with anti-V5 antibody.

C) S1 keratinocytes transfected with the hemagglutinin (HA) tagged CLIC4 ΔNLS mutant (^{199}KVVAKKYR^{206} → ^{199}TVVAITYG^{206}) and HA tagged wild type CLIC4 (HA-CLIC4). Cells have been either untreated (Un), treated with TNF-α (TNF) or treated with etoposide (Etop). Images inset represent Hoechst nuclear staining. Subcellular localisation has been detected by immunostaining using the HA tag antibody.

1.4 Summary and Research aims

In eukaryotic cells the genetic material is compartmentalised into the nucleus where it is separated from the cytoplasm by the nuclear envelope. Trafficking of molecules across this membrane is critical for the overall function of the cell. The nuclear import pathway is mediated by a group of soluble receptors known as karyopherins, which includes the import receptors importin-α and importin-β. The majority of nuclear bound cargo proteins are recognised by the adaptor molecule importin-α, which binds positively charged nuclear localisation signals (NLSs) in an extended conformation and facilitates transfer of the cargo across the nuclear envelope via a trimeric importin-α:β:NLS complex.

It has recently been demonstrated that a member of the chloride intracellular channel (CLIC) family of proteins (CLIC4) translocates to the nucleus in response to cellular stress, and that the transport process is mediated by a putative NLS which most likely interacts with importin-α. The CLICs are a highly conserved class of chloride ion channels that consist of an ~230 residue structural module that is related to the canonical GST fold. Despite existing in a monomeric, soluble form, it is hypothesised that the CLICs are capable of functionally relevant ion channel activity and have the ability to auto-insert into membranes. CLIC homologues in invertebrate species have also been identified, and in the case of the *C. elegans* CLIC-like protein, EXC-4, extensively characterised in regards to function.

This thesis aims to yield further knowledge of CLIC proteins by examining X-ray crystallography structures of the invertebrate CLIC from *D. melanogaster* (*Dm*CLIC) and importin-α complexed with the CLIC4 NLS. Additional studies seek a better understanding of the importin-α recognition of NLSs by exploring structures of importin-α with an empty binding site, bound to the SV40 NLS and bound to endogenous *E. coli* ribosomal proteins.

Chapter 2

Crystal Structure of *Drosophila melanogaster* CLIC

2.1 Introduction

Although the generic structure of CLIC family members has been elucidated thoroughly by X-ray diffraction techniques, the proteins studied thus far have been derived mainly from *Homo sapiens* and thus limit our ability to understand their broader role in an evolutionary context. A number of CLIC homologue proteins have been discovered in invertebrate species including the model organisms *Caenorhabditis elegans* (*C. elegans*) and *Drosophila melanogaster* (*D. melanogaster*), which have provided a platform to explore the functional roles of proteins containing the conserved CLIC module structure. The single CLIC-like protein in *D. melanogaster* will be referred to as *Dm*CLIC and the two CLIC-like proteins in *C. elegans* are known as EXC-4 and EXL-1.

The sequence similarity between each of the invertebrate CLICs is high, with *Dm*CLIC and EXC-4 sharing 39% sequence identity and *Dm*CLIC and EXL-1 sharing 35% sequence identity. These invertebrate proteins also share a high degree of sequence similarity compared to mammalian CLICs (*Dm*CLIC and human CLIC4 share 30% sequence identity).

The majority of the work presented in this chapter has resulted in the publication[*], Littler *et al*. 2008 [137]. This paper describes the structures of *Dm*CLIC (PDB:2YV7) and EXC-4 (PDB:2YV9), which both contain a bound Ca^{2+} cation in what appears to be a relatively weak metal binding site. Contradicting the assignment of Ca^{2+} as the bound metal, *Dm*CLIC and EXC-4 were only briefly exposed to low concentrations of

[*] My involvement in this publication focused on structural analysis of *Dm*CLIC, the results of which are presented here in Sections 2.3.1-2.3.3. Results presented in Sections 2.3.4-2.3.6 are original work performed by myself that enhance the already published data. I have no authorship of the EXC-4 structure or channel activity experiments, as published in Littler *et al.* 2008.

Ca^{2+} during purification and the *Dm*CLIC crystal was prepared in the presence of a metal chelator (100 mM EDTA).

In this chapter, we extend our investigation of the metal binding site and seek to resolve our concerns with the identification of the bound metal cation as Ca^{2+}. We describe here the work done in obtaining an X-ray crystal structure of *Dm*CLIC at a resolution of 1.80 Å, revealing *Dm*CLIC possesses the canonical glutathione S-transferase (GST) structural fold. There are some noteworthy structural characteristics including the invertebrate C-terminal extension that adopts a unique conformation found only in *Dm*CLIC and EXC-4. Perhaps the most important discovery in the *Dm*CLIC structure is the presence of the metal binding site in the N-terminal thioredoxin-like domain that includes residues from the N-terminus of α-helix 3 and β-strand 3. The metal is coordinated by four protein ligands including the highly conserved *cis*-Pro85 found in GSTs and thioredoxins. Drawing upon single-wavelength anomalous dispersion (SAD) data and the use of the bond valence method, the metal has been identified as the monovalent cation K^+. This assignment resolves the problems posed by earlier work that suggested the site may be occupied by the divalent cation Ca^{2+}, which is isoelectronic with K^+.

2.2 Methods

2.2.1 Cloning

The clone of *Dm*CLIC was received as a gift from Prof. Mark A. Berryman[†]. The plasmid consisted of the 1385bp Drosophila melanogaster gene CG10997 (NM 132700) inserted into the pGEX-2T GST-fusion vector (Amersham Biosciences). The *Dm*CLIC plasmid was transformed into the *E. coli* BL21(DE3) expression cell line.

2.2.2 Expression and Purification of *Dm*CLIC

The BL21(DE3) cells containing wild type *Dm*CLIC plasmid were grown overnight in 10 mL volumes of Luria-Bertaini (LB) media (100 μg/mL ampicillin) at 37 °C. An inoculation from this culture to a larger 500 mL volume of LB broth (100 μg/mL ampicillin) at a ratio of 1:50 was then performed into four, 2 L flasks to obtain a total cell culture volume of 2 L. The culture was then allowed to grow at 37 °C in a shaker-incubator until an optical density (O.D.) at 600 nm of 1.0 cm^{-1} was measured, at which point the culture was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was then lowered to 30 °C and growth continued for approximately 4 hours. Cells were harvested at the end of induction by centrifugation at 8,000 rpm (11,000 *g*) for 10 minutes in a Sorvall SLA-3000 fixed angle rotor. After discarding the supernatant, cells were resuspended in 30 mL PBS (137 mM NaCl, 2 mM KCl and 10 mM phosphate buffer at pH 7.4) and frozen at -80 °C overnight.

Upon resuming protein preparation, the frozen bacterial solution was thawed on ice and 2 mM dithiothreitol (DTT) was added. The cells were then passed through a French press twice using a 'high' ratio gauge pressure reading of 1000 psi, corresponding to a working cell pressure of 16,000 psi. The resulting cell debris was

[†] Department of Biomedical Sciences, Molecular and Cellular Biology Program, Ohio University College of Osteopathic Medicine, Athens, Ohio 45701, USA.

centrifuged at 18,000 rpm (39,000 *g*) for 45 minutes at 4 °C in a Sorvall SS-34 fixed angle rotor.

The supernatant containing soluble protein was added to 2 mL of glutathione sepharose 4B resin (Amersham Biosciences), and mixed under gentle rotation for one hour. The resin was then added to a Bio-Rad Econopak gravity flow column and unbound protein in the supernatant was eluted and discarded. The resin was washed with 200 mL of PBS containing 0.5 mM DTT and then fusion protein that had bound through a GST-glutathione interaction was cleaved from the sepharose beads by adding 50 μL of 1 NIH units/μL bovine plasma thrombin (Sigma-Aldrich) to a 10 mL solution of PBS with 2 mM DTT, which was added to the resin. The thrombin digestion proceeded for 16 hours at room temperature under gentle rotation. The cleaved protein was then allowed to elute from the column and was concentrated to 5 mL using an Amicon YM-10 Centriprep centrifugal concentrator.

The sample was loaded onto a HiLoad 26/60 Superdex 75 prep grade gel filtration column (Amersham Biosciences) and eluted with a sizing column buffer consisting of 20 mM Hepes pH 7.0, 100 mM KCl and 1 mM DTT. The elution was performed at a flow rate of 1 mL/min over a volume of 300 mL and was collected in 5 mL fraction volumes. The chromatogram shows *Dm*CLIC elutes as a single peak as judged by UV absorption at 280 nm. The three fractions corresponding to peak protein elution were pooled and concentrated using a 10 kDa cut-off filter (Amicon YM-10 Centriprep centrifugal concentrator), to a volume of approximately 1 mL. The protein yield from a 2 L cell culture was between 10-20 mg, with a final concentration of 10-20 mg/mL.

2.2.3 Crystallization and data collection

Crystallization trials of *Dm*CLIC were screened using the hanging drop vapour diffusion method with a 500 μL reservoir in Hampton 24 well plates. Drops consisted of 3 μL of protein at 12.7 mg/mL and 3 μL of reservoir.

The Hampton PEG/Ion screen yielded a hit in condition #11 which consists of 0.2 M KI and 20% w/v monodisperse polyethylene glycol 3350 (PEG3350). The

crystals deteriorate quickly and were found to be more stable when grown at 4 °C. Grid screening optimisation around the initial condition revealed an improvement to crystal morphology when the conditions were shifted to 0.14 M KI and 17% PEG3350. When this condition was supplemented with 100 mM ethylenediamine tetraacetic acid (EDTA) pH 5.8, crystal size and morphology was again improved. The crystals used in diffraction experiments grew to a size of 400 μm x 200 μm x 200 μm, over a period of two days.

The *Dm*CLIC crystals were gradually moved into a D-glucose cryoprotectant solution by a series of steps that transferred the crystals to solutions of progressively higher concentrations of D-glucose. All crystal manipulation was performed at 4 °C where possible. The gradual transfer process yields better crystal quality compared to a direct transfer. The final cryoprotectant solution consisted of the reservoir solution plus 250 mg/ml D-glucose. Crystals were flash cooled in liquid nitrogen and stored in a cryogenic dry shipping dewar before data collection.

Diffraction data was obtained at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 9.2 (λ = 0.98 Å) using a MarCCD detector. The crystals diffracted to 1.80 Å and were found to have symmetry in the orthorhombic space group $P2_12_12_1$ ($\alpha = \beta = \gamma = 90.0°$) with unit cell dimensions a = 39.3 Å, b = 63.4 Å and c = 115.0 Å (see Table 2.1).

2.2.4 Structure determination and refinement

A number of human CLIC protein structures have been solved revealing a high degree of structural uniformity. Therefore, there were a number of suitable molecular replacement starting models to choose from. The human CLIC1 reduced monomer structure (PDB:1K0M), sharing a sequence identity of 25% with *Dm*CLIC, was chosen as the molecular replacement probe. The CCP4 program CHAINSAW [153] was used to remove non-conserved residues between CLIC1 and *Dm*CLIC, preparing a more accurate model for calculating phases. The initial estimate of phases was found using the CCP4 program, PHASER [154]. Iterative refinement was then performed using multiple passes of the REFMAC5 maximum likelihood function [155] and manual

model building/refinement in COOT [156]. After the majority of the C_α backbone and residue side chains had been modelled, an initial solvent model was built using ARP/wARP [157]. Final refinement iterations included the use of TLS parameters [158].

The final *Dm*CLIC model was obtained from a crystal grown in 100 mM EDTA and consists of one *Dm*CLIC molecule within the asymmetric unit. Included in the model are residues 15-258, 163 waters, two iodide ions and one potassium ion. Residue Ile228 has been modelled with two alternative rotamers each at 50% occupancy. All other atoms have an occupancy of 1.0 apart from the iodide ions, which have an occupancy of 0.9 and 0.3. Only one residue adopts a *cis* peptide: Pro85.

Table 2.1: Data reduction and refinement statistics for *Dm*CLIC

Data Collection		
Source (λ)	SSRL BL 9.2 (0.98 Å)	Rotating anode (1.54 Å)
Detector	MarCCD	Mar345
Space Group	$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)		
a	39.3	39.4
b	63.4	62.6
c	115.0	114.1
Resolution (Å)[a]	1.80 (1.90-1.80)	1.90 (2.00-1.90)
Total Observations	96,281	68,536
Unique Observations	26,761	21,257
Completeness (%)[a]	98.0 (91.7)	92.9 (87.3)
Mean I/σ(I)[a]	15.5 (3.2)	11.3 (2.7)
R_{merge} (%)[a,c]	5.5 (31.6)	5.6 (40.2)
Wilson B-value ($Å^2$)	29.2	30.1
Anomalous Completeness (%)[b]	98.2 (98.4)	89.1 (79.7)
Refinement and Structure		
R factor (R_{free})[c]	19.2 (24.2)	20.9 (26.4)
Number of atoms (Avg. B-factor, $Å^2$)		
*Dm*CLIC	1,967 (26.8)	1916 (24.8)
Waters	163 (35.3)	141 (31.3)
Ions (K^+, I^-)	1 (34.5), 2 (38.0)	1 (46.0), 1 (44.6)
Ramachandran plot (%)[b]		
Favoured region	94.9	96.1
Allowed region	4.7	2.6
Disallowed	0.4 (E244)	1.3 (R76, E80, E244)

[a] Outer resolution shell statistics are shown in parenthesis.
[b] Calculated using MolProbity [159].
[c] The R_{merge} and R factor are defined in Section 2.2.7.1, p72.

2.2.5 Single-wavelength anomalous dispersion

Anomalous dispersion data was simultaneously collected at the SSRL (λ = 0.98 Å), along with X-ray diffraction data that was subsequently used to solve the *Dm*CLIC structure. Using single-wavelength anomalous dispersion (SAD) data and the fast Fourier transform (FFT) program from CCP4, we constructed Bijvoet difference

Fourier maps with coefficients ($|F^+|-|F^-|$, $\varphi+\pi/2$). The phases used for this map were calculated from the refined model.

Since anomalous scattering by heavy atoms is stronger at longer wavelengths, we also collected SAD data from our home source Nonius rotating anode generator using Cu K_α radiation where $\lambda = 1.54$ Å. This data was collected using a different *Dm*CLIC crystal than that used to solve the 1.80 Å structure. Another *Dm*CLIC model was solved using Cu K_α X-ray diffraction data and similar refinement methods as outlined in Section 2.2.4. The phases calculated from this data were used to construct new Bijvoet difference Fourier maps at $\lambda = 1.54$ Å. Therefore we have two SAD datasets: one that was collected at $\lambda = 0.98$ Å and one at $\lambda = 1.54$ Å.

Theoretical anomalous dispersion coefficients used in this chapter were taken from tables prepared by the Biomolecular Structure Center [160], using the theoretical approximation developed by Cromer and Liberman [161]. A plot of the scattering coefficients f′ and f″ for heavy atoms relevant to the current study (iodine, potassium, calcium and sulfur) is shown in Figure 2.1. Since the imaginary scattering component, f″, is proportional to the absorption of X-rays, it can thus be used as a guide to the expected values of relative experimental anomalous signals.

To calculate relative anomalous signals we have used density peaks in the Bijvoet difference Fourier map or integrated the density around the position of the peak. To do this integration we have used MAPMAN from the RAVE program suite to find peaks in the map and then integrate around these peaks in a spherical volume of either 1.8 Å (K^+ or sulfur peaks) or 2.0 Å (I^-).

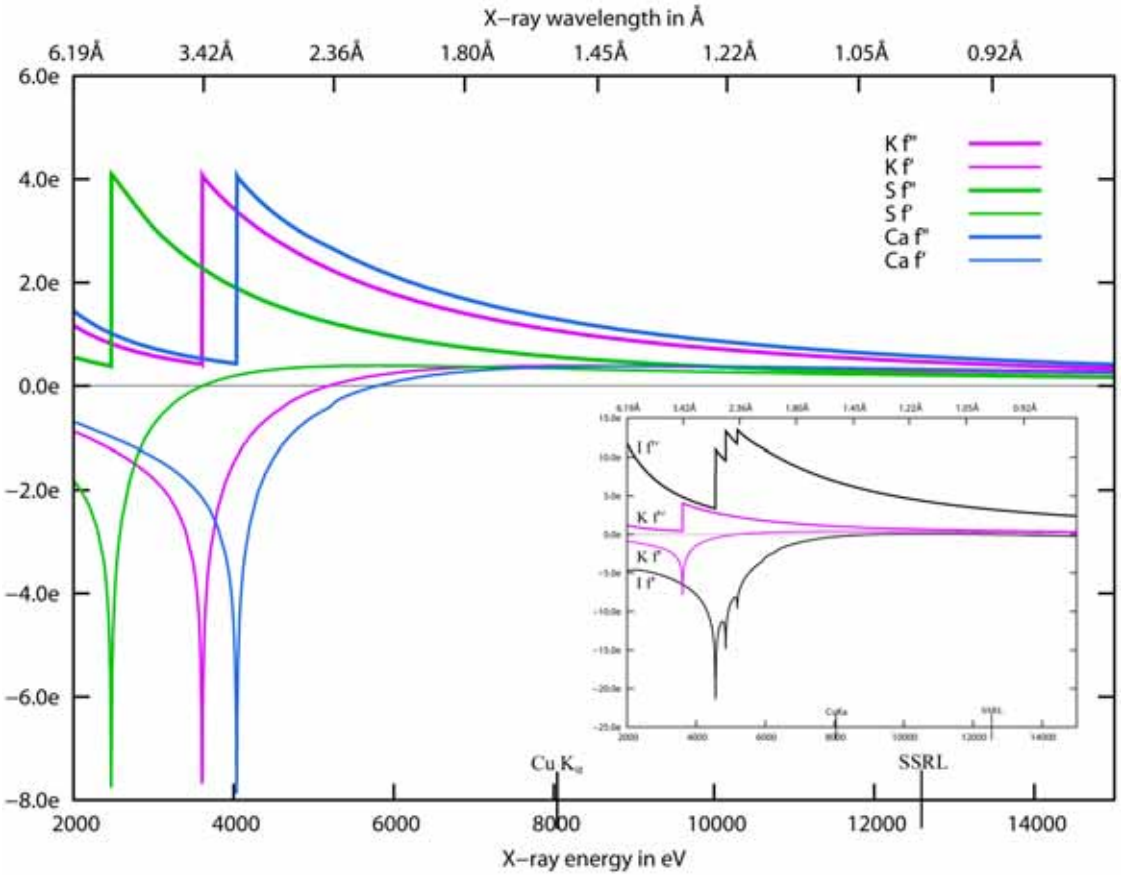

Figure 2.1: A plot of theoretical anomalous dispersion coefficients.

The anomalous dispersion coefficients, F″ and f′, are shown in units of electrons (e) for the elements K, Ca and S. The energy of Cu K_α (λ=1.54 Å) and Stanford Synchrotron Radiation Laboratory (SSRL, λ=0.98 Å) X-ray sources are labelled. The plot was generated using the Biomolecular Structure Center web interface [160].

Inset: The iodine and potassium theoretical anomalous dispersion coefficients are shown here on separate axes.

2.2.6 The bond valence method and the calcium bond valence sum

In order to identify the bound metal ligand in the *Dm*CLIC structure, we have made use of the bond valence method. This is an extension of Pauling's second rule involving the principle of electrostatic valency and its application to the identification of bound metal ions in protein structures has been investigated by Müller *et al*. [162]. As applicable to the *Dm*CLIC structure containing a bound metal cation (atom *i*), the bond valence (v_{ij}) of each bond is defined as,

$$v_{ij} = p_j \exp\left(\frac{d_0 - d_{ij}}{b}\right)$$

where d_0 is the bond valence parameter equal to the bond length of an ideal single bond between atoms *i* and *j*, d_{ij} is the measured bond length between the cation (*i*) and bonded oxygen atom (*j*), *b* is a constant equal to 0.37 and p_j is the occupancy of the atom *j*. The value of d_0 for K^+ corresponds to d_0(KO) = 2.13 and for Ca^{2+}, d_0(CaO) = 1.967. The valency of the cation *i*, is then the sum of all bond valences,

$$V_i = \sum_j v_{ij}$$

The calcium bond valence sum (CBVS) is introduced by Müller *et al*. mainly for the convenience of easily identifying metals in multiple structures using the bond valence equation above. This is done by setting the bond valance parameter to $d_0 = d_0$(CaO) = 1.967, which is not changed when dealing with a variety of metal ligands in multiple structures. The use of the CBVS gives rise to characteristic values of V_i that correspond to specific metal cations.

2.2.7 Structure analysis

2.2.7.1 The R-factor estimation of model quality

To estimate the quality of a structural model in comparison with our experimental observations, we have used two R-factor parameters which have both been calculated using REFMAC5 [155]. The first is referred to as the R-factor (or $R_{working}$) which is defined as,

$$R = \frac{\sum_{hkl} \left| \left| \mathbf{F}_{obs} \right| - \left| \mathbf{F}_{calc} \right| \right|}{\sum_{hkl} \left| \mathbf{F}_{obs} \right|}$$

where the summation is over all working reflections (*hkl*) used during refinement and $\mathbf{F}_{obs}$ and $\mathbf{F}_{calc}$ are the observed and calculated structure factors, respectively. This R-factor and the free R-factor (R_{free}) are equivalent, except that R_{free} is calculated using 5% of the total reflections that have been flagged and not included during refinement. The R_{free} has been calculated this way for all datasets used in this thesis.

The second R-factor we use is the merging R-factor (R_{merge}) also known as the symmetry R-factor (R_{sym}). This is defined as,

$$R_{merge} = \frac{\sum_{hkl} \sum_{i} \left| I_i - \langle I \rangle \right|}{\sum_{hkl} \sum_{i} I_i}$$

where the summation is performed over reflections (*hkl*), I_i is the observed intensity of an individual measurement of one reflection and <I> is the mean intensity of symmetry related observations of that reflection.

2.2.7.2 Electron density maps

Electron density in this thesis is represented by F_o-F_c and $2F_o$-F_c unweighted maps or $2mF_o$-DF_c and mF_o-DF_c sigma-A weighted maps. The *m* coefficient is a figure of merit for model phases and *D* is a general estimate of atomic coordinate errors in the model. They are defined as,

$$m = \left\langle \cos\left(\varphi^{true} - \varphi^{calc}\right)\right\rangle \quad \text{and} \quad D = \left\langle \cos\left(2\pi \mathbf{s}.\Delta r\right)\right\rangle$$

where φ^{true} corresponds to the true phases (a partial model plus missing atoms), φ^{calc} is the calculated phase (a partial model only), **s** is the reciprocal lattice vector and Δr is the coordinate error (see Read *et al*. 1986 for more details [163]). The sigma-A coefficients were calculated by REFMAC5 [155] and the Fourier transform was performed using FFT. The $2mF_o$-DF_c and mF_o-DF_c electron density maps are calculated by the Fourier summation shown here (refer to [164]),

$$\rho(x, y, z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\mathbf{F}\cdot e^{-2\pi i(hx+ky+lz)}$$

where $\mathbf{F} = \left(2m\left|F_o\right| - D\left|F_c\right|\right)\cdot e^{i\alpha_c}$ or $\mathbf{F} = \left(m\left|F_o\right| - D\left|F_c\right|\right)\cdot e^{i\alpha_c}$.

In the above equation, ρ(x,y,z) is the electron density function over real space, V is the volume of the unit cell, *hkl* are miller indices, F_o is the observed structure factor, F_c is the calculated structure factor and α_c is the calculated phase angle.

2.2.7.3 Structure alignments

Structural alignments have been performed and the Root Mean Square Deviation (RMSD) of the distance between equivalent atoms in two structures has been quoted throughout the thesis. This normally involves comparing selected C_α atoms that have

been aligned using PyMOL [165] or the Swiss-PdbViewer [166]. The formula for RMSD is given here,

$$RMSD = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (d_{ii})^2}$$

where d_{ii} is the distance between the i^{th} atom of the first structure and the i^{th} atom of the second structure and N is the number of atoms aligned.

2.2.7.4 Accessible surface area

Solvent accessible surface area calculations were performed using the CCP4 program AREAIMOL [167]. The solvent accessible surface area is defined as the set of points generated by the centre of a probe sphere rolling across the van der Waals surface of the protein. The molecular surface is different to the solvent accessible surface, as the set of points are taken from the inner edge of the probe sphere and not the centre. For our calculations the probe sphere has a radius of 1.4 Å. The buried solvent accessible surface area in importin-α:NLS interfaces are reported as the surface area obscured by the ligand on the surface of the protein.

2.3 Results

2.3.1 The crystal structure of *Dm*CLIC

The crystal structure of monomeric *Dm*CLIC has been determined at a resolution of 1.80 Å using synchrotron radiation (λ = 0.98 Å). Although there was no reducing agent added to the crystallization solution, the protein sample was kept in a reduced environment (1 mM DTT) prior to crystallization and so the structure represents a reduced monomeric form of *Dm*CLIC. The structure was solved in the $P2_12_12_1$ space group with one copy of monomeric *Dm*CLIC in the asymmetric unit. The final model shown in Figure 2.2 includes residues 15-258, 163 waters, two I^- ions and one K^+ cation. Although the entire C_α chain from Ser15 to Ile258 has been modelled, there are two loop regions where the electron density is too weak to unambiguously model side chains. This occurs in the connecting loop between β-strand 2 and helix 2 (Lys70), as well as the beginning of the C-terminal extension (Lys238, Met239, Lys240, Lys241 and Glu244).

The average atomic B-factor for protein main chain and side chain atoms is 26.8 $Å^2$, while for waters it is 35.3 $Å^2$. The potassium ion has been modelled with a B-factor of 34.5 $Å^2$ and an occupancy of 1.0. One iodide ion has a B-factor of 28.8 $Å^2$ and an occupancy of 0.9 and the other iodide ion has a B-factor of 47.2 $Å^2$ and an occupancy of 0.3. The Ramachandran plot shows 99.6% of residues (240/241) lie in favoured regions while one residue (Glu244) is in a disallowed region. Glu244 resides in the C-terminal extension (residues 236-258) where only weak electron density corresponds to the C_α main chain.

The structure confirms that *Dm*CLIC adopts the canonical GST fold which is maintained across all known CLIC structures in both vertebrate and invertebrate species. Compared to the *C. elegans* CLIC, EXC-4, there are few differences in the C_α backbone over the core structural regions with an RMSD of 1.2 Å over 218 C_α atoms (calculated using the Swiss-PdbViewer [166]). There is also a high degree of similarity when comparing *Dm*CLIC with vertebrate CLIC proteins such as human CLIC1, with an RMSD of 1.23 Å over 195 C_α atoms (see Figure 2.2).

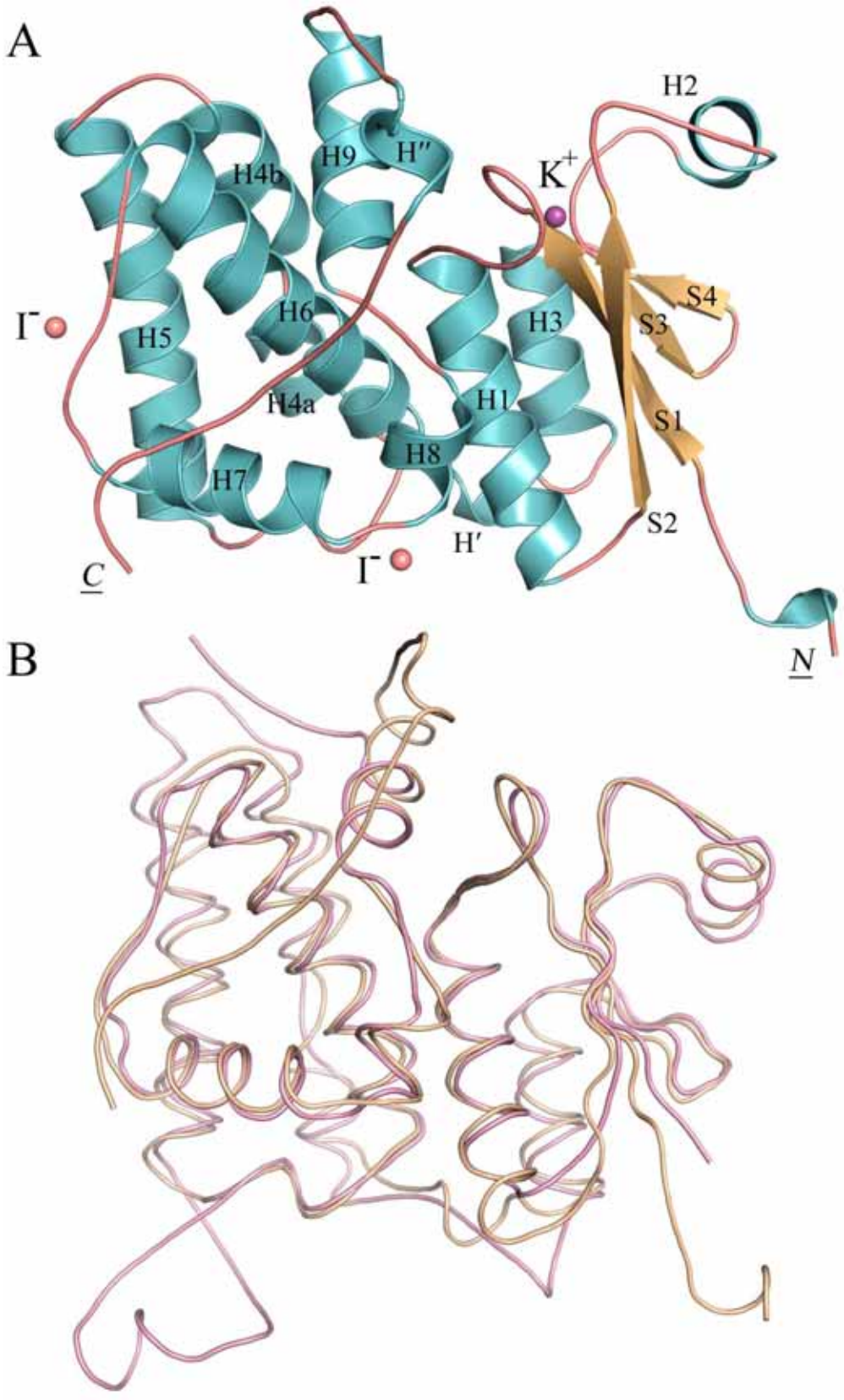

Figure 2.2: The structure of *Dm*CLIC.

A) The *Dm*CLIC structure is shown in cartoon representation. The α-helices are labelled H1 to H10 and β-strands S1-S4. A bound potassium ion (magenta) and iodide ions (pale pink) are shown as spheres.

B) Backbone alignment of *Dm*CLIC (light brown) and CLIC1 (pink) shown as C_α loops.

While many of the main structural elements of the vertebrate CLICs and EXC-4 are preserved in *Dm*CLIC, there are some minor differences observed at helix termini and in loop regions. What follows is a summary of these structural features (refer to Figure 2.2 and Figure 2.3).

Structural comparison between DmCLIC (& EXC-4) and vertebrate CLICs

- There is a four residue C-terminal extension of helix 1 in *Dm*CLIC, which is also present in EXC-4. Helix 1 constitutes part of the CLIC putative TMD that corresponds to residues 41-66 in *Dm*CLIC (Figure 2.3A).
- The connecting loop linking the N-terminal thioredoxin-like domain with the C-terminal all α-helical domain is shorter in the *Dm*CLIC structure and some of the flexible coil has adopted a short α-helical turn, labelled H′ (Figure 2.3B).
- *Dm*CLIC has a shorter helix 5 (five residues on the N-terminus) than the corresponding helix in vertebrate CLICs (Figure 2.3D).
- The connecting loop between helices 4 and 5 is shorter in *Dm*CLIC. The nematode CLIC, EXC-4, contains a unique sequence insertion of approximately 15 residues at this location (Figure 2.3D).
- Both *Dm*CLIC and EXC-4 do not contain the characteristic foot loop region located between helices 5 and 6, which is present in vertebrate CLICs (Figure 2.3E).
- *Dm*CLIC and EXC-4 share a C-terminal extension that snakes across the C-terminal domain (see Figure 2.3F and Figure 2.4).
- *Dm*CLIC possesses an N-terminal extension that forms a short α-helical turn over residues 16-18. It is likely this region is observed only due to fortuitous crystal packing interactions (see Section 2.3.6). EXC-4 possesses a similar but longer N-terminal extension observed in subunit A of the crystallographic dimer.

Structural comparison between DmCLIC and EXC-4

- While *Dm*CLIC shares the feature of a helix break at helix 4 with vertebrate CLICs, splitting the helix into H4a and H4b, EXC-4 is unusual in that helix 4 is continuous (Figure 2.3C).
- There is a 4 residue extension at helix 3 in the EXC-4 structure.
- EXC-4 possesses a slightly longer helix, labelled H″, in the C-terminal extension, (residues 242-245 in *Dm*CLIC and residues 256-262 in EXC-4, see Figure 2.3F).
- *Dm*CLIC does not possess a 10^{th} helix that is found at the C-terminus in EXC-4 (Figure 2.3F). This helix is presumably stabilised by crystal packing interactions unique to the EXC-4 crystal.

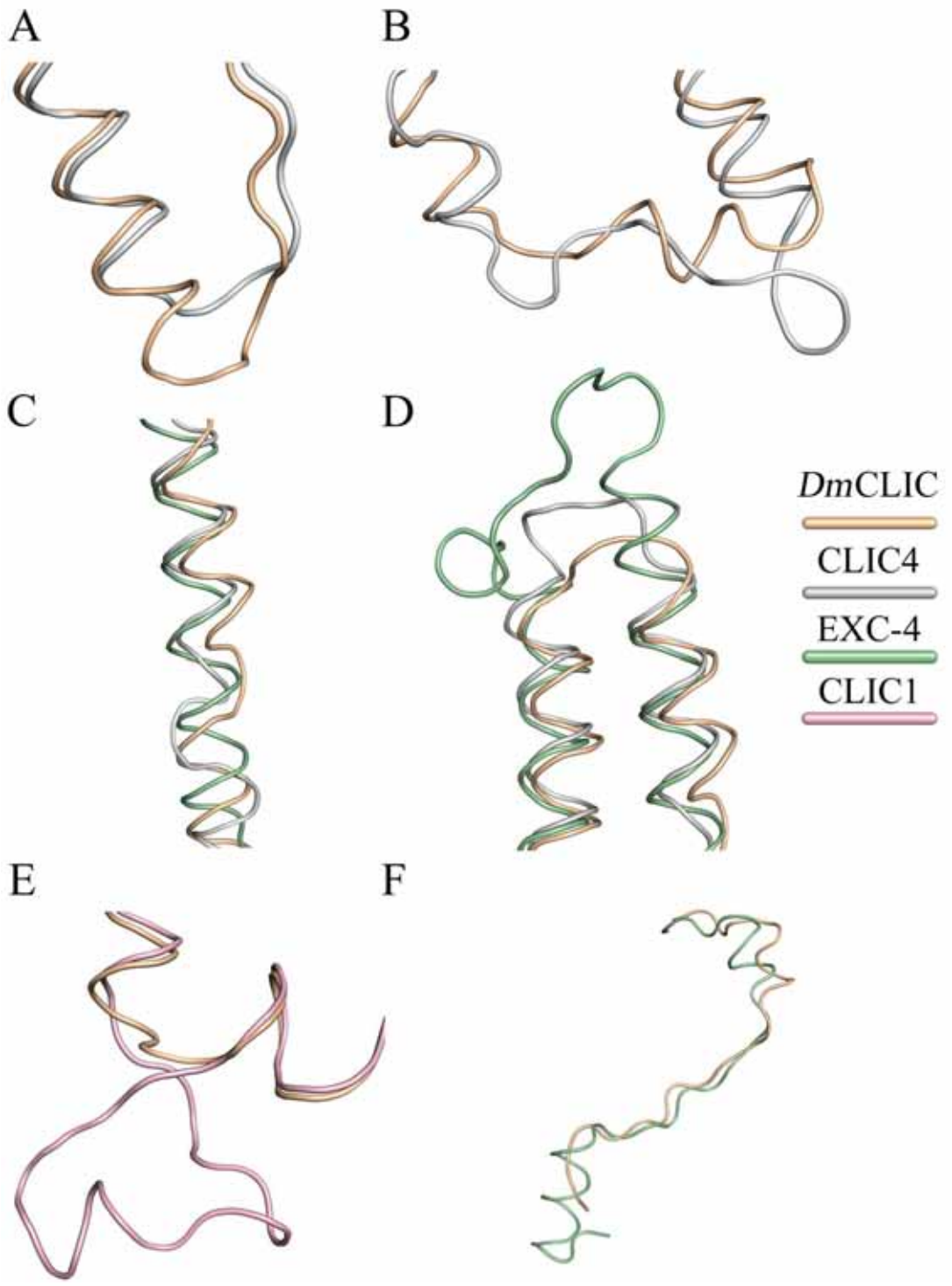

Figure 2.3: Structural differences between *Dm*CLIC, EXC-4 and vertebrate CLICs (CLIC1 & CLIC4).

A) C-terminus extension of helix 1 in *Dm*CLIC compared to CLIC4.
B) The interdomain loop and helix H′ in *Dm*CLIC, compared with the same region in CLIC4.
C) A comparison of helix 4 showing the helix splitting in *Dm*CLIC and CLIC4, but not EXC-4.
D) The connecting loop between helix 4 and helix 5 in *Dm*CLIC, CLIC4 and EXC-4.
E) The foot loop region in CLIC1 and *Dm*CLIC.
F) The C-terminal extension in *Dm*CLIC and EXC-4.

The colour scheme follows on from Figure 2.2 where applicable and is shown in the picture on the right: *Dm*CLIC is light brown, EXC-4 is green, CLIC4 is grey and CLIC1 is pink.

2.3.2 The C-terminal extension

Of the novel structural features found in the *Dm*CLIC and EXC-4 structures, perhaps the most notable is the extended C-terminal tail segment that stretches across the helical domain of the CLIC module away from the putative CLIC GSH binding site and metal binding site. This extension region deviates from the vertebrate CLIC C-termini after a conserved tyrosine residue in helix 9, which corresponds to Tyr231 in *Dm*CLIC (Tyr233 in CLIC1). Helix 9 continues past this point for another four residues before the C-terminal loop region begins. Thus, in *Dm*CLIC the C-terminal extension corresponds to residues 236-258. The tail extension forms a network of hydrogen bonds and hydrophobic contacts with residues spanning helices 6, 7 and 9, which are shown in Figure 2.4. The surface of the CLIC module in which the extension binds is relatively flat, with no significant electrostatic interactions. The side chains of residues 238 to 244 have not been modelled due to weak electron density in the region. The C_α backbone that has been modelled displays high B-factors and therefore this region is likely to be highly flexible, corresponding to the lack of significant interactions with the body of the *Dm*CLIC molecule. The C-terminal tail segment as shown in Figure 2.4 only encompasses residues 246-258 over which interactions take place with the *Dm*CLIC module and symmetry neighbours.

The residues in the C-terminal extension and *Dm*CLIC module that form hydrogen bonded pairs include: Glu246-Gln225, Thr249-Gln225/Asp224, Thr251-Phe250 and Gln183-Tyr253/Arg186 (see Figure 2.4). The majority of these hydrogen bonds involve side chain interactions although there is main chain hydrogen bonding occurring between a highly conserved aspartic residue (Asp224) at the N-terminus of helix 9 and Thr249 in the extension region.

There are also hydrophobic crystal packing interactions acting on the C-terminal extension involving Tyr253, Ile254 and Pro255. The buried surface area due to this packing is 80.7 $Å^2$, 10.1 $Å^2$ and 11.4 $Å^2$ respectively per residue, mostly accommodated by the C-terminus of helix 5 in a symmetry related molecule. The average atomic B-factors in the C-terminal extension are 26.3 $Å^2$ for main chain atoms, 28.5 $Å^2$ for side chain atoms and 27.4 $Å^2$ overall. These values are comparable to those

for the entire molecule, which have an average of 25.3 $Å^2$ for main chain atoms, 28.3 $Å^2$ for side chain atoms and 26.8 $Å^2$ overall.

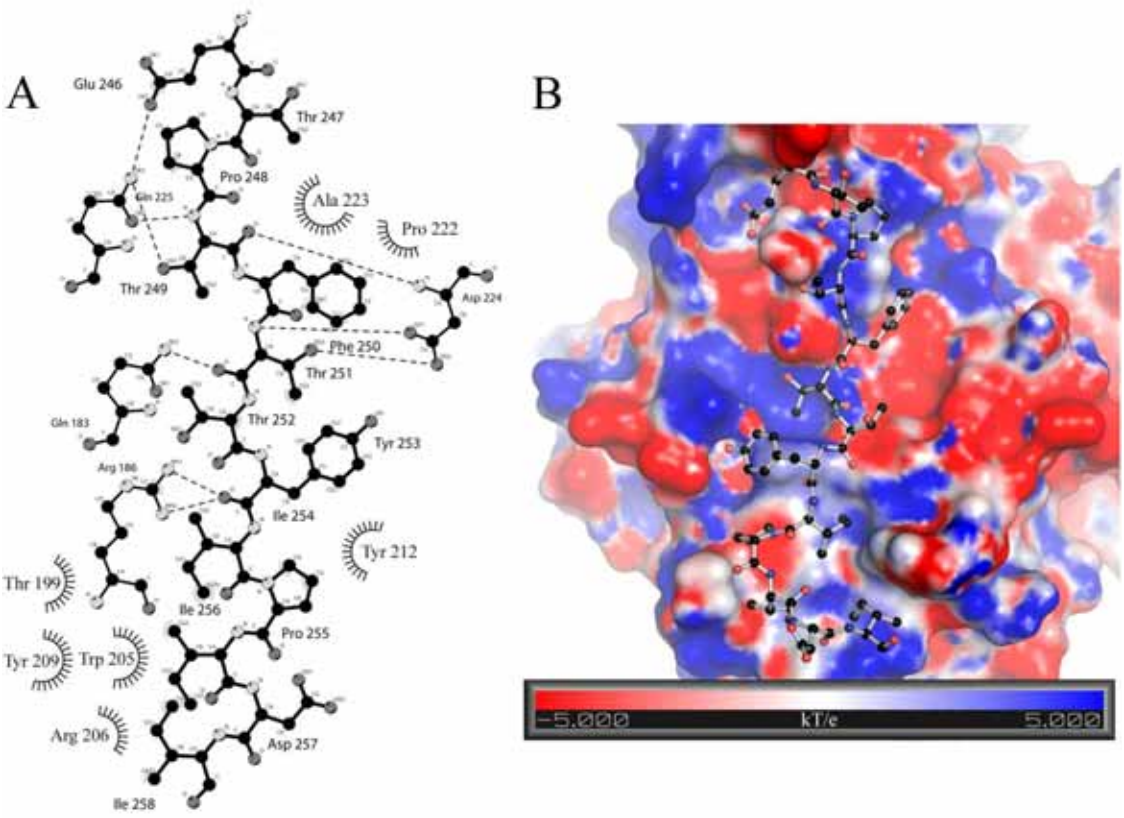

Figure 2.4: Interactions involving the *Dm*CLIC C-terminal extension.

A) Shown is a 2-dimensional schematic diagram of interactions between the *Dm*CLIC C-terminal tail (residues 246-258) and the rest of the *Dm*CLIC module. This does not include interactions involving symmetry molecules. Hydrogen bonds are represented by dashed lines while hydrophobic contacts are shown by arcs with radiating spokes. Figure generated with LIGPLOT [168].

B) The C-terminal extension is shown in ball and stick representation. The *Dm*CLIC structure is coloured by surface potential from -5 kT/e to +5 kT/e, calculated using APBS [169] and drawn with PyMOL [165].

2.3.3 The conserved CxxC motif

Similar to other CLIC proteins, *Dm*CLIC appears to have an intact putative redox site near the N-terminus of helix 1. This consists of a four residue motif, ^{40}CLFC^{43}, that is comparable to the CP[FY]C sequences of thioltransferases (glutaredoxins). In *Dm*CLIC, Cys40 corresponds to a highly conserved cysteine residue found in most CLICs at the N-cap position of helix 1 and in CLIC1 the equivalent cysteine (Cys24) is

found to bind GSH [87]. Typical of proteins with the thioredoxin fold, only the first cysteine, Cys40, is accessible while the other cysteine, Cys43, is buried.

Electron density around the CxxC motif is clear and shows the two cysteine residues within the glutaredoxin-like motif are reduced (Figure 2.5A). The side chain conformations orientate the sulfhydryl groups towards each other, with the *Dm*CLIC S-S distance equal to 3.6 Å. Although the *Dm*CLIC CxxC motif's cysteine side chain conformations are almost identical in CLIC2 and CLIC3 (Figure 2.5), the cysteines form a disulfide bridge in CLIC2 and CLIC3 with a bond length of ~2.2 Å. We note that compared to the invertebrate CLICs, the peptide bond preceding the first cysteine is rotated ~90° in vertebrate CLICs which may have an effect on the cysteinyl pK_a but is unlikely to be necessary for disulfide bridging.

Comparing the *Dm*CLIC CxxC motif to CLIC1 and CLIC4, both of which contain a CPFS motif, the conserved cysteine and conserved phenylalanine residues adopt the same side chain conformations. The serine side chain also adopts the same conformation as the equivalent cysteine.

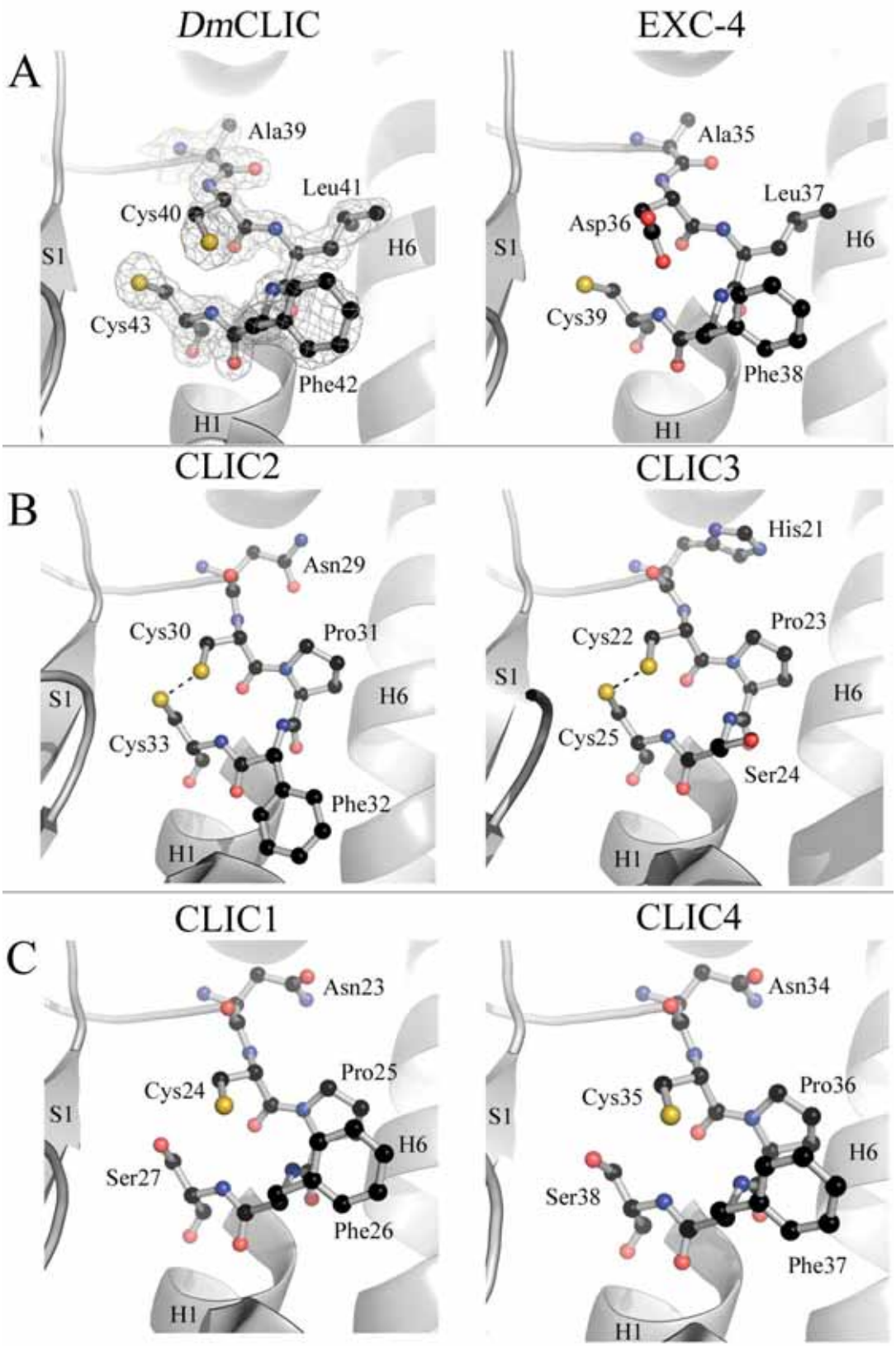

Figure 2.5: A structural comparison of the conserved CxxC motif in CLICs.

Residues in the motif are shown in ball and stick representation while the surrounding structure is shown in cartoon representation. Disulfide bonds are shown as a dashed line.

A) The CxxC motif of *Dm*CLIC and DxxC motif in EXC-4. Electron density ($2mF_o$-DF_c) is contoured at 1.5σ in grey. **B)** The CxxC motif of CLIC2 and CLIC3. **C)** The CxxS motif of CLIC1 and CLIC4.

2.3.4 The metal binding site

An interesting feature observed in the structure of *Dm*CLIC is the presence of a bound metal ion interacting with N-terminal domain residues located near the N-cap positions of α-helix 3 and β-strand 3 (Figure 2.6). The bound metal has been identified as the monovalent potassium cation (K^+) based on the following criteria: the electron density centred on the metal site is spherical at a sigma level well above that of well ordered water molecules, the number of neighbouring oxygen atoms is commensurate with metal coordination, there is an anomalous dispersion signal corresponding to the metal's position and metal to oxygen distances correspond to a K^+ cation (refer to Section 2.3.5). The K^+ cation has been modelled at full occupancy with a refined B-factor of 34.5 $Å^2$. This value is higher than the B-factor of its four protein ligand atoms which have an average atomic B-factor of 24.7 $Å^2$.

The K^+ ion is ligated by six electronegative oxygen ligands in total, four are protein ligands and two are solvent molecules (Figure 2.6). The protein ligands include backbone carbonyl oxygen atoms from His83 (bond length of 2.64 Å), *cis*-Pro85 (2.51 Å), Leu95 (2.67 Å) and the side chain oxygen (O^δ) of Asn97 (2.76 Å). In addition, there are two water molecules at a distance of 3.19 Å and 3.21 Å. The triangle formed by oxygen atoms from His83, Leu95 and Asn97 has the K^+ ion sitting approximately along the same plane in the centre. The oxygen from *cis*-Pro85 sits below this plane almost perpendicular to the centre of the triangle. The two water ligands are on the opposite side of the triangular plane from *cis*-Pro85, almost equally offset from the centre. The geometry of this hexameric coordination sphere could be described as pseudo-trigonal bipyramidal symmetry, where the positions of two water ligands substitute for a single ligand.

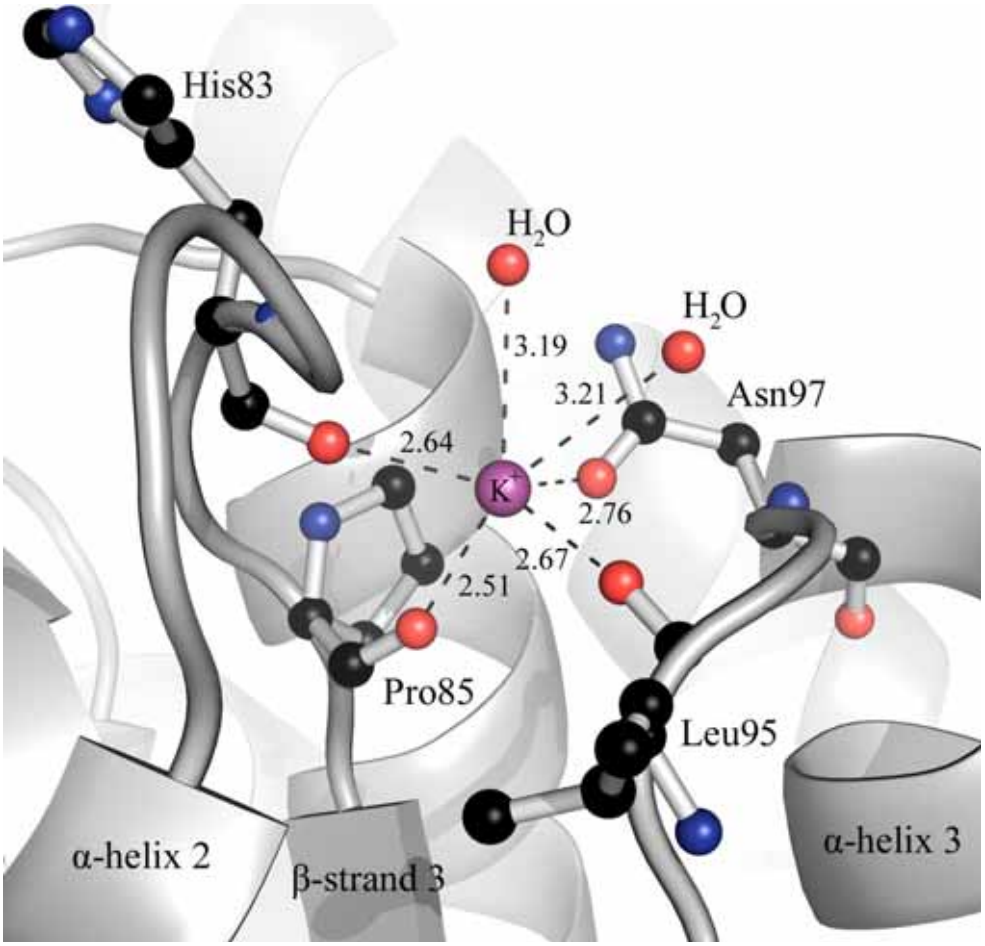

Figure 2.6: The bound potassium ion in *Dm*CLIC.

The K^+ coordination sphere is shown in ball and stick representation with carbon (black), oxygen (red) and nitrogen (blue) atoms. The potassium cation is shown in magenta. The structure of *Dm*CLIC is shown in cartoon representation (grey). The six K^+-O interactions are represented by dashed lines and the bonding distances are shown in units of ångströms.

2.3.5 Identification of the bound K^+ metal cation.

2.3.5.1 Electron density analysis

To identify the bound metal ion, we first rule out the possibility that electron density in the binding site corresponds to an ordered water molecule. A water molecule modelled in the site refines to a B-factor of 12.0 $Å^2$, significantly lower than local protein atoms including the four protein oxygen ligands which have an average B-factor of approximately 24.7 $Å^2$. We also note that the R-factor and R_{free} values do not display appreciable changes when the bound ligand is modified from a water molecule to K^+, and so these R-factor assessments of model quality have not been used

to aid identification. Refinement of a water at the binding site clearly reveals this assignment is inadequate to match the strength of electron density corresponding to the ligand. Protein-ligand bonding distances vary only slightly (≤ 0.02 Å) with a water oxygen at the metal site compared to when K^+ is modelled there. Hence we see no effects from van der Waals bumping constraints that may be applied during refinement due to the differences in atomic sizes of oxygen and potassium.

We have considered atoms heavier than oxygen including period 3 metal ions such as Na^+ and Mg^{2+} and period 4 ions such as K^+ and Ca^{2+}. Using these ligand candidates in a trial and error approach to refinement showed that the electron density at the binding site is most closely accounted for by either a K^+ or Ca^{2+} (see Figure 2.7).

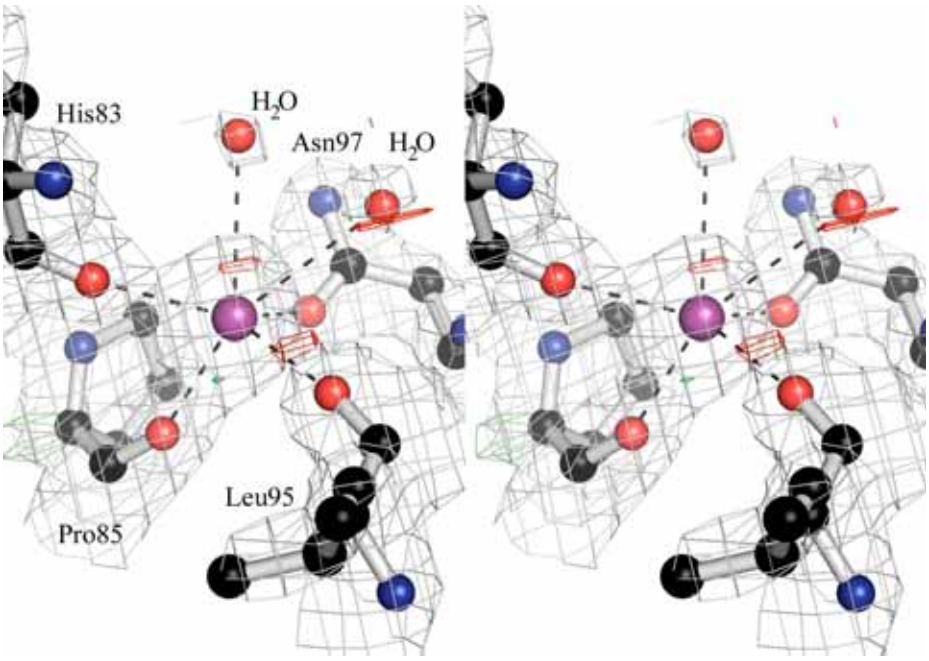

Figure 2.7: Stereo view of electron density around the bound K^+ ion in *Dm*CLIC.

The electron density maps shown are the $2mF_o$-DF_c map (grey 1.3σ), and the mF_o-DF_c difference map (green +3σ, red -3σ). Residues participating in ligand interactions are shown in ball and stick representation.

2.3.5.2 SAD data analysis

More compelling evidence that rules out a water molecule or other light atoms such as Na^+ and Mg^{2+} as ligand candidates in the binding site is the presence of a weak

anomalous signal centred on the bound metal (see Figure 2.8). To analyse anomalous dispersion from our *Dm*CLIC crystal, we have used single wavelength radiation from either the Stanford Synchrotron Radiation Lab (SSRL, $\lambda = 0.98$ Å) or from our home X-ray source which produces Cu K_α radiation ($\lambda = 1.54$ Å). Both of these sources are far from the absorption edges associated with the heavy atoms we expect could be present in the crystal, such as K^+ (K-absorption edge, $\lambda_K = 3.44$ Å), Ca^{2+} ($\lambda_K = 3.07$ Å) and I^- (L-absorption I, $\lambda_{L-1} = 2.39$ Å). Hence the anomalous signal from these atoms will be weak, particularly at synchrotron wavelengths.

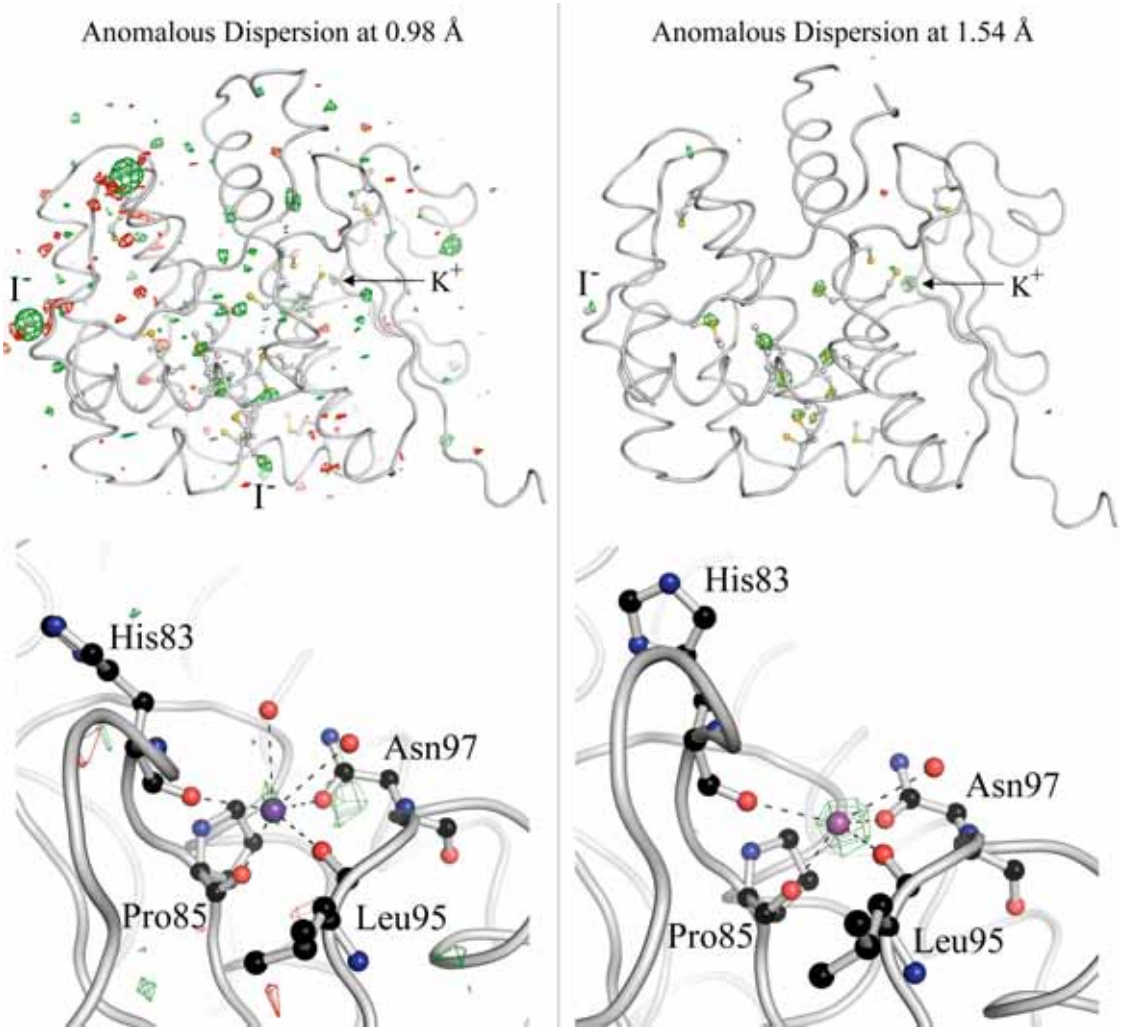

Figure 2.8: The *Dm*CLIC Bijvoet anomalous difference Fourier map.

Both Bijvoet difference maps have been calculated using anomalous data to 2.4 Å resolution. The synchrotron anomalous map has been contoured at ± 3σ and the home source anomalous map has been contoured at ± 4σ. Positive contours are green and negative contours are red. Heavy atoms associated with anomalous scattering are shown as spheres: sulfur is coloured yellow, the iodide ion is light brown and the potassium ion is magenta.

In the λ = 0.98 Å anomalous map there is a strong positive peak in the density (25.1σ) that corresponds to an iodide ion at 90% occupancy. The theoretical imaginary term of anomalous scatter from iodine at this wavelength is f″ = 3.2e. There are more positive peaks at 5.7σ and 5.1σ which cannot be unambiguously accounted for in the model. The next peak at 4.5σ corresponds to a well ordered sulfur atom (S^{δ}) in Met179 and a peak at 4.2σ corresponds to the second iodide ion binding site at 30% occupancy. By comparison, the metal binding site has a relatively weak anomalous signal with a density peak at 3.5σ (see Figure 2.8).

We have estimated the relative anomalous scattering strength of the bound metal by comparing the intensity of Bijvoet difference map peaks corresponding to the metal cation and a representative sulfur atom in *Dm*CLIC. We have also integrated the anomalous density in the Bijvoet difference map centred on each atom and compared these values. The anomalous signal of a scattering atom is known to be dependent on B-factor and therefore the S^{δ} atom in Met68 has been chosen as the internal standard since it has been modelled with a similar atomic B-factor to the bound metal. Properties of the anomalous scatterers in *Dm*CLIC are shown in Table 2.2.

For λ = 0.98 Å anomalous data, the relative magnitude of Bijvoet map peaks corresponding to the bound metal (0.026 e/Å^3) and S^{δ}-Met68 (0.020 e/Å^3) is 1.3. The relative integrated density at the position of the bound metal (0.41e) and S^{δ}-Met68 (0.28e) is 1.5. These values are comparable to the theoretical anomalous scattering power of potassium (K, f″ = 0.46e) relative to sulfur (S, f″ = 0.23e), which is a value of 2.0.

Table 2.2: Properties of anomalously scattering heavy atoms in *Dm*CLIC.

Radiation, λ	Anomalous Scatterer[a]	B-factor (Å^2)	$2mF_o$-DF_c Density Peak (σ)	Anomalous Peak (σ, x10^{-3} e/Å^3)	Integrated Signal (e)	Theoretical f″ (e)[b]
0.98 Å	I^-	28.8	28.0	25.1, 183	2.62	3.17
	K^+	34.5	8.1	3.5, 26	0.41	0.46
	S^{δ}-Met68	40.7	4.2	2.9, 21	0.16	0.23
1.54 Å	I^-	44.6	5.1	5.0, 33	1.03	6.91
	K^+	46.0	5.5	6.6, 44	1.28	1.08
	S^{δ}-Met68	47.4	4.0	3.6, 23	0.67	0.56

[a] The λ=0.98 Å iodide ion has an occupancy of 0.9. The λ=1.54 Å iodide has an occupancy of 0.3.
[b] Theoretical f″ values correspond to element scattering at full occupancy, taken from tables prepared by the Biomolecular Structure Center [160].

Although the final *Dm*CLIC model was built using data collected by means of synchrotron radiation (λ = 0.98 Å) with anomalous scattering data collected concurrently, this wavelength produces negligible anomalous dispersion for most of the candidate metal ions (see Figure 2.1 and Figure 2.8). The strength of anomalous signal for sulfur atoms at an equivalent B-factor to K^+ is close to noise. For this reason *Dm*CLIC diffraction data that had previously been collected on a home source Cu K_α rotating anode X-ray machine (λ = 1.54 Å) was used to confirm the presence of an anomalous signal corresponding to the bound metal ion.

Using λ = 1.54 Å anomalous data, the atoms under scrutiny have a significantly stronger signal. Potassium has a theoretical anomalous scattering factor approximately twice as strong at 1.54 Å (f″ = 1.08e) compared to 0.98 Å (f″ = 0.46e), which is similar for sulfur atoms (f″ = 0.56e at 1.54 Å, f″ = 0.23e at 0.98 Å) and iodine (refer to Table 2.2). The increased strength of the anomalous signal is observed in Figure 2.8, where the 1.54 Å Bijvoet anomalous map is contoured at 4σ showing well ordered sulfur atoms and the metal binding site with corresponding anomalous signals well above background noise. Again the S^{δ} atom in Met68 has been chosen as the internal standard since it has been modelled with an atomic B-factor similar to the bound metal.

For λ = 1.54 Å anomalous data, the relative magnitude of Bijvoet map peaks corresponding to the bound metal (0.044 e/Å^3) and S^{δ}-Met68 (0.023 e/Å^3) is 1.9. The

relative integrated density at the position of the bound metal (1.28e) and S^{δ}-Met68 (0.67e) is 1.9. These values closely correspond to the theoretical anomalous scattering power of potassium (K, f″=1.08e) relative to sulfur (S, f″=0.56e), which is a value of 2.0. These measurements of the bound metal's anomalous scattering strength are consistent with the theoretical value expected for K^+ and compare more favourably in comparison to the metal's anomalous scattering strength calculated in the 0.98 Å anomalous map. This is probably a result of the higher signal to noise at 1.54 Å.

The structure of *Dm*CLIC solved using diffraction data collected from a Cu K_α rotating anode X-ray source has a variation in iodide assignment when compared with the final structure solved using synchrotron radiation. We note that in both structures the space group is identical ($P2_12_12_1$) and unit cell dimensions are similar (λ=1.54 Å, *a* = 39.4, *b* = 62.6, *c* = 114.1 compared with λ=0.98 Å, *a* = 39.3, *b* = 63.4, *c* = 115.0). However the electron density maps and Bijvoet anomalous maps show that the iodide ion, which has an occupancy of 0.9 in the synchrotron *Dm*CLIC structure, has a significantly lower occupancy, 0.3, in the Cu K_α structure. This is despite the iodide binding site residues adopting similar conformations in both structures. There is still a relatively weak anomalous signal at the position of this iodide ion (5.1σ) which supports the assignment of the iodide ion at a lower occupancy. There is also a low occupancy (0.3) iodide ion modelled in the synchrotron structure with a corresponding anomalous signal (4.3σ), which is not present in the Cu K_α structure.

Despite the differences in the two structures, they should still be satisfactory for structural comparisons particularly in regard to the anomalous signal at the metal binding site. Based on the results from SAD analysis, the metal's anomalous signal corresponds closely to a potassium ion and we can rule out the bound metal being a light atom. However, the anomalous data may not be sufficient to unambiguously distinguish between K^+ and Ca^{2+} so further methods were used to differentiate between these isoelectronic cations.

2.3.5.3 The bond valence method

The final technique used to clarify the identity of the bound metal ion is the bond valence method, which calculates the oxidation state of an atom and thus provides a clue to its identity. In our case, the power of this method lies in its ability to distinguish between the K^+ and Ca^{2+} isoelectronic cations which have a similar anomalous scattering power and electron density strength. Metal to oxygen (M-O) bond lengths have been measured in the binding site using the final synchrotron *Dm*CLIC model and are shown in Table 2.3. All ligands including water molecules are present in the model at full occupancy, as is the case in the EXC-4 structure (PDB:2VY9) which has also been considered in this analysis. It should be noted that although bond distances have been measured using the final *Dm*CLIC model that comprises a bound potassium ion at the metal binding site, substituting this for a calcium ion does not alter bond distances more than ~0.02 Å after refinement.

The calculation of the bond valence sum (BVS) is described in Section 2.2.6. For a K^+-O bond, the bond valence parameter is equal to d_0(KO)= 2.13 Å, which equates to an expected bond length of 2.79 Å for symmetrical octahedral coordination. The BVS for the bound metal ion in *Dm*CLIC after the summation over the six metal ligands gives a value of 1.14. This is close to the expected value of unity for singly ionised K^+ and strongly suggests the bound metal corresponds to K^+, rather than Ca^{2+}. This calculation was repeated for the EXC-4 structure which contains two assigned calcium ions per asymmetric unit. The BVS using the bond valence parameter of d_0(KO), gives a value of 1.07 for chain A and 1.25 for chain B, which again suggests the identity of the bound metal is indeed a potassium cation.

Table 2.3: Identification of the bound metal ion in *Dm*CLIC and EXC-4, using the bond valence method.

	Ligand		Bond Length (Å)		Bond Valence		CBVS	
	Residue	Atom						
*Dm*CLIC	His83	C=O	2.64		1.14		0.73	
	Pro85	C=O	2.51					
	Leu95	C=O	2.67					
	Asn97	O^{δ}	2.76					
	-	H_2O	3.19					
	-	H_2O	3.21					
			Chain A	Chain B	A	B	A	B
EXC-4	Gln77	C=O	2.88	2.79	1.07	1.25	0.69	0.81
	Pro79	C=O	2.66	2.57				
	Thr91	C=O	2.60	2.52				
	Asn93	O^{δ}	3.06	3.35				
	-	H_2O	2.94	2.73				
	-	H_2O	2.68	2.73				

In order to further support the results obtained using the BVS, we have repeated these calculations to coincide with the concept of the calcium bond-valence sum (CBVS) as outlined by Müller *et al*. [162]. This method is identical to the normal bond valence method except that the ion being investigated is assumed to be a calcium and thus the bond valence parameter is set to d_0(CaO) = 1.967 Å. This equates to bond lengths of 2.373 Å for Ca^{2+} with symmetrical octahedral coordination. Therefore in the case of the divalent cation Ca^{2+}, the CBVS value should correspond to a value of 2.0. In the case that there is a different cation, the CBVS value will differ and reflect a characteristic value consistent with that particular ligand. For example, the expected CBVS for some common cations are: NH^{4+} (0.51), K^{+} (0.64), Na^{+} (1.57), Ca^{2+} (2.00), Mn^{2+} (3.23), Zn^{2+} (4.07) and Mg^{2+} (4.19).

We find the *Dm*CLIC bound metal ion corresponds to a CBVS value of 0.73. Applying the CBVS method to EXC-4, the bound metal ions have similar values of 0.69 and 0.81. This is close to the expected value of 0.64 for a K^{+} ligand and clearly rules out Ca^{2+}.

Based on all of the above results presented in Section 2.3.5, we assign the observed metal ion to K^+ in both invertebrate CLIC structures: *Dm*CLIC and EXC-4.

2.3.6 Crystal packing interactions

The structure of *Dm*CLIC shows there are a number of significant packing interactions in the crystal. These interfaces have been analysed using PISA [170]. There are four unique interface contacts formed between the *Dm*CLIC structure and symmetry mates, making up a total interface surface area of 2,926 $Å^2$, which is equal to 23.3% of the total *Dm*CLIC surface area (Figure 2.9).

The *Dm*CLIC N-terminal extension fits into the interdomain cleft which contains the GSH binding site in CLIC1 and is in the vicinity of the metal binding site (Figure 2.9, interface is coloured blue). It should be noted however that no symmetry related interactions involve residues in the metal binding site and so the bound K^+ is unlikely to be effected by crystal packing. There is a key residue that stabilises this interface at Lys17, which forms five salt bridge bonds shared between Glu45, Glu101 and Glu177 in the symmetry neighbour. There are also four hydrogen bonds involving: Ser16-Glu126/Asn127, Asn90-Lys133 and Asn113-Arg76. The interface covers a total solvent accessible surface area of 1,583.1 $Å^2$, calculated by adding the buried surface area on each molecule.

The opposite side of the N-terminal tail also forms a small interface involving just four residues: Arg35 in one *Dm*CLIC molecule and Ser15, Ser16 and Asp19 in the symmetry molecule (Figure 2.9, interface is coloured green). A salt bridge is formed between Arg35 and Asp19, and a weak hydrogen bond is formed between Arg35 and Ser15. This interface covers a total of 128.0 $Å^2$ in solvent accessible surface area.

The next interface mainly involves residues in the C-terminal extension, His242-Phe250, and residues in the C-terminal domain helices 4a and 4b from a symmetry molecule related by translation only (Figure 2.9, interface is coloured orange). There is

only one hydrogen bond formed by the interface involving Glu244 and Asn127. The interface covers a total solvent accessible surface area of 520.4 Å^2.

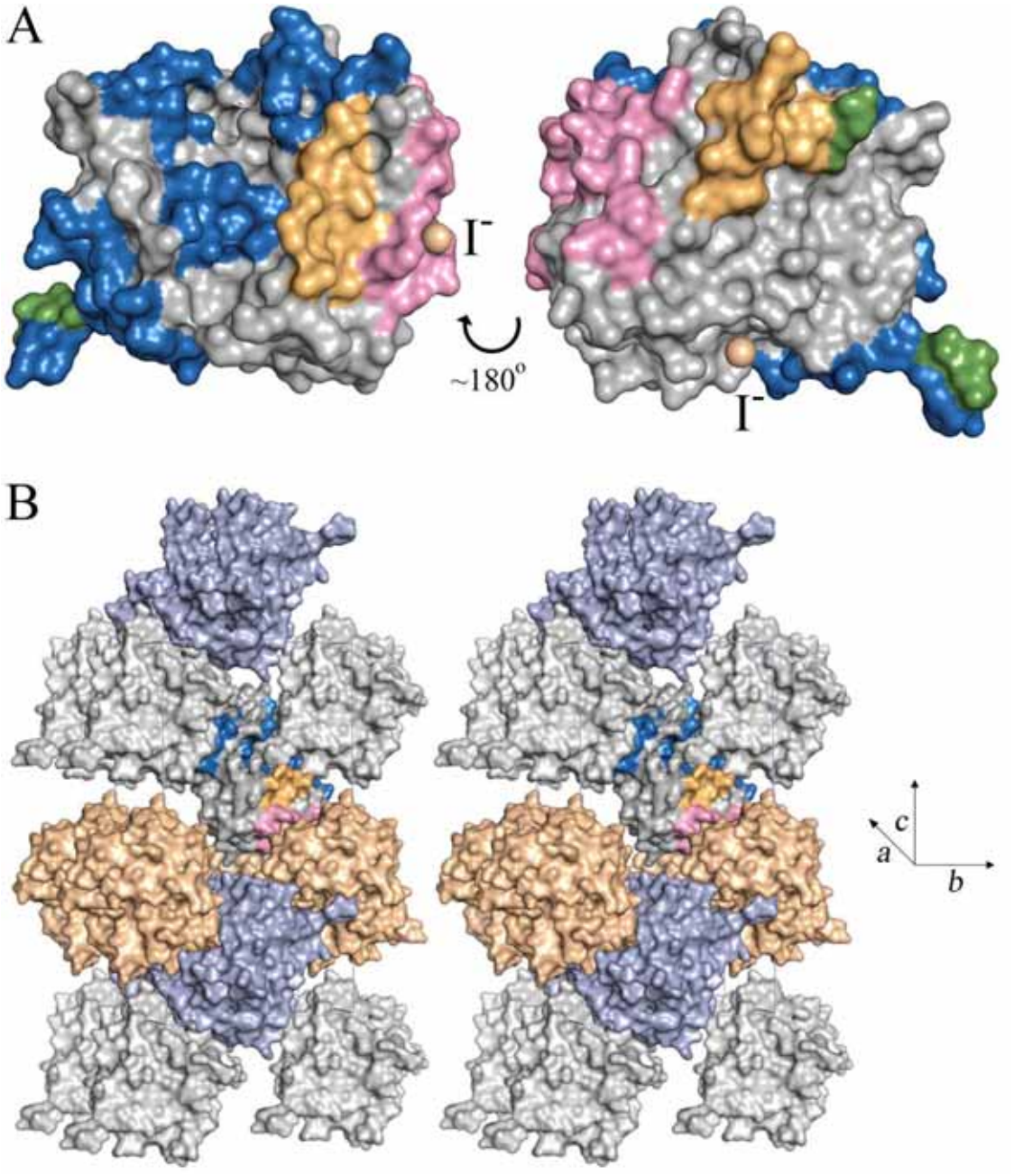

Figure 2.9: Crystal packing interactions in *Dm*CLIC.

A) Images of the *Dm*CLIC molecular surface are shown, related by a rotation of ~180° in the viewers y axis. The four symmetry contacts are denoted by surface colour. Iodide ions are also shown as spheres. Blue represents the packing of the N-terminus into the active site cleft, pink represents C-terminal domains packing over the iodide binding site, orange represents the C-terminal tail packing contacts and green represents the small packing interactions with the opposite side of the N-terminus.

B) A stereo image of the *Dm*CLIC packing diagram over the unit cell. Symmetry molecules related by translation are coloured alike. At the centre is a *Dm*CLIC molecule, coloured as in panel A. The orientation of the unit cell axes are shown on the right.

The last interface involves the packing of C-terminal domains and is shown in Figure 2.9, coloured in pink. This includes exposed residues in the amphipathic helix 5, the C-terminal tail and the C-terminus of helix 6. The interface includes three salt bridge bonds in total (two involve Lys153 and Asp194, the other involves Asp156 and Lys139) and four hydrogen bonds that all involve Arg152, bonding to Lys190, Tyr191 and Phe192. The interface covers a total solvent accessible surface area of 694.8 $Å^2$.

Although this last interface also encloses an iodide binding site, there are no symmetry related interactions with the iodide ion under 4 Å (see Figure 2.10, panel A). The iodide ion sits in a well defined surface depression in *Dm*CLIC near helix 5 and the connecting loop between helix 6 and helix 7. In Figure 2.10 panel B, I^- is shown as a sphere at half of its ionic radius (1.10 Å) for clarity. The ion makes favourable polar contacts with the nitrogen group of Arg152 and also interacts with two water molecules within 4 Å. Considering His200 is likely to be protonated in the crystal condition at pH 5.8, the proximity of the iodide ion and histidine is also likely to complement local electrostatics.

With all of the above interfaces, the accessible surface areas appear small except for the N-terminal extension which packs into the interdomain cleft of a neighbouring molecule like a ‘finger’. Hence this interface corresponds to a significantly higher buried surface area.

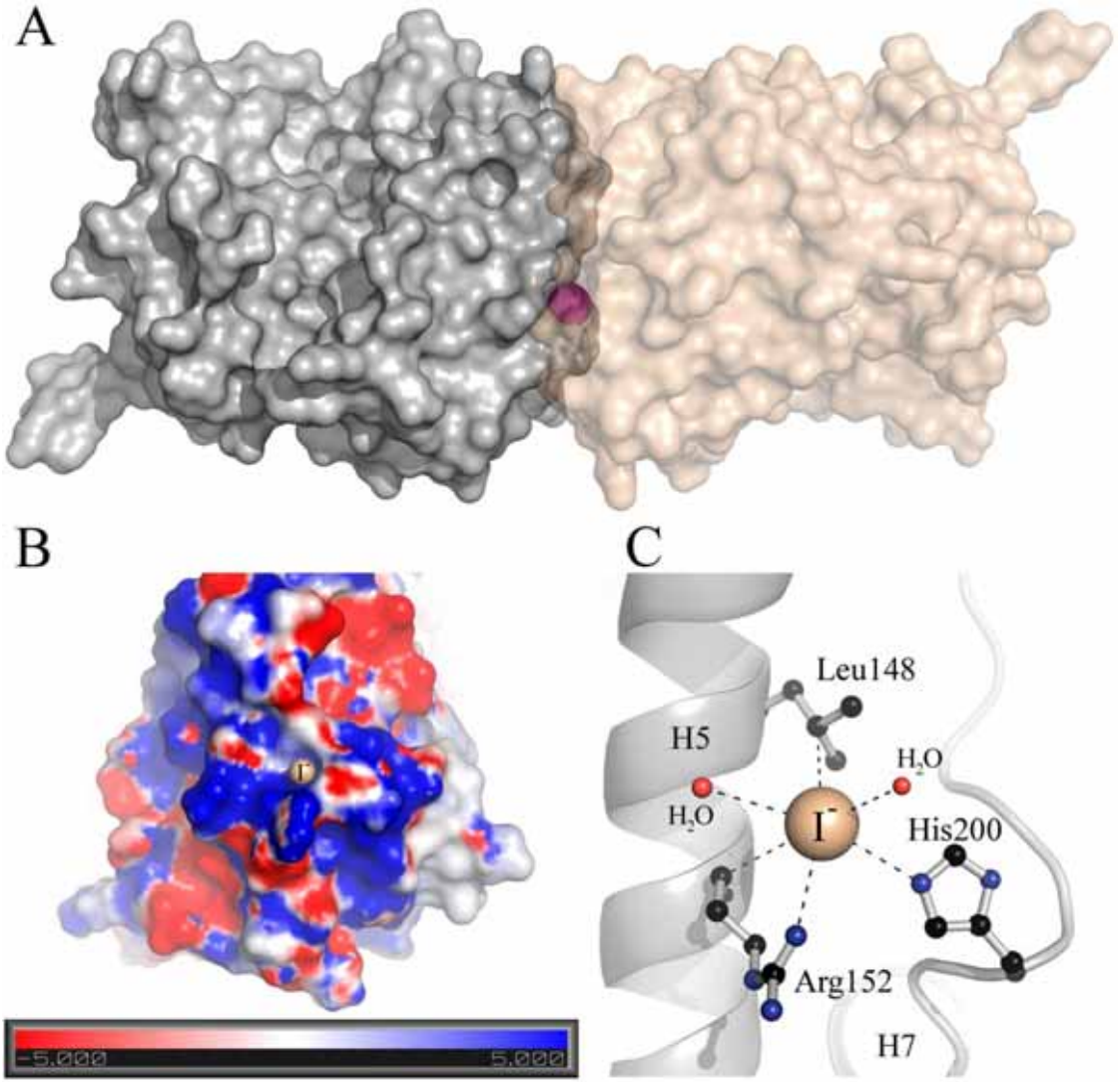

Figure 2.10: The iodide ion binding site in *Dm*CLIC

A) The iodide ion (magenta sphere) binds in a crystal packing interface. *Dm*CLIC is shown in surface representation (grey) while the symmetry related molecule is shown as a transparent surface (wheat).

B) The iodide ion (wheat) sits in a van Der Waals surface cavity formed by *Dm*CLIC and coloured by electrostatic potential mapped to the surface. The iodide ion is shown as a sphere with a radius (1.1 Å) equal to half its expected van Der Waals radius.

C) The binding site showing nearby residues in ball and stick representation. Atomic contacts within 4 Å to the iodide ion are shown as dashed lines. *Dm*CLIC is shown in cartoon representation (grey) with helix 5 (H5) and helix 7 (H7) labelled. Apart from a single water molecule (shown), no other iodide interactions are observed within 4 Å to the symmetry related molecule shown in A.

2.4 Discussion

The structure of the *D. melanogaster* CLIC protein, *Dm*CLIC, has been solved at a resolution of 1.80 Å displaying a structural fold that is very similar to other CLIC proteins, including members from vertebrate and invertebrate species. In comparison with the structure that we have previously published (PDB:2VY7), the structure discussed in this chapter has had a number of revisions including: the modelling of helix 2, the inclusion of residues 238-244 that form a short helix similar to that seen in EXC-4, the addition of an iodide ion with an occupancy of 30% and the amendment of the bound metal ion assignation as the monovalent K^+ cation rather than the divalent Ca^{2+} cation.

Consistent with previously solved CLIC structures, *Dm*CLIC adopts the characteristic GST structural fold with a thioredoxin-like N-terminal domain and all α-helical C-terminal domain. The high degree of similarity between *Dm*CLIC and other vertebrate CLICs could suggest a common cellular function, although this is yet to be confirmed. In the case of *Dm*CLIC and EXC-4, there is an overlap of function as shown by the rescue ability of *Dm*CLIC in *C. elegans exc-4* deleted mutants [134].

2.4.1 The *Dm*CLIC C-terminal extension

Electron density after helix 9 becomes weak beginning at Lys238, a position that is almost equivalent to the C-terminus of vertebrate CLICs 1-6. The weak electron density continues until Glu244, where density again becomes strong and unambiguous through until the last residue modelled, Ile258, which means there are only two residues missing at the C-terminus. In comparison with the shorter helix 9 and C-termini of vertebrate CLICs, in which the tail continues away from helix 9 in between helices 5 and 6, the C-termini of *Dm*CLIC and EXC-4 turns back towards helix 9 before running across the face of the C-terminal domain helices 6 and 7. This arrangement does not appear to be due to the constraints imposed by crystal packing as there is space in the *Dm*CLIC crystal for the C-terminal tail to follow the same path as seen in vertebrates CLICs. The extension of helix 9 in *Dm*CLIC and EXC-4 may

contribute to the deviation of the main chain conformation when compared to vertebrate CLICs.

The C-terminal extension has a very short α-helical turn from His242-Leu245 in what is otherwise a loosely structured loop region. A network of hydrogen bonds stabilises the interaction between the tail region and the body of *Dm*CLIC, and there are also crystal packing interactions involving residues His242 to Phe250. The effect of crystal packing may be significant since strong and unambiguous electron density returns at Leu245. A conserved tyrosine at the C-terminus of helix 9 (Tyr233 in CLIC1) marks the point where the C-terminal tail begins in vertebrate CLICs. After this point, there remains a degree of sequence conservation in CLICs 1-6 (but not CLIC3) where a short conserved motif YxxVAK is present. All solved vertebrate CLIC structures have the C-terminal tail in the same conformation, which includes the recently solved CLIC3 structure (PDB:3FY7). The conserved tyrosine residue in *Dm*CLIC corresponds to Tyr231, but this does not designate the beginning of the tail extension as helix 9 continues past this point for another four residues before leading into the unstructured tail segment. The invertebrate CLICs have a lower degree of conservation in the C-terminal extension, but also share a similar structural conformation.

*Dm*CLIC also has an extended N-terminal tail segment modelled in the final structure, similar to that seen in subunit A of the EXC-4 crystal dimer. The structural CLIC module has a well defined starting point at a conserved proline residue (Pro21 in *Dm*CLIC) that is present at the N-terminus of β-strand 1. There is a very low level of sequence conservation prior to this proline in the CLICs and it is likely that the N-terminus is flexible in solution. As has been shown, the N-terminus of *Dm*CLIC forms a significant packing interface by fitting into the interdomain cleft of a symmetry molecule. Hence the N-terminal extension in *Dm*CLIC is stabilised by crystal packing only, as is the case in EXC-4.

2.4.2 The potassium cation binding site

The presence of a bound metal in CLIC proteins is a novel feature that is found only in the invertebrate CLICs, *Dm*CLIC and EXC-4. The metal binding site is located at the N-terminus of α-helix 3 and β-strand 3 in a large interdomain cleft that also accommodates the putative CLIC redox active site involving the CxxC motif. The metal is coordinated by three main chain oxygen donors from His83, *cis*-Pro85, Leu95 and a side chain oxygen donor from the O^{δ} atom of Asn97. Together with two water molecules, these protein ligands form a coordination sphere with pseudo-trigonal bipyramidal symmetry. While the binding site is clearly electronegative, there are no nearby acidic residues that could compensate for the more positively charged cations.

In contrast to the earlier structures of *Dm*CLIC and EXC-4, which were modelled with a Ca^{2+} divalent cation in the metal binding site, the analysis presented in this chapter suggests a revised identification of the metal ion as the K^{+} monovalent cation. The earlier work focused on two main criteria to identify the bound metal: electron density and an anomalous signal. While there has been a more detailed approach to both these criteria described in this chapter, they are not sufficient to unambiguously distinguish between the isoelectronic Ca^{2+} and K^{+} cations, and so a further test was carried out in this current study which relates to the use of the bond valence method.

Anomalous diffraction data was analysed at both synchrotron (λ = 0.98 Å) and Cu K_{α} (λ = 1.54 Å) wavelengths by comparing the anomalous signal from the bound metal to an internal standard represented by the S^{δ} atom in Met68. This sulfur was chosen since it is modelled with an atomic B-factor similar to K^{+}. At the synchrotron and Cu K_{α} X-ray energies being used, we are far from the anomalous absorption edge of most metals and therefore can ignore the oxidation state of candidate metal ligands. We have estimated the anomalous scattering strength of the bound metal by comparing density peaks in the Bijvoet anomalous difference map corresponding to the metal and S^{δ}-Met68. Since this method may be sensitive to differences in B-factor we therefore also estimated the relative anomalous scattering strength by integrating the anomalous density centred on the metal and S^{δ}-Met68.

Using the λ=0.98 Å dataset we showed that the relative anomalous signal of the metal to sulfur (M/S) is 1.3 via the density peak method and 1.5 via the density integration method. Hence the values obtained using the two methods are comparable. This low-wavelength dataset is not optimal for anomalous scattering which is demonstrated by significant noise in the Bijvoet anomalous difference map at the level of sulfur atom scattering. Although the relative anomalous signal (1.3-1.5) theoretically corresponds to chlorine (Cl/S=1.3) or argon (Ar/S=1.6), the nearest metal ligands are potassium (K/S=2.0) and calcium (Ca/S=2.4), which are capable of binding in the *Dm*CLIC metal site. This information is only sufficient to identify the bound metal as a period 4 element that is likely to correspond to the weakest anomalous scatterer in the group, namely potassium. However due to the weak anomalous data obtained at this wavelength we are careful not to emphasise this result.

The stronger anomalous scattering data obtained at λ=1.54 Å is demonstrated in the Bijvoet anomalous difference map where ordered sulfurs and the bound metal have an anomalous signal well above background noise. Using this dataset there is excellent agreement between the anomalous density peak method and density integration method, for estimating the relative anomalous signal of metal to sulfur, which is 1.9 in both cases. The bound metal's anomalous signal now closely corresponds to potassium (K/S=2.0), tentatively differentiating the identity of the metal from calcium (Ca/S=2.4). Thus we surmise that the SAD data analysis supports the assignment of the bound metal as the K^+ cation.

An approach that is often used in the literature to support the identification of metal ligands in protein structures is the bond valence method, or equivalently the calcium bond valence sum CBVS as proposed by Müller *et al*. [162]. In order to use the CBVS, metal to oxygen ligand distances (M-O) must be accurately determined and it is also a requirement that the coordination sphere is complete without any missing ligands, such as water molecules with half occupancy. This latter necessity led Müller *et al*. to formulate the vector sum of bond valences or VECSUM concept to predict the completeness of a coordination sphere. Since we have closely examined the metal site in *Dm*CLIC and EXC-4 and concluded there are no missing ligands, the VECSUM is unnecessary for our analysis. We have ensured M-O distances in the *Dm*CLIC

structure have not been affected by anti-bumping restraints that may be used during refinement with REFMAC5 and are thus confident these distances are a true representation of the binding site.

The CBVS value for the metal cation in the *Dm*CLIC structure has been calculated as 0.73. The value correlates well with the expected value for a potassium ion of 0.64, easily distinguishable from the expected value of other plausible metal candidates including calcium (CBVS = 2.00). This calculation was repeated with the published structure of EXC-4 which contains two copies of the metal binding site in the asymmetric unit. In this case, the CBVS value is 0.69 and 0.81, again corresponding to the expected value of K^+.

The minimum required structure resolution stated by Müller *et al*. to apply the CBVS calculation accurately is between 1.5 Å and 1.8 Å, meaning our *Dm*CLIC structure solved at 1.8 Å is on the edge of this resolution limit. However, it is stated that the distinction between K^+ and Ca^{2+} is clear using the CBVS even towards the lower end of the resolution range. We have also analysed the published EXC-4 structure (PDB: 2VY9) where a metal ligand is also bound in the equivalent position to the metal ion in *Dm*CLIC, with an almost identical coordination sphere consisting of four protein ligands and two waters. The EXC-4 structure has been solved at 1.6 Å and is thus within the recommended resolution range.

Since we have previously eliminated other metal candidates using SAD analysis, the results obtained from the CBVS strongly suggest the metal bound to *Dm*CLIC and EXC-4 is a potassium cation. Furthermore, there is no evidence to suggest the metal binding site is heterogeneously occupied by other cations.

In the case of *Dm*CLIC, the protein has been exposed to potassium ions throughout purification, with the affinity column buffer (PBS) containing 2 mM KCl and the sizing column buffer containing 100 mM KCl. In addition, the crystallization solution consisted of KI at a concentration of 140 mM. Therefore, there was ample opportunity for *Dm*CLIC to scavenge potassium ions and maintain the bound

potassium site at full occupancy, particularly during crystallization where [K^+] approaches 240 mM.

The *Dm*CLIC crystal grew in the presence of a high concentration (100 mM) of the metal chelating agent, EDTA. The EDTA formation constant (log K_f) at 20 °C and 0.1 M ionic strength, is 0.8 for K^+ and 10.7 for Ca^{2+} [171], hence EDTA binds calcium significantly stronger than it does potassium. For the original structure modelled with a bound Ca^{2+} cation this poses a contradiction, since the binding site appears weak yet EDTA at a high concentration does not strip Ca^{2+} from the site. We have resolved the paradox by reassigning the metal ion as K^+, which clarifies the negligible effect EDTA has on metal binding in *Dm*CLIC. The presence of EDTA should prevent highly charged metal cations from binding *Dm*CLIC, if they were able, during crystallization. If this effect is occurring, it may not have a significant impact since EXC-4 binds K^+ in a similar manner to *Dm*CLIC but was not crystallized using EDTA in the crystal condition.

In the case of EXC-4, the crystallization solution included sodium formate (HCOONa) at 150 mM, thus crystals were grown in the presence of Na^+. EXC-4 was also exposed to potassium ions during size exclusion chromatography (100 mM KCl) and thus the crystallization solution also contained 100 mM K^+.

The role of the K^+ binding site in the invertebrate CLICs is not clear. It may serve to stabilise the local structural region, which may have a more global effect on the N-terminal domain in which it resides. Electrophysiology experiments have determined that *Dm*CLIC and EXC-4 have channel activity [137]. By chance, these experiments were performed in the presence of ~150 mM KCl and so most likely both of these proteins were in the potassium bound form when found to produce anion channels. If invertebrate CLICs require a large structural change to transition from the soluble form to integral membrane form, as expected for vertebrate CLICs, then it appears that the metal binding site does not interfere with this structural rearrangement.

2.4.3 The metal binding site in other CLICs

The metal binding site observed in the *Dm*CLIC structure is almost identical to the metal binding site in EXC-4, where the bonding carbonyl oxygens in the backbone adopt the same conformation despite sequence variations in the binding site residues (His83→Gln77, Leu95→Thr91). The carbonyl group of the conserved *cis*-proline (Pro85 in *Dm*CLIC) completes the backbone interactions with the metal. It is noted that an asparagine residue shared between *Dm*CLIC (Asn97) and EXC-4 (Asn93) appears to be required to complete the K^+ coordination sphere by facilitating a fourth oxygen donor through a side chain interaction with a bonding distance of 2.8 Å in *Dm*CLIC, and 3.1 Å or 3.4 Å in the EXC-4 subunits.

Comparing the invertebrate CLIC metal binding site with the equivalent site in vertebrate CLICs reveals a relatively conserved local structural region (see Figure 2.11). All human CLIC structures show the *cis*-proline residue at the N-terminus of β-strand 3 which adopts the same conformation as observed in the invertebrate CLICs, regardless of whether it is part of a double proline (CLIC2, CLIC4) or single proline (CLIC1, CLIC3) motif. Therefore they each satisfy the requirement of the *cis*-proline carbonyl ligand for an invertebrate-like metal binding site (see Figure 2.11). The backbone conformation of the preceding carbonyl ligand (His83 in *Dm*CLIC) is also conserved, although the equivalent CLIC1 residue Glu63 does not satisfy the required carbonyl orientation. A final backbone ligand is supplied by Leu95 in *Dm*CLIC that orientates the carbonyl group towards the metal binding site. A similar conformation is found in CLIC2 and CLIC3 for the equivalent residue. However, the structures of CLIC1 and CLIC4 show that peptide bond corresponding to Leu95-Glu96 in *Dm*CLIC has undergone a peptide-plane flip, which orientates the amide nitrogen towards the metal site. This is also observed in alternative structures of CLIC2. The only side chain oxygen ligand to the metal is supplied by an asparagine residue in *Dm*CLIC and EXC-4. The human CLICs cannot replicate this interaction since their equivalent residue corresponds to hydrophobic side chains (phenylalanine and valine) or the shorter polar side chain of threonine. The O^{γ} atom of threonine is orientated towards the metal site in CLIC1 and CLIC3, but probably cannot replicate the O^{δ}-asparagine interaction due an increased distance.

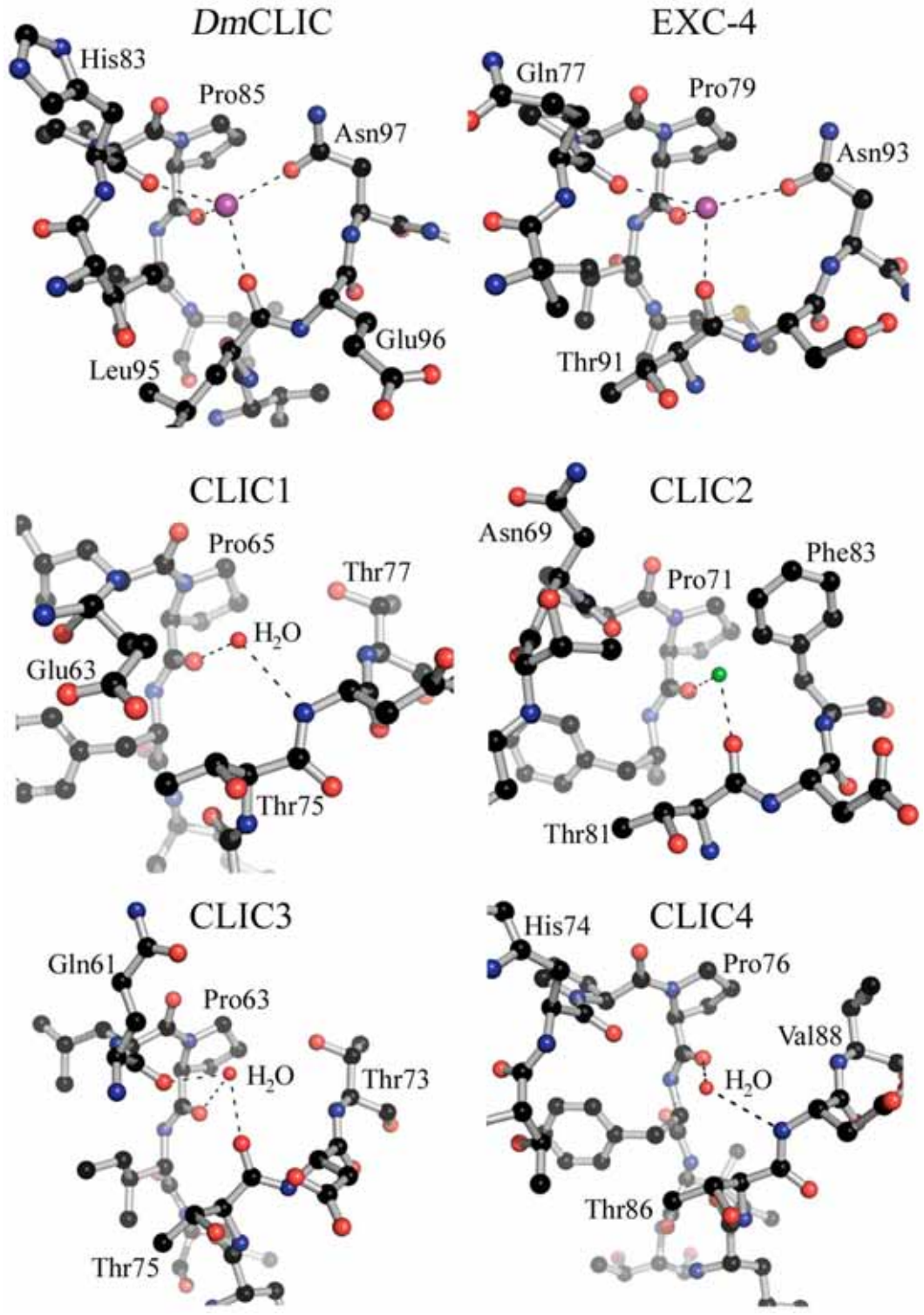

Figure 2.11: A comparison of the metal binding site in invertebrate and vertebrate CLICs.

Residues are shown in ball and stick representation. The K^+ cation is shown as a magenta sphere. Water molecules that occupy the metal site in other structures are labelled, and where the metal site is unoccupied a green sphere is shown in the equivalent position. Bonds are denoted by dashed lines. The structures shown correspond to CLIC1 (PDB:1K0M), CLIC2 (PDB:2R4V), CLIC3 (PDB:3FY7), CLIC4 (PDB:2AHE), EXC-4 (PDB:2YV9) and *Dm*CLIC presented in this chapter.

In all published structures of monomeric CLIC1, CLIC3 and CLIC4, the metal binding site is occupied by a water molecule which has been modelled with full occupancy. In the structure of dimeric CLIC1 [102], the cysteine residues Cys24 and Cys59 form a disulfide bond and CLIC1 undergoes a reversible structural transition. This structural rearrangement mainly affects the N-terminal thioredoxin domain and obliterates the putative metal binding site in CLIC1. The sequence of residues between helix 2 and helix 3 that form β-strands 3 and 4 as well as the metal binding site in monomeric CLIC1, transit to an unstructured loop in the dimeric form. The CLIC1 N-terminal extension separates metal binding residues by fitting between helix 2 and helix 3.

CLIC2 (PDB:2R4V, [112]) is the only CLIC structure that has an empty metal binding site, possibly the result of the large hydrophobic side chain of Phe83 blocking access to water molecules (Figure 2.11). This phenylalanine is in the equivalent position to Asn97 in *Dm*CLIC. It has been suggested that Phe83 will also block the γ-glutamyl portion of GSH binding to CLIC2 [112]. Otherwise the binding site is well formed with carbonyl oxygen donors from the three equivalent binding residues all oriented in the correct direction for metal binding. An alternative structure of CLIC2 (PDB:2R5G, [112]) shows there is a structural rearrangement to the backbone conformation near Phe83 corresponding to the Thr81-Asp82 peptide-plane flip, similar to that observed in CLIC1 and CLIC4 structures. The change to the backbone conformation permits Phe83 to adopt a rotamer that frees up the potential metal site allowing a water molecule to occupy it. The ability of this peptide bond to rotate may suggest that K^+ in *Dm*CLIC and EXC-4 stabilises the local structural region, as it seems reasonable that all CLICs can undergo this peptide-plane flip.

A structure of CLIC2 solved by an independent group is also missing a water molecule in the equivalent metal binding site (PDB:2PER, [172]). The site is again well formed although Phe83 now adopts an all together different rotamer that does not obscure the region and should not prevent a water molecule occupying the site. However, after inspecting the corresponding electron density maps reconstructed using structure factors deposited in the PDB it is clear that the binding site and surrounding

residue side chains have been modelled in weak density that may lead to the modelling of ambiguous side chain conformations.

The protein purification and crystallization protocols described in the literature reveal that CLIC2 and CLIC3 have not been exposed to K^+ ions during preparation; however both have been exposed to Na^+ ions in the crystallization solution [112]. CLIC1 was crystallized in a solution consisting of Na^+ ions [87], and CLIC4 was crystallized in a solution of NH_4^+ [86]. Only CLIC1 and CLIC4 have been exposed to K^+ ions during purification (SEC). This would seem to provide ample opportunity for these proteins to bind K^+, but the crystal structures of CLIC1 and CLIC4 do not contain a bound cation. This information may suggest the vertebrate binding site is not capable of recognising metal ligands, not withstanding the possibility that a structural rearrangement may be required for K^+ binding. In the case of CLIC2 and CLIC3, where the binding site is preformed and closely resembles the coordination sphere in *Dm*CLIC, the lack of exposure to K^+ may be the reason there is no bound metal in the crystal structures.

2.4.4 The CxxC motif in *Dm*CLIC

*Dm*CLIC contains two glutaredoxin-like CxxC sequence motifs: a conserved ^{30}CLPC^{33} motif at the N-terminus of helix 1 and a unique ^{172}CCFDC^{176} motif at the N-terminus of helix 6. The unique motif at helix 6 contains the N-terminal capping residue Cys172 and the highly conserved Cys176. However the sulfhydryl groups of the three cysteines are orientated away from each other, suggesting the site is not redox active. But while the site may not contain active cysteines, it has been shown that the structural integrity of helix 6 in the related GST proteins is affected by the N-cap residue and surrounding motif, particularly since its position corresponds closely to the GST-motif II region formed by helix 6 and the preceding loop [111]. The motif is conserved in GST proteins with a sequence that typically corresponds to Gxxh(T/S)xxDh, where x is any residue and h is hydrophobic. If *Dm*CLIC is regulated by redox conditions than the helix 6 CxxC motif may lead to structural changes in the C-terminal domain unique within CLIC-like proteins.

The other CxxC motif in *Dm*CLIC is mostly conserved across vertebrate and invertebrate CLICs and forms the putative GSH binding site in CLIC1. In structurally related GST proteins, this region is equivalent to the G-site that confers thioltransferase activity and the recognition of GSH at this site is integral to their main functional role as detoxifying agents in the cell. It has been noted that the pK_a value of these CxxC cysteine residues in proteins that contain a thioredoxin fold is unusually low [173, 174]. In vertebrate CLICs, a conserved proline residue follows the first cysteine in the CxxC motif, while in the invertebrate CLICs this proline is replaced by a leucine. Interestingly, it has been shown using synthetic peptides that a proline residue acts to lower the pK_a of a preceding cysteine at the N-terminus of a helix, along with the decrease of pK_a associated with a charge-dipole interaction [175]. The proline to leucine replacement however does not seem to instigate any significant changes to the main chain conformation as seen in crystal structures.

The reducing environment of the *Dm*CLIC crystal ensures that Cys30 and Cys33 do not form a disulfide bridge which has been observed in the structures of CLIC2 and CLIC3. It is reasonable to assume that under a favourable oxidative environment *Dm*CLIC will also possess this intramolecular disulfide bond.

The proximity of the putative CLIC GSH binding site to the metal binding site suggests the K^+ cation may be a cofactor or facilitate an allosteric interaction if indeed GSH binds in a manner similar to that observed in CLIC1 and the G-site of GSTs [87, 89]. The GSH sulfhydryl group forms a disulfide bond to the conserved CxxS cysteine (Cys24) in CLIC1, while the GSTs normally do not contain an equivalent cysteine and thus the sulfhydryl is not disulfide linked. There are other notable interactions involving GSH and GSTs in the conserved double proline loop connecting helix 2 and β-strand 3. This loop contains the invariant *cis*-proline that allows a suitable conformation for the preceding residue to form backbone hydrogen bonding to the cysteinyl amide and carbonyl groups of GSH. However, in the CLIC1:GSH complex, only weak hydrogen bonding to the amide group is observed due to deviations in the main chain moving the preceding residue away from the GSH molecule. The presence of a preceding proline in *Dm*CLIC and CLICs 2, 4, 5 and 6, makes backbone hydrogen bonding to the cysteinyl carbonyl impossible. However in *Dm*CLIC a histidine residue

at His83 could hydrogen bond the carbonyl group of GSH through the N^{δ} side chain atom (Figure 2.12). Other CLICs may also be able make up for the lost hydrogen bond by bonding GSH through a side chain interaction involving the equivalent His83 residue (histidine in CLIC4 and CLIC5, asparagine in CLIC2 and CLIC6, glutamine in CLIC1 and CLIC3). Such an interaction could well have been missed in the CLIC1:GSH complex due to a PCR mutation error (Q36E) in CLIC1 [87].

The superposition of the GSH molecule from the CLIC1:GSH complex onto *Dm*CLIC shows steric interference between the γ-glutamyl moiety and Asn97 (see Figure 2.12). In CLIC1, Asn97 corresponds to Thr77, which forms an amide backbone hydrogen bond to a GSH carbonyl group in the γ-glutamyl moiety. This could suggest that *Dm*CLIC GSH binding and K^+ binding cannot occur concurrently, or at least K^+ may act as an allosteric switch that works through Asn97. A similar mechanism could be utilised by a shared interaction with His83 between GSH and K^+.

Proteins which use K^+ as a cofactor or allosteric effector are defined as either K^+-activated type I or II enzymes respectively [176]. In the case of *Dm*CLIC, K^+ could act as a cofactor with substantial changes to the mode of GSH binding in order to bring GSH closer to the metal site. This could occur if the carbonyl groups of the GSH γ-glutamyl moiety could be repositioned to replace the two K^+ water ligands, a rearrangement that would only require a 10° rotation over the long axis of GSH towards the metal binding site. Such a rotation by GSH should be able to maintain a disulfide bond to Cys40 and eliminate steric hindrance from Asn97. While there are surprisingly few obstacles to this rearrangement, this action would break possible hydrogen bonding between the backbone carbonyl of Pro84 and the cysteinyl amide group. Although it appears more plausible that K^+ will have an allosteric function, future studies will need to elucidate whether *Dm*CLIC can be classified as a K^+-activated type I or type II enzyme.

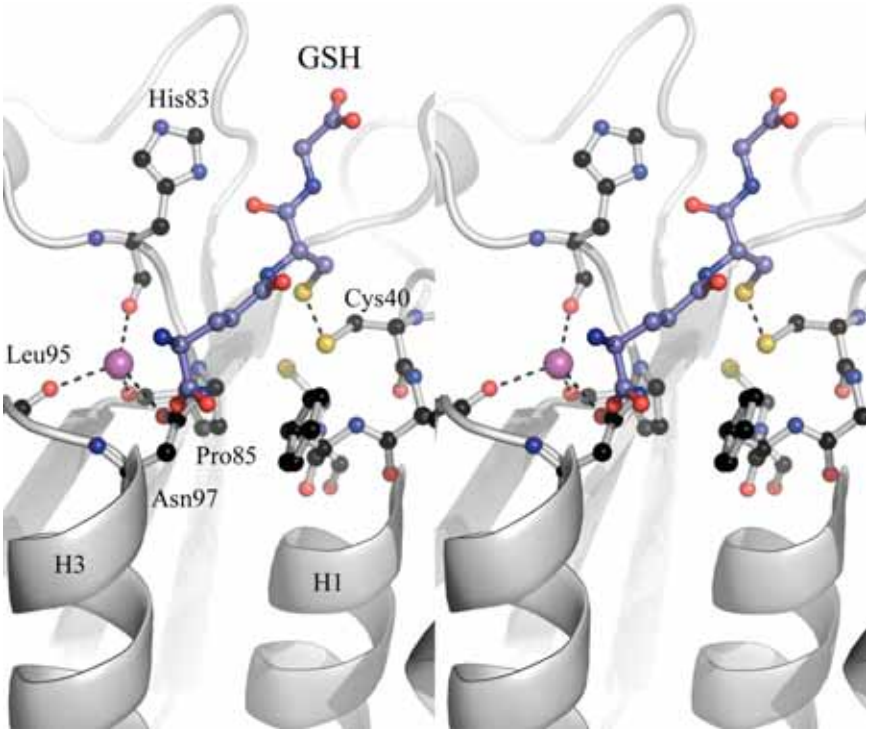

Figure 2.12: Stereo superposition of GSH with the *Dm*CLIC structure.

The GSH molecule bound to CLIC1 (PDB:1K0N), is superimposed onto our *Dm*CLIC model after an alignment of the two protein structures. The K^+ cation is shown as a magenta sphere. Metal ligand residues, the CxxC motif and GSH are all shown in ball and stick representation. *Dm*CLIC is shown in cartoon representation (grey) and GSH carbon atoms are light blue.

2.5 Conclusion

In this chapter we have presented a structure of the soluble form of *Dm*CLIC, a CLIC-like protein from the model organism *D. melanogaster*. *Dm*CLIC possesses the canonical GST fold that is similar to other known CLIC structures from invertebrate (EXC-4) and vertebrate (CLICs 1-4) species. The structure presented here is a revised version of the model that has previously been published [137]. It includes additional loop regions near helix 2 and helix 9, but most importantly the metal binding site at the N-terminus of β-strand 3 and α-helix 3 has been analysed using single wavelength anomalous dispersion data and the bond valence method. Through these methods we have identified the bound metal as the K^+ cation rather than the isoelectronic Ca^{2+} cation. This metal assignment clears up the inherent contradiction posed by the presence of Ca^{2+} in the original *Dm*CLIC structure, where a high concentration of the metal chelating agent, EDTA, in the crystallization buffer does not appear to affect Ca^{2+} binding. The abundance of K^+ ions during purification and crystallization fits in with the assignation. We also show through similar considerations that the metal binding site in EXC-4 is also likely to correspond to a homogeneously bound K^+ cation rather than Ca^{2+}.

The possible function of the K^+ cation in invertebrate CLICs is a not clear. The location of the binding site partially overlaps the GSH binding site in CLIC1 and notably if GSH were to bind *Dm*CLIC, it would sterically interfere with residues that participate in K^+ binding. Although the expected configuration of GSH suggests it will have an allosteric effect on K^+ binding, small changes to the geometry of GSH could impose a direct interaction between the two indicating it may function as a cofactor.

These findings appear to have opened up an excellent avenue for further study. The presence of K^+ cations and the effect they may have on invertebrate structures could be investigated by depriving the proteins of abundant potassium ions by a simple modification to purification protocols. It is likely that the substitution of NaCl for KCl in purification buffers is all that is required to observe any changes, although the absence of K^+ may necessitate a new crystallization condition. Whether or not the

preformed metal binding site can bind K^+ in other CLICs should also be investigated. The CLIC2 and CLIC3 structures that contain an intramolecular disulfide bridge in the CxxC motif near the metal binding site have not been exposed to K^+ ions during crystallization. CLIC1 which also forms a disulfide bridge upon oxidation, experiences a structural transition that destroys its putative metal binding site. There is the possibility that oxidative environments will alter metal binding in CLICs in different ways and this could be explored in both vertebrate and invertebrates CLICs.

The presence of K^+ during electrophysiology experiments also needs to be considered since its presence in the N-terminal domain could feasibly have an effect on the required structural rearrangement before membrane insertion. Previous tip-dip and chloride efflux experiments have measured *Dm*CLIC and EXC-4 channel activity in the presence of KCl [137]. Further experiments will need to examine any changes to the behaviour of invertebrate CLIC channels in the absence of K^+.

Chapter 3

Preliminary importin-α structures

3.1 Introduction

In this chapter we will describe a number of crystal structures of importin-α that were solved prior to the importin-α:NLS complexes presented in Chapters 4 and 5. We examine here the early difficulties involved in obtaining unambiguous importin-α:NLS complexes due to the presence of a rogue ligand that was later identified as the endogenous *E. coli* 30S ribosomal subunit S21. In these initial experiments, we closely followed importin-α purification methods that are routinely reported in the literature and have previously been used to obtain co-crystal structures of importin-α:NLS peptide complexes [50, 75, 177, 178]. By examining electron density maps of importin-α co-crystallized with an SV40 NLS or CLIC4 NLS peptide, we found that density in the major binding site corresponds to a bound peptide that lacks well defined side chain features. For this reason, electron density in the binding site cannot be unambiguously attributed to the co-crystallized peptide.

The recombinant mouse importin-α (70-529) clone used here is a truncated version that lacks the N-terminal auto-inhibitory IBB domain (residues 1-69) and hence an importin-α (70-529) crystal structure is not expected to contain a bound peptide corresponding to the auto-inhibitory domain NLS-like motif. However, after solving the structure of importin-α (70-529) without co-crystallization of NLS peptides, electron density corresponding to an unknown bound peptide in the importin-α (70-529) major binding site was unexpectedly observed. The density does not unambiguously define residue side chains, however a preliminary sequence motif of ^{P1}AKAA(K/R)A^{P6} could be ascertained, where 'A' corresponds to an alanine or longer disordered side chain.

After noting a matching sequence motif, AKFERQ, in the artificial purification tag at the N-terminus of the importin-α experimental construct, a double point mutant

was cloned to obliterate the putative NLS-like sequence. The importin-α (K25E, R28G) mutant was crystallized and the structure determined. It was also found to contain electron density that corresponds to a bound peptide. We are thus able to rule out the possibility that the NLS-like motif in importin-α (70-529) forms an intrasteric interaction with the binding site.

These preliminary experiments serve to focus our attention on attaining a ligand free importin-α structure (*apo* importin-α (70-529)) that would be necessary to unambiguously determine importin-α:NLS structures.

3.2 Methods

3.2.1 Cloning

The original clone of mouse importin-α (70-529) was received as a generous donation from Bostjan Kobe[‡] (see reference [178] for clone details). DNA sequencing was used to confirm the identity of the clone, which corresponds to a truncated version of importin-α (*Mus musculus*, isoform α2 [179]) which includes residues 70-529 inserted into the pET-30a N-terminal His-tagged expression vector (Novagen). The cloning and expression region includes: a hexa-histidine affinity tag, a thrombin cleavage site, an affinity S-tag, an enterokinase cleavage site and restriction enzyme sites. Including this 50 residue long artificial N-terminal region and the 460 residues of importin-α (residues 70-529), the full length expression product consists of 510 amino acids with a molecular weight of 55.3 kDa and an isoelectric point (pI) of 5.1. The N-terminal region is referred to in this thesis as the *linker* region, signifying its position in the sequence where it links the hexa-histidine tag to importin-α. The *linker* consists of the following amino acid sequence:

```
                    10        20        30        40        50
MHHHHHHSSG LVPRGSGMKE TAAAKFERQH MDSPDLGTDD DDKAMADIGS
His-tag   Thrombin  S-tag               Enterokinase
```

3.2.1.1 The importin-α double-point mutant

A double point mutation construct was created using the Quikchange™ site directed mutagenesis kit (Stratagene). The standard manufacturer's protocol was followed using the importin-α (70-529) plasmid as a template. The mutagenesis primer 5′- CC GCT GCT GCT **G**AA TTC GAA **G**GC CAG CAC A TG G -3′ and its complimentary pair were designed and used to substitute a lysine at position 25 and an arginine residue at position 28 in the *linker* region, to glutamic acid and glycine respectively.

[‡] School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia QLD 4072

This double mutation (K25E, R28G) serves to remove a weak NLS consensus sequence in the truncation construct located towards the middle of the *linker*:

^{1}MHHHHHH – *linker* – ^{24}A**K**FE**R**Q^{29} to ^{24}A**E**FE**G**Q^{29} – *linker* – ^{51}importin- α (70-529)

3.2.1.2 The importin-α deletion mutant

A deletion mutation was performed to remove 40 extraneous residues in the N-terminal *linker* region. The Phusion mutagenesis kit (New England Biolabs) was used and the standard protocol followed as per the manufacturer's instructions. Primers were designed to border the deletion area which corresponds to 120 bp between the His-tag and the beginning of importin-α (70-529). The 5′ phosphorylated forward and reverse primers (5′-GGA TCC AAC CAG GGT ACT GTA AAT-3′ and 5′ –AGA ATG ATG ATG ATG ATG GTG CAT- 3′ respectively) were used to PCR amplify the fragment using the pET-30a truncated importin-α plasmid as a template. The PCR product was checked on an agarose gel prior to ligation, after which the ligation product was transformed into chemically competent BL21(DE3) *E. coli* cells.

The final purified plasmid was checked on an agarose gel and the mutation was confirmed by DNA sequencing. The importin-α (70-529) deletion construct corresponds to:

^{1}MHHHHHH – ^{8}SGS – ^{11}importin-α (70-529)

3.2.2 NLS peptide synthesis

All peptides used in these experiments were synthesised by Sigma-Genosys as custom peptides (Table 3.1). The peptides contain a free amide N-terminus group (NH_2), and a carboxyl terminating group (COOH) at the C-terminus. CLIC4 peptide sequences were designed to contain residues flanking the NLS that increase solubility, while the SV40 NLS peptide sequence is identical to that used in previous studies of

importin-α:SV40 complexes [178]. Impurities resulting from peptide synthesis possibly include: deletion and truncation sequences, trifluoroacetic acid, acetic acid and peptides that have experienced side chain reactions. Characteristics of each peptide used are summarised in Table 3.1.

Lyophilised peptides were first dissolved in water or a suitable assay buffer before use. The stock concentration of each peptide sample varied between 1-3 mg/ml for different crystallization trials.

Table 3.1: The sequences and properties of three synthetic NLS peptides used in this study.

Source	Sequence	Purity[a]	MW[b]	pI[b]
SV40	NH_2-^{126}PKKKRKV^{132}-COOH	>50%	883.1 Da	11.3
CLIC4	NH_2-^{58}DLKRKPAD^{65}-COOH	>50%	942.1 Da	9.9
CLIC4	NH_2-^{198}VKVVAKKYRN^{207}-COOH	>95%	1204.5 Da	10.9

[a] Peptide purity has been assessed by analytical RP-HPLC.
[b] The molecular weight (MW) and isoelectric point (pI) have been theoretically calculated.

3.2.3 Importin-α (70-529) expression

BL21(DE3) cells containing mouse importin-α (70-529) (or importin-α *linker* mutants) in the pET-30a expression vector were grown overnight in 10 mL volumes of Luria-Bertani (LB) media (30 μg/mL kanamycin), at 37 °C with shaking at 200 rpm. Flasks of 400 mL LB media (30 μg/mL kanamycin) were then inoculated with 10 mL of the overnight culture. The cells (2 L) were grown at 37 °C until an optical density (O.D.) at 600 nm of 1.0 cm^{-1} was reached, at which point the culture was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was then lowered to 30 °C and growth continued for approximately 5 hours. Cells were harvested by centrifugation at 8,000 rpm (11,000 *g*) for 8 minutes in a Sorvall SLA-3000 fixed angle rotor and were then resuspended in 35 mL PBS and frozen at -80°C.

The frozen bacterial sample was thawed on ice and 2 mM dithiothreitol (DTT) was added. At this stage protease inhibitors were also added (*Complete protease*

inhibitor cocktail, Roche Applied Science). The cells were then passed through a French pressure cell two to three times using a 'high' ratio gauge pressure reading of 1,000 psi, corresponding to a working cell pressure of 16,000 psi. The resulting cell debris was centrifuged at 18,000 rpm (39,000 *g*) for 45 minutes at 4 °C in a Sorvall SS-34 fixed angle rotor and the supernatant containing soluble protein kept on ice until purification.

3.2.4 Importin-α (70-529) purification

The supernatant was added to a 3 mL solution of either Ni-NTA agarose resin (Qiagen) or Profinity IMAC resin (Bio-Rad), and was mixed under gentle rotation for one hour. The resin was then loaded into an Econopak gravity flow column (Bio-Rad) and unbound protein was eluted and discarded. Resin was washed with 200 mL of 20 mM HEPES pH 7.0, 500 mM NaCl, 1mM $MgCl_2$, 0.1 mM DTT and up to 20 mM imidazole. Bound protein was then step eluted with 10 mL of a buffer containing 20 mM HEPES pH 7.0, 500 mM NaCl, 0.1 mM DTT and 150 mM imidazole.

In addition to step elution of the importin-α (70-529) after binding to IMAC resin, an imidazole gradient was at times employed for greater control over protein elution and higher purity. A HisTrap^{TM} FF column (GE Healthcare) with 1 mL volume and prepacked with Ni Sepharose^{TM} 6 Fast Flow resin was first pre-equilibrated with a buffer consisting of 20 mM Tris pH 8.0 and 500 mM NaCl before binding, and then washed with 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM $MgCl_2$ and 20 mM imidazole. The protein was eluted over an imidazole gradient from 20 mM to 250 mM at a flow rate of 0.5 mL/min over a volume of 40 mL.

Fractions with the majority of protein as assessed by UV absorption at 280 nm (A_{280}) were then pooled and dialysed into 20 mM Tris pH 8.0, 100 mM NaCl and 2 mM DTT. Figure 3.1 shows a typical chromatogram of an imidazole gradient used in preparing importin-α, along with SDS-PAGE analysis. The purification protocol described here follows the methods reported in the literature [50, 75, 177, 178] and does not include size exclusion chromatography (SEC). We therefore refer to the purified protein used in this chapter as '(no-SEC) importin-α (70-529)'.

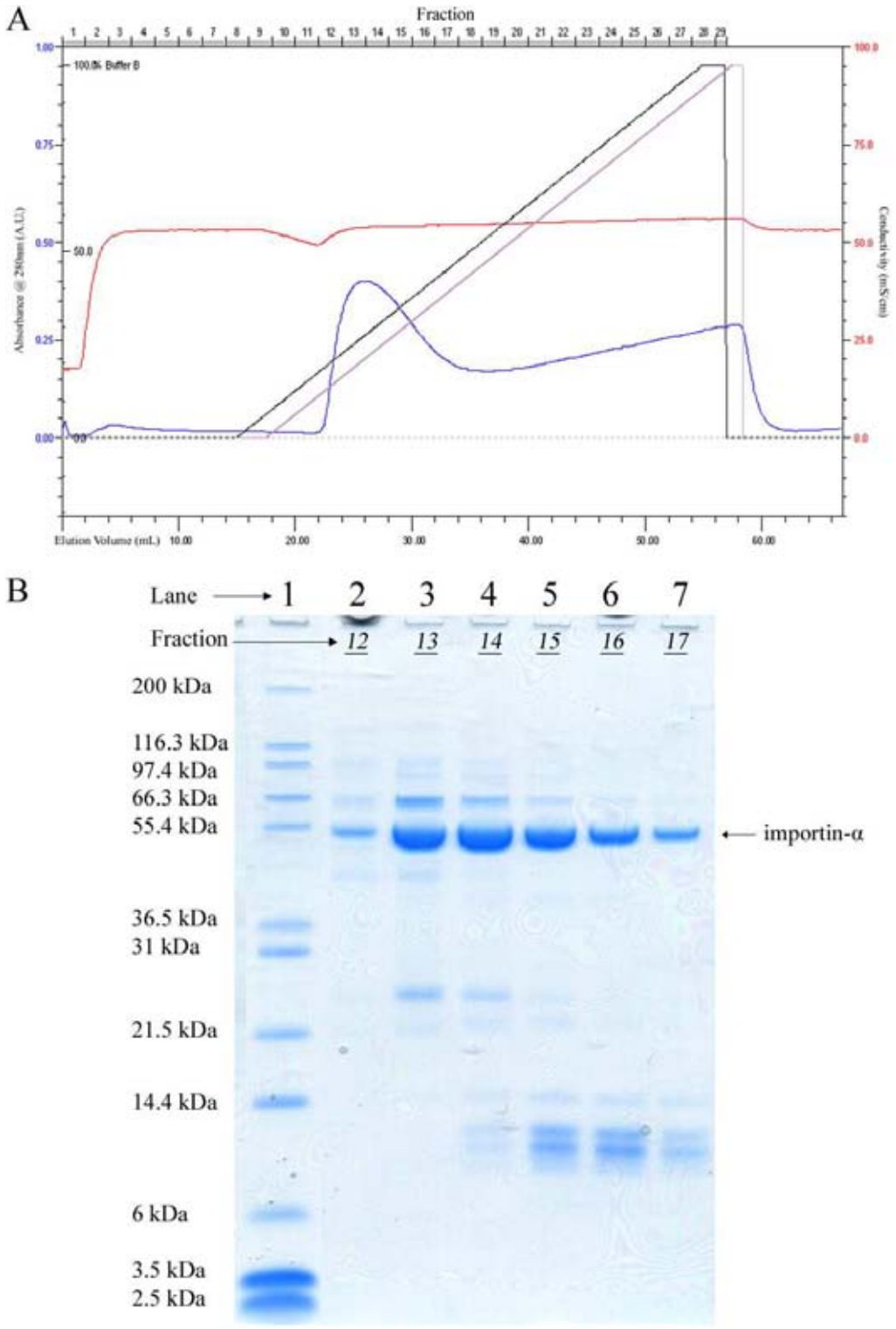

Figure 3.1: IMAC imidazole gradient elution of importin-α (70-529) and SDS-PAGE analysis.

A) The blue curve represents A_{280}, the red curve represents conductivity (mS/cm), the black line represents the pump gradient and the purple curve represents the estimated imidazole gradient from 0% (20 mM) to 100% (250 mM). The gradually increasing tail of the A_{280} curve is due to the absorption properties of imidazole.

B) SDS-PAGE of samples collected from the imidazole gradient in panel A. Lane 1 contains the Mark12™ unstained protein standard (Invitrogen), while lanes 2-7 contain fractions 12-17.

3.2.5 SDS-PAGE

Protein samples were analysed using denaturing Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE), or non-denaturing (native) PAGE. For SDS-PAGE, the protein was first diluted into a 2x loading buffer consisting of: 20% glycerol (v/v), 2% SDS, 0.02% bromophenol blue and 90 mM Tris-HCl, pH 6.8. A reducing agent (DTT) was added to the final solution to 2 mM. The running buffer was prepared by diluting 50 mL of 20x NuPAGE® MES in 950 mL of Milli-Q purified H_2O (MQ-H_2O) which was used to fill both the inner and outer chamber.

The samples were loaded onto a NuPAGE® Novex® 10% Bis-Tris gel with 1.00 mm thickness and 12 wells (Invitrogen). The voltage was set to a constant 200 V and the gel was run for approximately 45 minutes. The Mark12™ unstained protein standard (Invitrogen) was used to estimate the apparent MW on the gel. Gels were stained for approximately 30 minutes with a buffer consisting of 0.2% Coomassie G-250, 10% acetic acid and 30% ethanol. Destaining was performed overnight in a buffer consisting of 10% acetic acid and 5% ethanol.

For blue native-PAGE (BN-PAGE), the protein was added to 2.5 μL of 4x NativePAGE™ Sample Buffer which was made up to 10 μL using MQ-H_2O. The sample buffer was not supplemented with Coomassie G-250. The anode buffer was prepared by diluting 50 mL of 20x NativePAGE™ Running Buffer in 950 mL MQ-H_2O. The cathode buffer (containing Coomassie G-250) was prepared by adding 50 mL of 20x NativePAGE™ Running Buffer to 50 mL of 20x NativePAGE™ Cathode Additive, which was made up to 1 L using MQ-H_2O. The samples were loaded onto a NativePAGE™ 4-16% Bis-Tris gel with 1.00 mm thickness and 15 wells (Invitrogen). The voltage was set to a constant 150 V and the gel was run for approximately 110 minutes. The gel was stained using the same protocol described above for SDS-PAGE.

3.2.6 Crystallization and data collection

Crystallization trials of importin-α (70-529) were performed using a starting condition determined by Teh, T. *et al*. 1999 [177], which consisted of 0.8 M sodium citrate, 10 mM DTT and 100 mM HEPES at pH 6.0. Each of our importin-α (70-529) crystals were optimised by systematic grid screening around this original condition. We found that at a higher pH of 7.0 and a lower sodium citrate concentration of 0.7 M, both crystal and diffraction quality improved.

All crystals were grown using the hanging drop vapour diffusion technique in 24 well plates with 3.5 mL well capacity (Hampton) and siliconised square glass cover slides (Hampton). In all cases the well was filled with a 500 mL reservoir solution that also corresponds to the crystallization solution.

Co-crystallization of (no-SEC) importin-α (70-529) + SV40/CLIC4 NLS peptides

Co-crystals grew in a condition consisting of 0.6 - 0.7 M sodium citrate, 80 mM HEPES at pH 6.5 and 10 mM DTT. Drops consisted of 1 μL of (no-SEC) importin-α (70-529) at 20 mg/mL, 2 μL of either the SV40 NLS peptide or CLIC4 NLS peptide at 1 mg/mL (equivalent to a 6.3 x molar excess or 4.6 x molar excess, respectively), and 3 μL of reservoir.

No-SEC importin-α (70-529)

Crystals grew in 0.7 M sodium citrate, 80 mM HEPES at pH 6.5 and 10 mM DTT. Drops consisted of 1 μL of (no-SEC) importin-α (70-529) at 20 mg/mL, and 3 μL of reservoir solution.

Importin-α (70-529) double point mutant (K25E, R28G)

Crystals grew in a condition consisting of 0.7 M sodium citrate, 70 mM HEPES at pH 7.2 and 10 mM DTT. Drops consisted of 3 μL of the (no-SEC) importin-α (70-529)

double point mutant at 23 mg/mL and 3 μL of reservoir. Crystals did not nucleate as frequently compared to *apo* importin-α with brown precipitate forming in some wells.

General crystal handling and data collection

In all cases, crystals were highly susceptible to cracking and physical degradation upon transfer to a cryoprotectant, which was not helped by a heavy skin which consistently developed on the surface of the crystal drop. To minimise degradation as much as possible, crystals were gradually transferred to the cryoprotectant solution (consisting of the reservoir solution supplemented with 20% glycerol), in a 3 step process. Using a cryoloop, the first step was to move the crystal to a fresh reservoir solution and allow it to stand for one minute. The crystal was then moved to a 50%/50% reservoir/cryoprotectant solution where it was allowed to stand again for one minute. The crystal was then finally moved into the 100% cryoprotectant solution.

Crystals were either directly flash cooled in a nitrogen cryostream at 100 K or flash cooled in liquid nitrogen and stored in a cryogenic dry shipping dewar, before data collection. Auto-indexing and initial processing was performed using MOSFLM [180], and the rotation range for each dataset was determined by running the "Strategy" command from within MOSFLM. Processing was invariably performed in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of approximately a = 79 Å, b = 90 Å and c = 100 Å (see Table 3.2).

While a number of different X-ray sources were utilised to collect diffraction data, synchrotron radiation was used whenever possible to obtain high resolution structures with a high signal to noise. There were two synchrotron sources used in the current study: the Advanced Photon Source (APS) which consists of a 1,104 m storage ring with nominal electron energy of 7.0 GeV, and the Australian Synchrotron (AS) which consists of a 216 m storage ring with nominal electron energy of 3.0 GeV. At the APS, experiments were performed on beamline 23 ID-B using a MarCCD detector. At the AS, experiments were performed on beamline PX-1 (3BM-1B) using an ADSC Quantum 210 (Q210) detector. Translational movement across the crystal was performed during data collection to reduce radiation damage when using APS

synchrotron radiation. Our laboratory X-ray source (or home source) used in these experiments is a Nonius rotating anode generator using Cu K_α radiation and osmic confocal mirror optics with a Mar345 image plate detector.

3.2.7 Structure determination and refinement

After datasets were initially processed and integrated in MOSFLM, the reflection data file was then passed through a suite of programs in the CCP4 crystallography package [167]. Scaling of intensities and inspection of data quality was performed in SCALA. In particular, the resolution cut-off was determined by analysing the signal to noise parameter, $\langle I/\sigma(I) \rangle$, as it varies against resolution. The resolution corresponding to a $\langle I/\sigma(I) \rangle$ value of 2.0 was used as an approximate limit to the high resolution bin, as described in the literature [181]. The fraction of data used for the R_{free} calculation was set to 5%.

Phases were solved using molecular replacement with the maximum likelihood search function in PHASER [182]. The first importin-α dataset obtained was solved using the full length importin-α structure (PDB: 1IAL [37]) as a search probe, with the solvent and NLS peptide model removed, and atomic B-factors reset to an appropriate value as suggested from the Wilson plot. All subsequent datasets were then solved using a similar molecular replacement method with our importin-α structure as the probe. An initial solvent model was built using ARP/wARP [157], which was then manually checked and validated using the visualisation program COOT [156]. Water molecules that are not well ordered, as well as waters placed in or near the importin-α binding site, were removed.

Iterative refinement of the model was performed with multiple passes of REFMAC5 maximum likelihood refinement [155] and COOT manual refinement. Electron density for a bound peptide was generally much clearer in the major binding site compared with the minor binding site. Data reduction and refinement statistics are summarised in Table 3.2.

Table 3.2: Data reduction and refinement statistics for importin-α structures

	Importin-α (K25E, R28G)	***(no-SEC) importin-α (70-529)***	***(no-SEC) importin-α:SV40 NLS***	***(no-SEC) importin-α:CLIC4 NLS***
Data Collection				
Source (λ)	Home Source (1.54 Å)	APS BL 23 ID-B (0.98 Å)	APS BL 23 ID-B (0.98 Å)	APS BL 23 ID-B (0.98 Å)
Detector	Mar345	MarCCD	MarCCD	MarCCD
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)				
a	78.7	78.5	78.0	78.8
b	90.1	90.0	89.5	90.0
c	99.4	99.6	99.3	99.6
Resolution (Å)[a]	2.00 (2.11-2.00)	2.00 (2.10-2.00)	2.40 (2.53-2.40)	2.52 (2.66-2.52)
Total Observations	178,280	107,170	95,608	88,320
Unique Observations	48,308	43,650	25,409	24,549
Completeness (%)[a]	99.7 (100)	91.0 (82.6)	92.5 (95.5)	99.8 (100)
Mean I/σ(I)[a]	14 (2.9)	16.6 (3.3)	5.7 (2.6)	10.1 (1.9)
R_{merge} (%)[a,c]	6.1 (47.8)	3.8 (24.5)	16.9 (74.6)	9.4 (85.9)
Wilson B-value ($Å^2$)	32.7	37.4	40.7	57.4
Refinement and Structure				
R factor (R_{free})[c]	19.8 (23.7)	22.7 (24.8)	18.4 (24.2)	19.0 (24.3)
Number of atoms (Avg. B-value, $Å^2$)				
Importin-α	3,255 (40.6)	3,256 (43.6)	3,256 (47.1)	3,256 (53.1)
Waters	196 (46.0)	182 (46.3)	65 (42.4)	46 (49.0)
Ramachandran plot (%)[b]				
Favoured region	98.4	98.6	96.5	92.2
Allowed region	1.6	1.4	3.3	7.3
Disallowed	0	0	0.2	0.5

[a] Outer shell statistics shown in parenthesis.
[b] Calculated using MolProbity [159].
[c] The R_{merge} and R factor are defined in Section 2.2.7.1, p72.

3.3 Results

3.3.1 The (no-SEC) importin-α:SV40 NLS and importin-α:CLIC4 NLS structures

Importin-α (70-529) was initially co-crystallized with either the SV40 NLS peptide or the CLIC4 NLS peptide in a crystal condition that corresponds to previously reported crystallization experiments of importin-α (70-529). Using these co-crystals, we successfully obtained high quality diffraction datasets sufficient for structure determination. Molecular replacement was used to solve the structures which have been named (no-SEC) importin-α:SV40 NLS and (no-SEC) importin-α:CLIC4 NLS. In reference to the fact no ion exchange chromatography (IEC) or size exclusion chromatography (SEC) was used during purification of importin-α (70-529) in preparation for these structures, the prefix '(no-SEC)' has been added.

Electron density in the $2mF_o$-DF_c map (refer to Section 2.2.7.2) for the importin-α molecule is clear and unambiguous, including all residues from the beginning of the truncation at Asn70, up to Phe496 where density becomes too weak to continue modelling the protein main chain. Not accounted for in the electron density are C-terminal residues 497-529 and the 50 residue artificial *linker* region containing the N-terminal hexa-histidine tag. It is likely these residues are disordered due to inherent flexibility at the N-terminus and C-terminus of the importin-α (70-529) construct.

Inspection of the mF_o-DF_c difference electron density map, in the region of the importin-α (70-529) major binding site, clearly reveals the presence of a peptidic molecule in both the SV40 NLS and CLIC4 NLS structures (see Figure 3.2). Our first assumption upon observing density in the binding site was that our co-crystal peptides were successfully binding importin-α (70-529) in an NLS-like fashion. This result was almost expected after following standard protocols that are regularly used to produce importin-α:NLS structures.

Although the peptide side chain density is weak and ambiguous, apart from the lysine in binding slot P2, main chain density is strong with carbonyl and amide groups featuring prominently and allowing the peptide chain direction to be determined

unambiguously. The peptide runs anti-parallel to importin-α in an extended conformation as expected for an importin-α:NLS interaction. Furthermore, preliminary modelling using a poly-alanine peptide showed there is a pattern of main chain hydrogen bonding to the importin-α tryptophan and asparagines arrays that is also consistent with peptide binding observed previously [39].

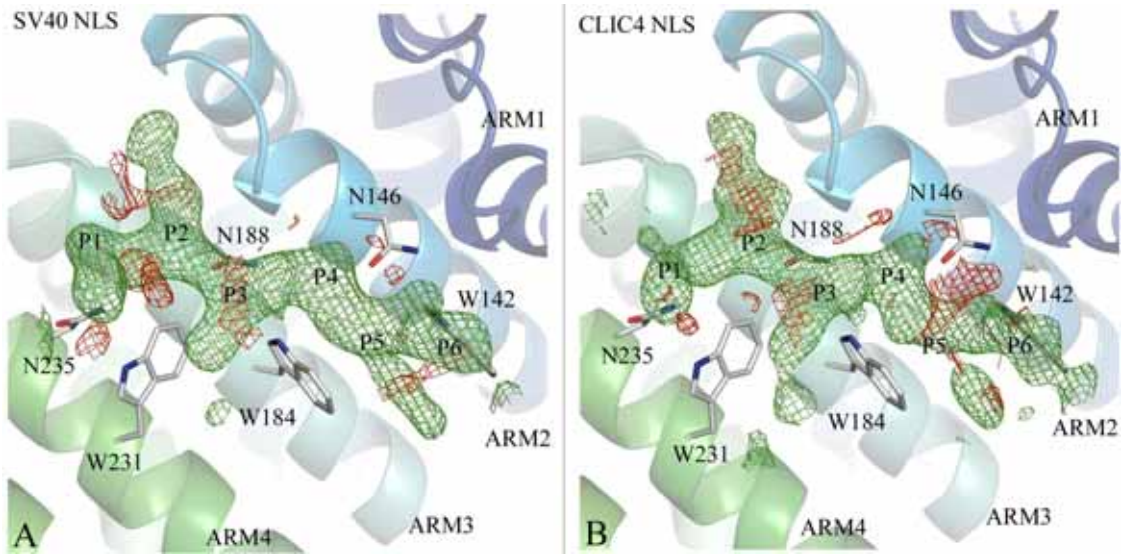

Figure 3.2: Electron density in the (no-SEC) importin-α:SV40 NLS and (no-SEC) importin-α:CLIC4 NLS structures.

Importin-α is shown in cartoon representation near the major binding site, coloured by armadillo (ARM) repeats. Tryptophan and asparagine arrays are shown in stick representation, defining binding slot positions P1 - P6. Difference map density (mF_o-DF_c) is contoured at +2.8σ (green) and -2.8σ (red) and carved around the binding site in both panels.

A) The (no-SEC) importin-α:SV40 structure, obtained by co-crystallization of importin-α and the SV40 peptide.

B) The (no-SEC) importin-α:CLIC4 NLS structure, obtained by co-crystallization of importin-α and the CLIC4 peptide.

The peptidic electron density seems to suggest the straightforward conclusion that our importin-α:NLS structures show that peptides added during co-crystallization are bound to importin-α (70-529). Indeed other labs have apparently also made this premature interpretation, going on to publish what are likely to be erroneous results (this is discussed in Section 4.4.5). However, after comparing the (no-SEC) importin-α:SV40 NLS and (no-SEC) importin-α:CLIC4 NLS structures, it is clear that features of the peptide electron density are essentially the same (Figure 3.2).

We proceeded by attempting to identify density features that could only be explained by the particular NLS peptide added during co-crystallization. We found that while it is possible to model a simple poly-alanine peptide into the binding site, modelling in complete side chains leads to an unsatisfactorily poor fit and highly ambiguous residue assignment. In the case of the SV40 NLS peptide, we expect electron density to indicate the presence of the ^{SV40}Arg130 side chain at binding slot P4 [178], however an alanine is the best fit in this location. For the CLIC4 NLS peptide, we were looking for an indication of its characteristic aromatic residue, ^{CLIC4}Tyr205, which must be present at P3 or P4 if ^{CLIC4}Lys202 or ^{CLIC4}Lys203 occupies the critical P2 slot. Like ^{SV40}Arg130, there is no side chain density that corresponds to ^{CLIC4}Tyr205.

There are mF_o-DF_c electron density stubs present in binding slots P1, P3, P4 and P6, which could indicate alanine residues are indeed present in these positions, or that longer side chains are disordered. Binding slot P5 has weak density for a long side chain in the (no-SEC) importin-α:SV40 NLS difference map, most likely corresponding to a lysine residue, but this density is weak and fragmented in the (no-SEC) importin-α:CLIC4 NLS difference map (see Figure 3.2).

Weak side chain density in the difference maps not only makes peptide residue assignment difficult, but indeed makes it impossible to unambiguously conclude that our particular NLS peptides have bound. As expected, creating a model of the SV40 NLS peptide or CLIC4 NLS peptide in the binding site produces a poor fit to the electron density. Since each structure contains density that corresponds to a peptide of the same length that lacks discernable side chain features, there is no evidence to indicate these are not the same peptide. This being the case, there may be an entirely different ligand that is binding importin-α (70-529) in both of these structures and that the electron density instead corresponds to it. If there were a molecule binding importin-α (70-529) other than our artificial peptides, it would likely have a high affinity for the binding site since it is able to compete with the classical NLS motif of SV40. Rather than occupying the site fully, this rogue peptide may share the binding site with the SV40 NLS and CLIC4 NLS across different importin-α (70-529) units in the crystal. Hence it is plausible that the electron density we observe is the result of a

mixture of peptides. Such a scenario may help to explain the ambiguous features seen in the density map.

In order to clarify this point, our experiments then focused on obtaining a structure of importin-α (70-529) free from any bound ligands, a structure we refer to as *apo* importin-α. After obtaining this, our thoughts were we should then able to begin solving importin-α:NLS peptide complexes with unambiguous electron density that could be used to identify the bound peptide.

3.3.2 The (no-SEC) importin-α (70-529) structure

Inspecting (no-SEC) importin-α:NLS structures revealed the possibility that rogue proteins or peptides are capable of binding importin-α with a high enough affinity that they compete with, and at least partially prevent, the distinguishable binding of the SV40 NLS and CLIC4 NLS peptides. In order to clarify this point, a structure of (no-SEC) importin-α (70-529) was obtained by crystallizing the same importin-α sample as that used in the (no-SEC) importin-α:NLS crystals. Crystallization was performed under similar crystallization conditions, but without co-crystallizing peptide additives. The (no-SEC) importin-α (70-529) structure was solved using molecular replacement at a resolution of 2.00 Å.

Similar to structures solved after co-crystallization with NLS peptides, the termini of the importin-α (70-529) construct were not observed in the electron density map. This includes residues from the artificial N-terminal *linker* sequence and importin-α residues Asn70, Gln71 and 497-529. The final model thus includes residues 72-496 and 182 waters. We also note the presence of diffuse difference electron density near the indole side chains of Trp273 and Trp357, which cannot be trivially accounted for by ordered water molecules (Figure 3.3). We reason there is a metal cation - π interaction with these aromatic side chains involving Na^+ ions that are abundant in the crystallization buffer. Although a modelled Na^+ ion refines to a distance of ~4 Å from the aromatic plane, consistent with a cation - π bond [183], it is unclear from examining the electron density that this is an unambiguous assignment. To satisfy the density, multiple cations need to be included at each site with varying levels of

occupancy. Due to these complications, no Na^+ cations have been modelled in our importin-α (70-529) structures.

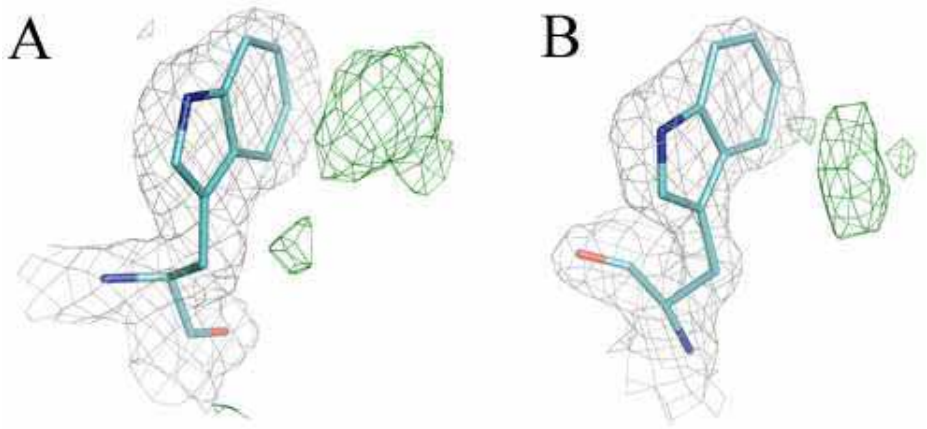

Figure 3.3: Cation - π interactions involving importin-α tryptophan residues.

The (no-SEC) importin-α (70-529) tryptophan residues corresponding to Trp273 (panel A) and Trp357 (panel B) are shown in stick representation. $2mF_o$-DF_c electron density is contoured around the tryptophan at 1.5σ (grey). Difference electron density (mF_o-DF_c) is contoured at 3.0σ.

Unexpectedly, the (no-SEC) importin-α (70-529) structure displays unaccounted for electron density in the major binding site (Figure 3.4), similar to that seen in the (no-SEC) importin-α:NLS structures (compare with Figure 3.2). The mF_o-DF_c difference map clearly shows the presence of a rogue peptide binding importin-α (70-529). This finding confirms that our (no-SEC) importin-α:NLS structures do not correspond to the SV40 NLS or the CLIC4 NLS binding importin-α (70-529).

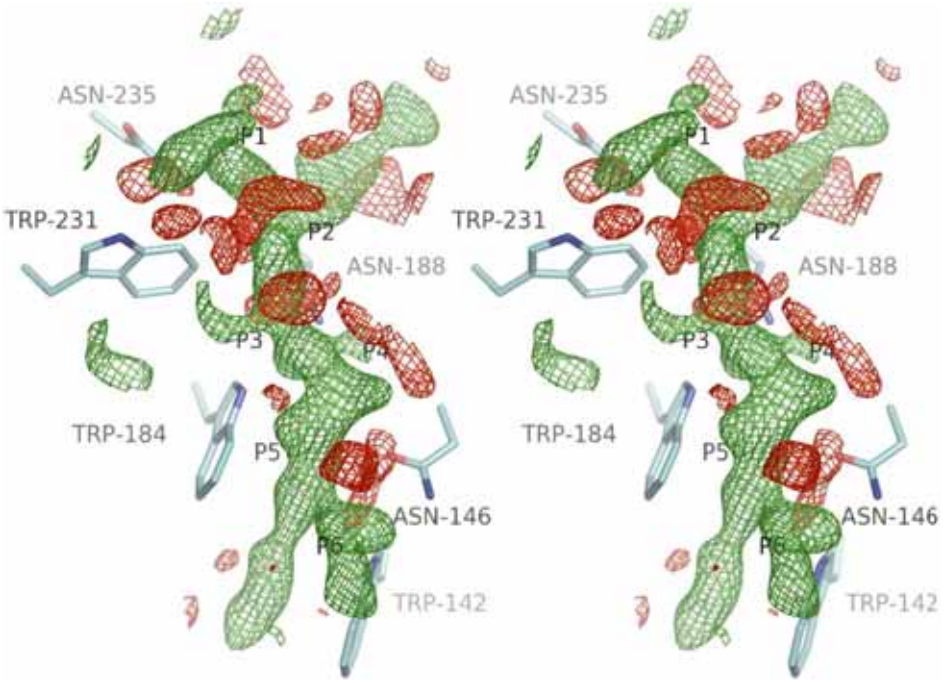

Figure 3.4: Stereo image of electron density near the (no-SEC) importin-α (70-529) major binding site.

The mF_o-DF_c difference map is shown, contoured at +2.8σ (green) and -2.8σ (red). Tryptophan and asparagines arrays are shown as sticks (cyan). A molecule with peptidic features is clearly defined by the positive density.

Covering the length of the binding site over P1 - P6, electron density for main chain atoms is again strong and unambiguously defines the rogue peptide orientation as running anti-parallel to importin-α (70-529). The binding slot at P2 contains density corresponding to a lysine side chain, although the fit is not as convincing as that seen in the (no-SEC) importin-α:NLS structures. Irregular density at P2 could suggest this binding slot is heterogeneously occupied. P5 now contains strong density that corresponds to either arginine or lysine, however the density here also displays globule-like features commensurate with a heterogeneously occupied site. At P1, P3 and P4, side chain density corresponds to alanine sized density stubs.

Poly-alanine modelling of the unknown peptide shows the sequence length corresponds to at least six C_α atoms. By adjusting the occupancy of the alanine residues and observing changes to R_{free} and electron density, we find the peptide is bound to importin-α at ~60% occupancy. Continuing this analysis by filling in longer side chains where density permits, the most plausible residue assignment for the rogue

peptide is of the form, ^{P1}AKAA(K/R)A^{P6}, where 'A' corresponds to alanine but may represent a longer disordered side chain.

3.3.3 The (no-SEC) importin-α (K25E, R28G) mutant structure

After noting a rogue peptide binding importin-α (70-529) in the absence of NLS peptides, our intention was to identify it and if possible, remove it. Since the peptide is able to compete with classical NLS sequences for the binding site, as seen in the (no-SEC) importin-α:SV40 NLS structure, and since it occupies the site at ~60% in the absence of added NLS peptides, it suggests the rogue peptide and importin-α (70-529) are in an environment where the molar ratio of the two molecules is approximately 1:1. Therefore we sought to rule out the possibility that the rogue peptide is in fact a part of the importin-α (70-529) molecule itself. If this was the case, it would explain the high molar ratio and enhanced affinity that would be necessary for the rogue ligand to competitively bind importin-α (70-529) in the presence of a high concentration of the SV40 NLS peptide.

The importin-α (70-529) construct is a 510 amino acid sequence that can be described as consisting of two distinct parts: an unfolded N-terminal *linker* domain 50 residues long and a C-terminal domain 460 residues long that is folded into a series of α-helical ARM repeats. The *linker* corresponds to the cloning and expression region in the pET-30a vector, while the C-terminal domain corresponds to residues 70-529 in importin-α. The N-terminal *linker* domain is disordered in our importin-α crystal structures suggesting a degree of flexibility and thus is a likely location for a tethered NLS-like motif.

A manual search of the importin-α (70-529) sequence for an NLS-like motif reveals a suspicious basic cluster in the N-terminal *linker* domain corresponding to ^{24}A**K**FE**R**Q^{29}. Although this motif contains two basic residues, with lysine and arginine capable of occupying the more significant P2 and P5 binding slots respectively, it is not immediately apparent that it would function as a high affinity NLS. The sequence does however correspond to the preliminary model of the unknown peptide (^{P1}A**K**AA(**K**/**R**)A^{P6}), where Lys25 fits in the P2 binding slot. This

would allow Arg28 to occupy the P5 slot and residues Phe26 and Glu27 would then occupy P3 and P4, positions at which there is no side chain electron density in the (no-SEC) importin-α (70-529) difference map. If the hydrophobic side chain of Phe26 and the negatively charged side chain of Glu27 do not make favourable interactions with importin-α, it is plausible that they are disordered in the crystal. This would then require the P2 lysine and P5 arginine interactions with importin-α to be strong enough to overcome these energetically unfavourable residues.

In order to eliminate this possibility, two importin-α (70-529) mutant constructs were designed: a double-point mutant (K25E, R28G) described in Section 3.2.1.1 and a deletion mutant that removed 40 residues from the *linker* domain (see Section 3.2.1.2). The properties of the deletion construct are sufficiently modified compared to the original importin-α (70-529) construct that it no longer crystallizes under the standard crystallization condition. A thorough screening procedure over new conditions resulted in no successful hits. Therefore the more specific (K25E, R28G) double-point mutant was used to investigate interactions with the binding site posed by the importin-α N-terminal *linker* region. The K25E point mutation by itself should be sufficient to obliterate any binding affinity, since Lys25 should occupy the critical binding slot P2 [60]. The double-point mutations change the N-terminal *linker* NLS-like motif to ^{24}A<u>E</u>FE<u>G</u>Q^{29} (underlined residues mutated), and the resulting construct is named the importin-α *linker* mutant.

While the *linker* mutant crystallized in similar conditions to that of the original importin-α (70-529) construct, it was more difficult to obtain good quality diffraction. Using molecular replacement, the structure of the importin-α *linker* mutant with no added NLS peptides was solved at 2.00 Å. There is clear electron density for importin-α residues 72-496, but no density for the N-terminal *linker* domain as per the (no-SEC) importin-α (70-529) structure. This means the two mutated residues (Glu25 and Gly28) could not be observed in the crystal structure.

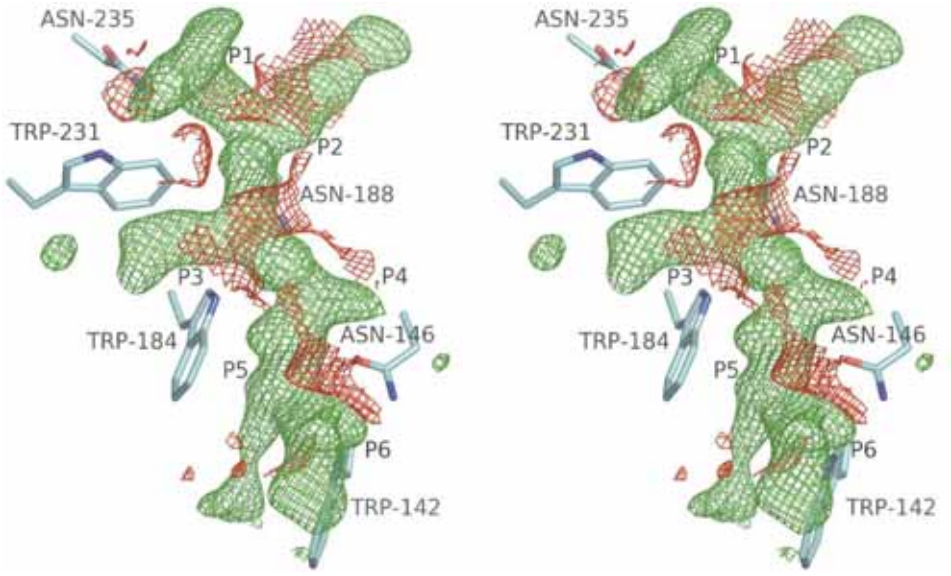

Figure 3.5: Stereo image of electron density near the (no-SEC) importin-α (K25E, R28G) mutant major binding site.

The mF_o-DF_c difference map is shown, with positive density contoured at +2.8σ (green) and negative density contoured at -2.8σ (red). The tryptophan and asparagine arrays are shown as sticks. A molecule with peptidic features is clearly defined by the positive density.

In the mF_o-DF_c difference map, once again there is clear electron density in the importin-α (70-529) major binding site corresponding to the presence of a peptide (see Figure 3.5). Comparing the density to earlier structures, in particular the (no-SEC) importin-α (70-529) structure, we see certain density features have become better defined. This is most evident in binding slot P3 where electron density now corresponds to more than a side chain stub, although no residue could be fit unambiguously. At P2, density once again clearly matches a lysine residue. Curiously, density at P5 is slightly weaker although it still appears to correspond to a long side chain.

The structure of the importin-α *linker* mutant shows that the N-terminal (K25E, R28G) mutation has no significant weakening effect on rogue peptide binding. Since there are no other NLS-like sequences within the importin-α (70-529) tail regions, we surmise that the rogue peptide cannot be fully explained by intrasteric binding. This does not exclude the possibility that in the (no-SEC) importin-α:NLS structures, the electron density represents a mixture of intrasteric binding from the *linker* region as well as rogue peptide binding.

3.4 Discussion

3.4.1 Preliminary (no-SEC) importin-α structures

Our preliminary attempts to determine importin-α (70-529):NLS structures followed a standard importin-α (70-529) purification protocol that is routinely reported in the literature, omitting ion exchange chromatography (IEC) and size exclusion chromatography (SEC) [50, 75, 177, 178]. To signify that the importin-α (70-529) sample was prepared using this protocol we refer to these structures with the preceding (no-SEC) descriptor. After co-crystallizing (no-SEC) importin-α (70-529) with the SV40 NLS or CLIC4 NLS peptides, it appeared that we had successfully obtained importin-α:NLS complexes with mF_o-DF_c difference electron density corresponding to a bound peptide, albeit with ambiguous side chain density. Peptide density is stronger in the major binding site compared to the minor binding site and therefore our analysis focuses on the more prominent major site.

Complications become apparent after comparing these initial co-crystal structures, (no-SEC) importin-α:SV40 NLS and (no-SEC) importin-α:CLIC4 NLS, where peptide side chain density features are virtually identical to each other. Poorly defined side chain density means the conservative crystallographer should refrain from building in a complete peptide model including all side chain atoms. Indeed this will leave the peptide unidentified without sufficient information to accurately determine the amino acid sequence. In our case there are two exceptions to weak side chain density: the critical binding slot P2 which clearly fits a lysine in all structures, and binding slot P5 which generously fits a lysine or arginine. The residue assignment at P2 and P5 is common to most NLS sequences and thus is not useful in identifying the bound peptide in our (no-SEC) importin-α:NLS structures.

Rather than over-interpret ambiguous electron density, we proceeded to show that in the absence of co-crystallized NLS peptides, the (no-SEC) importin-α (70-529) structure also contains peptidic electron density in the major binding site. This finding supports the hypothesis that a rogue peptide is binding importin-α (70-529) and ultimately serves as a crucial experimental control. Unfortunately it is still difficult to

identify the rogue peptide sequence in this structure. Electron density at the P2 binding slot is less well defined than previously observed, without a clearly defined outline of the lysine side chain. At P5, density clearly corresponds to a long side chain that may be lysine or arginine. The differences observed in the (no-SEC) importin-α (70-529) structure could be due to a heterogeneously occupied binding site in (no-SEC) importin-α:NLS structures, where the SV40 NLS peptide or CLIC4 NLS peptide bind importin-α (70-529) in some unit cells, while the rogue ligand binds importin-α (70-529) in others.

Furthermore, importin-α (70-529) structures determined by co-crystallization with an SV40 NLS or CLIC4 NLS peptide, may actually bind an impurity from peptide synthesis that corresponds to a truncated NLS peptide sequence. Such a possibility is less likely in the case of the CLIC4 NLS peptide, which has been purified more rigorously than other NLS peptides used in this study (>95% purity for the CLIC4 NLS peptide, see Section 3.2.2). The structure of (no-SEC) importin-α (70-529) qualitatively shows that even in the presence of a 7 x molar excess of the SV40 NLS peptide, importin-α (70-529) binds a rogue ligand with a high enough affinity that it can compete for the binding site, making interpretation of electron density difficult. These problems stem from the acute ability of the importin-α binding groove to bind non-specific targets.

3.4.2 An NLS-like motif in the importin-α (70-529) construct

For crystallization of importin-α complexes, we used a truncated form of importin-α (70-529) lacking the 69 residue long N-terminal auto-inhibitory domain, to avoid binding competition between it and NLS peptides. However the artificial N-terminal *linker* domain of importin-α (70-529) also contains a weakly basic NLS-like motif, ^{24}AKFERQ^{29}. Although the Phe26 and Glu27 side chains, which should occupy binding slots P3 and P4 respectively, have no corresponding electron density in the (no-SEC) importin-α difference maps, the long basic side chains Lys25 and Arg28 do fit density at P2 and P5. The absence of density for phenylalanine or glutamic acid side chains could be explained by disorder in these residues. Although the NLS-like motif is apparently weak, containing only two basic amino acids, it is

covalently part of the artificial importin-α (70-529) construct and proximal to the binding site, which could account for the high occupancy of the rogue ligand.

The proposed scenario is very similar to what takes place with the wild type N-terminal auto-inhibitory domain of importin-α, in which the apparently weak NLS sequence, ^{48}LKRRNV^{53}, competitively binds to the NLS recognition site [37, 58]. It is possible that by removing the auto-inhibitory domain in the importin-α (70-529) construct, another problem has been introduced with the artificial *linker* domain. Models of the truncated importin-α (70-529) construct do not normally include amino acids from the artificial N-terminal domain, due to disorder in what is presumably a flexible and unstructured region of the protein. An exception to this is the full length importin-α (1-529) structure, which includes residues 70-496 of the folded C-terminal domain, as well as residues 44-54 from the N-terminal auto-inhibitory domain [37]. It is probable the N-terminal domain residues are not disordered solely because of their intrasteric interaction with the major binding site. A flexible *linker* domain could also allow N-terminal residues free access to the binding site. Since the artificial NLS-like motif occurs 20 residues from the beginning of the importin-α (70-529) folded domain at residue 70, the *linker* should be sufficiently long to access the binding site in a favourable orientation.

To investigate the possibility that the *linker* region binds the importin-α (70-529) major binding site, we created a new importin-α (70-529) construct with the *linker* domain deleted. Unfortunately the deletion mutant did not crystallize, so to proceed with a structural method of analysis we performed a double point mutation leaving the *linker* domain otherwise intact, which appears to be required for crystallization. Mutagenesis of the importin-α (70-529) construct involved the double point mutation, K25E/R28G, that obliterates the NLS-like motif to become ^{24}A<u>E</u>FE<u>G</u>Q^{29}. We solved the structure of this mutant construct, finding that it also contains electron density corresponding to a rogue peptide. The mF$_o$-DF$_c$ difference electron density map shows features of the rogue peptide more clearly than observed in (no-SEC) structures, particularly at binding slot P3 where longer side chain density is visible. Not in line with the generally stronger density, P5 contains weak density corresponding to a long

basic side chain. This could be an indication that a lysine or arginine from the NLS-like *linker* motif occupies this site in the (no-SEC) importin-α (70-529) structure.

We hypothesise that the mutation of the importin-α *linker* NLS-like motif was successful in preventing intrasteric binding, allowing a rogue ligand to homogenously occupy the major binding site without competition. The structure of the importin-α (K25E, R28G) *linker* mutant serves as important evidence that the rogue peptide does not correspond to an intrasteric interaction, leaving the possibility that endogenous *E. coli* peptides or proteins are binding importin-α (70-529). This appeared to be an unlikely scenario which would require the *E. coli* peptides/proteins to bind in an environment where the molar ratio to importin-α molecules is ~1. This could occur during expression in BL21(DE3) cells, or just after cell lysis, requiring the complex to remain stable throughout subsequent purification and crystallization. We explore this possibility in the Chapter 4.

3.5 Conclusion

Our preliminary investigations into importin-α:NLS structures reveal the presence of unaccounted for electron density that corresponds to a bound peptide in the major binding site. Although certain features of the density in (no-SEC) importin-α:SV40 NLS and (no-SEC) importin-α:CLIC4 NLS structures could correspond to the SV40 NLS or CLIC4 NLS peptides, residue side chains in core binding slots are unidentifiable. It is likely this density corresponds to a mixture of the co-crystallized peptide, an NLS-like motif in the artificial *linker* region and the 30S ribosomal S21 subunit, later identified as the rogue ligand (described in Chapter 4). The (no-SEC) importin-α (70-529) crystal structure showed that the unaccounted for electron density is still present without co-crystallized peptides. The importin-α (K25E, R28G) *linker* mutant structure confirms that an intrasteric interaction is not wholly responsible for the density in (no-SEC) structures, suggesting a peptide or protein from *E. coli* is binding importin-α (70-529).

The obstacles we explore in this chapter reveal the importance of acquiring an importin-α (70-529) structure containing an empty binding site, which we refer to as *apo* importin-α (70-529). The (no-SEC) structures described here are all derived from an importin-α sample that has been purified according to protocols in the literature that are commonly used for importin-α (70-529):NLS structures. Our findings demonstrate that this purification protocol is not sufficient to obtain *apo* importin-α (70-529), and so we were required to develop a more intensive purification protocol. An analysis of the importin-α:PLSCR1 NLS structure in Chapter 4 demonstrates that it is likely the problems we discovered in our (no-SEC) importin-α (70-529) structures were missed by another research group. Our experimental focus was thus directed into obtaining an *apo* importin-α (70-529) structure with an empty binding site, which would allow co-crystallization of importin-α (70-529) with SV40 NLS and CLIC4 NLS peptides, so that electron density can be unambiguously interpreted.

Chapter 4

Importin-α and NLS peptide interactions

4.1 Introduction

The import of proteins across the nuclear membrane in eukaryotic cells is a process regulated by Nuclear Localisation Signals (NLSs) that designate the cargo protein ready to be transported to the nucleus. In the classical nucleocytoplasmic pathway, an NLS normally consists of one (monopartite) or two (bipartite) basic clusters of amino acids that are recognised by the nuclear import receptor, importin-α. Importin-α consists of two structural domains: an N-terminal auto-inhibitory domain also known as the importin-β binding (IBB) domain, and a C-terminal α-helical domain made up of repeating armadillo (ARM) motifs. Importin-α acts as an adaptor protein for importin-β by binding NLS bearing cargo and forming the trimeric importin-α:β:NLS complex. This complex transports the cargo through the NPC where interactions with FG-nucleoporins are mediated by importin-β.

Our preliminary structural studies of a truncated form of importin-α missing the N-terminal auto-inhibitory domain, importin-α (70-529), revealed the presence of a rogue peptide interacting with the importin-α major binding site in the absence of added NLS peptides. By extending the standard importin-α (70-529) purification protocol as reported in the literature, we have obtained a high purity sample of importin-α (70-529) with a vacant binding site and subsequently solved the first mammalian *apo* importin-α (70-529) crystal structure at 1.77 Å. This is the highest resolution structure of any importin-α currently determined. The binding site is occupied by waters and other small molecules, and the electron density corresponding to the presence of a rogue peptide has been significantly reduced.

Using *apo* importin-α (70-529), the importin-α:SV40 large T-antigen NLS peptide structure is solved at 1.85 Å. We confirm the SV40 NLS peptide is unambiguously modelled in the binding site by calculating the $F_o^{sv40} - F_o^{apo}$ data-data

difference Fourier. Although the SV40 NLS is perhaps the most widely studied importin-α binding partner, our structure, the highest resolution achieved to date for an importin-α:NLS complex, reveals new insights into various modes of side chain binding. By comparing normalised atomic B-factors in the importin-α:SV40 NLS complex and *apo* importin-α, we introduce the B_z^{-apo} score as a measure of changes in residue flexibility. This is the first time such analysis has been performed on importin-α complexes and proves to be a useful method for determining the localised impact of NLS binding on importin-α surface residues.

After separation of importin-α:*E. coli* protein complexes from *apo* importin-α (70-529) during purification, we were able to obtain a relatively high concentration of importin-α (70-529) bound to a rogue peptide and successfully solved the complexed structure at 1.95 Å. Using structural and mass spectrometry data, the rogue peptide was identified as a six residue sequence, ^{45}RKRAKA^{50}, belonging to the *E. coli* 30S ribosomal subunit S21. We show that it is likely that importin-α (70-529) is bound to the intact full length S21 protein in a heterodimeric state with the *E. coli* 30S ribosomal protein S11. Although the S21 NLS-like sequence corresponds to the observed electron density, there is still the possibility that there is a mixture of bound peptidic species confusing the interpretation of electron density.

The interference from ribosomal proteins in our structure based experiments may have important repercussions in the field. It is likely that other labs working on importin-α:NLS peptide structures have observed the binding of S11:S21, but have mistaken the electron density for their co-crystallized peptide. As we demonstrate, there is already an example of this in the literature, where an importin-α (70-529) bound peptide has been ambiguously modelled in electron density that most likely corresponds to S11:S21. It is difficult to ascertain just how widespread the problem is in the field and it is important that future importin-α:NLS peptide studies consider the implications of our results.

4.2 Methods

4.2.1 Protein expression and purification

The standard protocol of importin-α (70-529) expression and purification has been described previously in Chapter 3. In the following sections we describe two extra purification steps that are relevant to this chapter: ion exchange chromatography and size exclusion chromatography.

4.2.2 Ion exchange chromatography

Following the standard expression and purification protocol outlined Chapter 3, crystallization of importin-α (70-529) was achievable which was suitable for X-ray diffraction. However, the standard protocols left endogenous *E. coli* proteins in the final sample. It was found that an improved purification protocol was necessary to separate *apo* importin-α (70-529) from importin-α (70-529):*E. coli* protein complexes. This was achieved with ion exchange chromatography (IEC) described here, and size exclusion chromatography described in the next section.

For IEC, protein was loaded onto a Hiload^{TM} 26/10 Q-sepharose^{TM} anion exchange column (GE Healthcare) immediately after imidazole gradient IMAC purification. Importin-α was allowed to bind to the $N^{+}(CH_3)_3$ charged groups of the Q-sepharose^{TM} HP resin in Buffer A (Buffer A: 50 mM NaCl, 20 mM Tris pH 8.0, 2 mM DTT), before being eluted over a salt gradient from 50 mM to 1 M at a flow rate of 4 mL/min (total volume of 400 mL). Figure 4.1 shows a typical chromatogram obtained from ion exchange purification of importin-α (70-529), including SDS-PAGE analysis. Eluted fractions with the highest concentration of protein were pooled and concentrated using an Amicon YM-10 Centriprep centrifugal concentrator in a Sorvall SH-3000 swinging bucket rotor at 3,000 rpm (1,800 *g*). Concentrated protein was then passed through a size exclusion column as outlined in the next section.

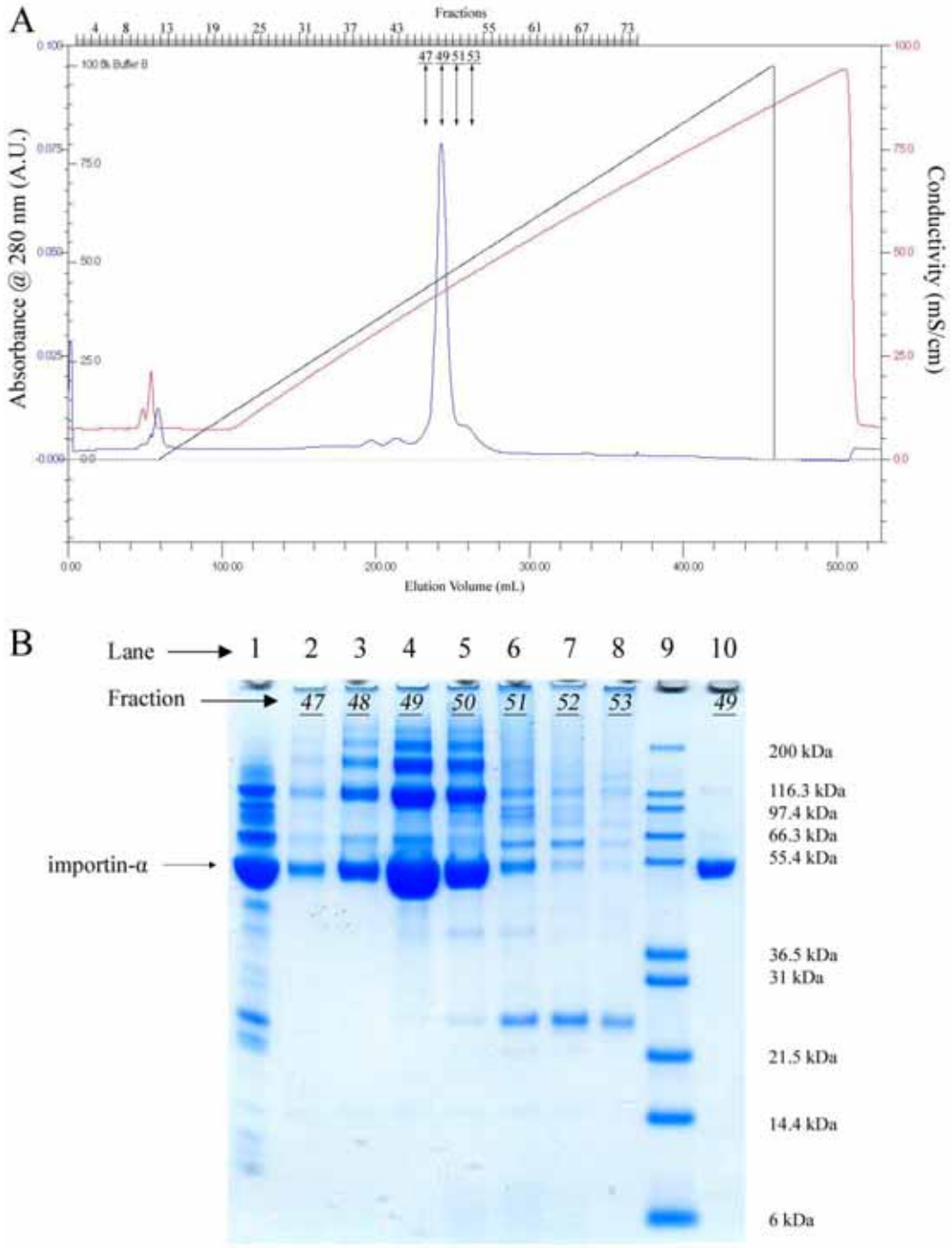

Figure 4.1: Ion exchange chromatography of importin-α (70-529) and SDS-PAGE analysis.

A) The blue curve represents A_{280}, the red curve represents conductivity (mS/cm) and the black curve represents the pump gradient. The conductivity gradient increases from 50 mM NaCl (100% Buffer A, 0% Buffer B) to 1 M NaCl (0% Buffer A, 100% Buffer B). Note there is a time lag between the salt gradient (black line) and the reading from the blue and red curves. Importin-α elutes at a conductivity reading of ~40 mS/cm (fraction 49 or 42%/58% Buffer B/A, which equates to ~450 mM NaCl). Fractions have been labelled at the top of the graph and key fractions are underlined.

B) SDS-PAGE of fractions collected in panel A. Lane 1 contains the loaded sample and lane 9 contains Mark12™ unstained protein ladder. Lanes 2-8 contain fractions 47-53 which have been prepared by trichloroacetic acid (TCA) precipitation. The sample in lane 10 corresponds to lane 4 prior to TCA precipitation.

4.2.3 Size exclusion chromatography

The final step in the improved protocol was size exclusion chromatography (SEC). The concentrated protein from IEC was loaded on a Superdex 200 10/300 GL (GE Healthcare) analytical size exclusion column and eluted with 20 mM Tris pH 8.0, 100 mM NaCl and 2 mM DTT (Figure 4.7, page 154). Fractions 15 and 16 corresponding to importin-α (70-529) were pooled and concentrated to approximately 18 mg/mL using an Amicon Ultra concentrator (Millipore) with a 10 kDa molecular weight cut-off. Fractions 13 and 14 corresponding to importin-α (70-529) with endogenous *E. coli* proteins were pooled and concentrated to 6.9 mg/mL. Aliquots of 75 µL were then pipetted, snap frozen in liquid nitrogen and stored at -80 °C until use.

To help in the qualitative analysis of SEC results, a standard calibration curve was created as shown in Figure 4.2. The calibration was performed on a Superdex 200 10/300 GL (GE Healthcare) analytical size exclusion column. Data points have been plotted using a Low Molecular Weight gel filtration calibration kit (Amersham) with the following protein standards: ribonuclease A, 13.7 kDa; carbonic anhydrase, 29 kDa; ovalbumin, 43 kDa; albumin, 66 kDa. Blue dextran 2000 (2,000 kDa) was used to determine the column's void volume equal to 8.65 mL. Values for the partition coefficient (K_{av}) were calculated using the formula,

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

where V_e is the elution volume, V_0 is the void volume (8.65 mL) and V_t is the total column volume (24 mL).

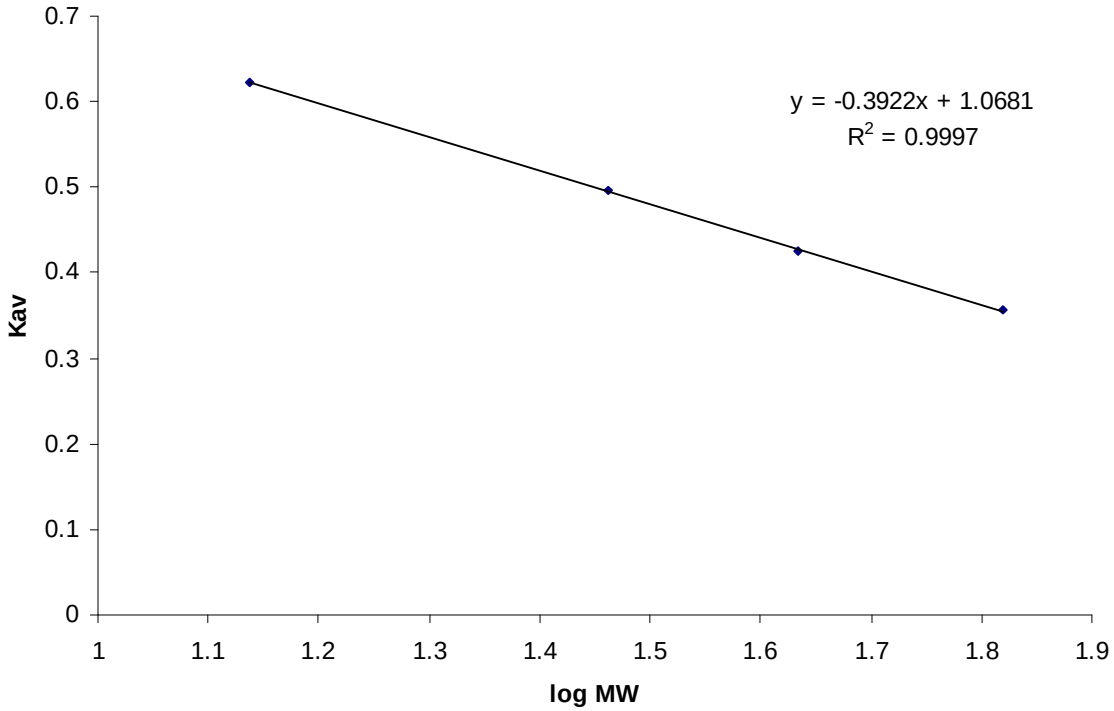

Figure 4.2: Calibration curve for the Superdex 200 10/300 GL SEC column.

The graph plots the partition coefficient K_{av} against the log of molecular weight. Four data points have been plotted at 13.7 kDa, 29 kDa, 43 kDa and 66 kDa. The equation of best fit is shown, along with an R^2 correlation coefficient defined by the square of the Pearson product moment.

4.2.4 Mass spectrometry

To determine the identity of proteins in the importin-α (70-529) sample containing *E. coli* proteins, electrospray ionisation mass spectrometry (ESI-MS) experiments were performed using an Applied Biosystems QSTAR Pulsar in positive ion mode. Time of flight (TOF) mass spectrometry was used to accurately determine the molecular weight of intact protein, as shown in Figure 4.3 for importin-α (70-529). Peptide mass finger printing was performed on excised SDS-PAGE gel bands after being subjected to trypsin digestion, where experimentally determined peptide masses were compared to theoretical values using the Mascot search engine [184].

The Mascot search uses a probability based scoring algorithm to rank protein hits based on an accumulation of ion scores defined as $-10 \cdot \log(P)$, where P is the probability that the observed ion is a random match to the sequence database. In our experiments, the significance threshold for ion matches was set to P = 0.05. Total

Mascot protein scores as referred to in this thesis are obtained by adding all matched ions for the protein, which means the score is non-probabilistic but useful for comparing protein hits.

Mass spectrometric results were obtained at the Bioanalytical Mass Spectrometry Facility within the Analytical Centre of the University of New South Wales with assistance from Dr. Mark Raftery. Subsidised access to this facility is gratefully acknowledged.

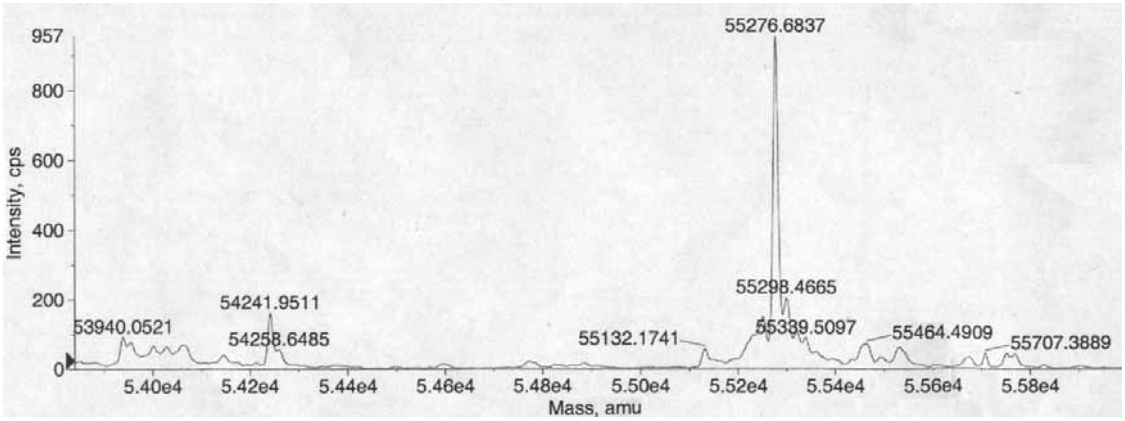

Figure 4.3: ESI-TOF-MS mass spectra of importin-α (70-529).

Shown is the ESI-TOF-MS mass reconstruction of *apo* importin-α (70-529). The peak value of 55,276.7 Da corresponds closely to the theoretical value of 55,275.8 Da.

4.2.5 Crystallization and data collection

Apo importin-α (70-529)

The best crystals of *apo* importin-α formed in a condition consisting of 0.7 M sodium citrate, 10 mM DTT and 70 mM HEPES at pH 7.0. Drops consisted of 2 µL of importin-α (70-529) at 18 mg/mL and 2 µL of reservoir solution. Crystals grew to maturity over a period of 20 days at a temperature of 20 °C (see Figure 4.4).

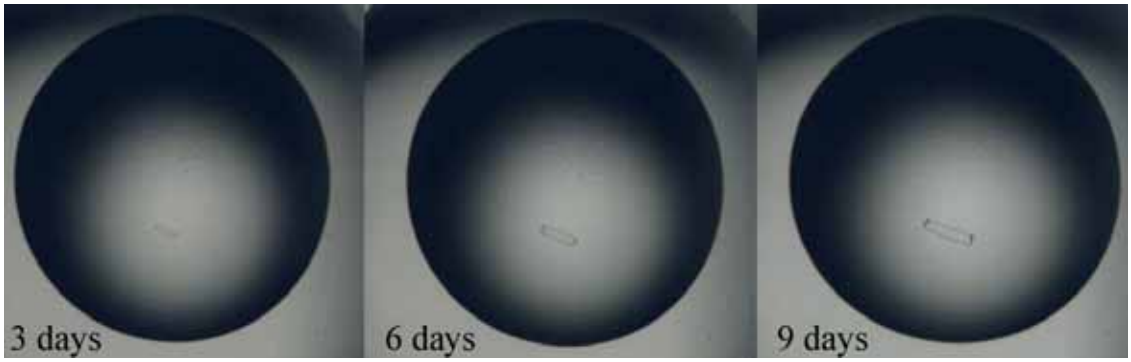

Figure 4.4: A time lapse image of an *apo* importin-α (70-529) crystal.

This crystal was grown in 0.7 M sodium citrate, 10 mM DTT and 70 mM HEPES at pH 7.0.

Importin-α (70-529) + SV40 NLS large T-antigen peptide (126-132)

The importin-α (70-529) + SV40 large T-antigen NLS peptide (residues 126-132) crystals were obtained by co-crystallization using hanging drop vapour diffusion. The reservoir consisted of 500 µL of 0.72 M sodium citrate, 10 mM DTT and 72 mM HEPES at pH 7.4. Drops consisted of 0.75 µL of importin-α (70-529) at 18 mg/ml, 0.75 µL of the SV40 NLS at 2 mg/ml and 1 µL of reservoir. This gives a 7 x molar excess of the SV40 NLS peptide over importin-α (70-529). There was no incubation of the peptide with importin-α before co-crystallization.

Importin-α (70-529) + 30S ribosomal subunit S21

Crystals grew in a condition consisting of 0.9 M sodium citrate and 90 mM HEPES at pH 7.4. Drops consisted of 2 µL of importin-α (70-529) at 6.9 mg/mL and 2 µL of reservoir. Ribosomal proteins were already present in the importin-α sample as contaminants remaining after purification by SEC. The lower importin-α concentration meant that crystals took a longer time to grow, reaching maturity in two months at 20 °C.

Crystal handling and Data Collection

General procedures used in crystal handling and data collection have been previously described in Section 3.2.6. An X-ray diffraction image of the *apo* importin-α (70-529) crystal is shown in Figure 4.5.

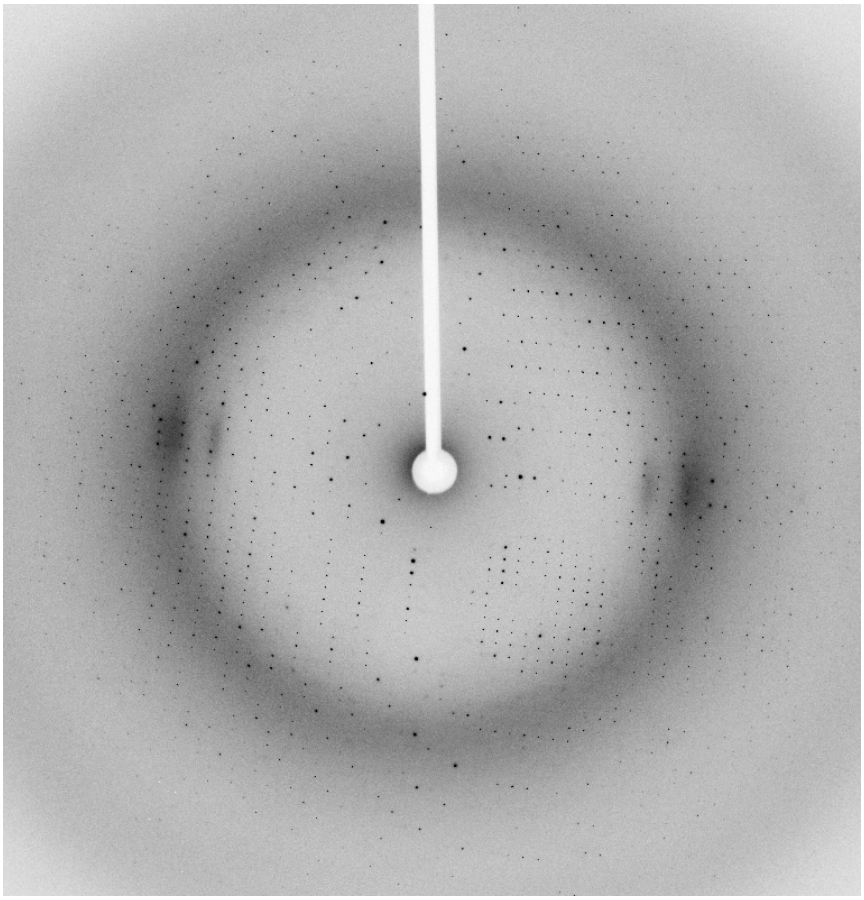

Figure 4.5: An X-ray diffraction image of the *apo* importin-α (70-529) crystal.

Shown is a 1° oscillation image taken on a MarCCD detector at the Advanced Photon Source (APS) synchrotron in Argonne, Illinois. The image has been cropped and corresponds to a resolution of 1.75 Å at the corners. The crystal to detector distance is set at 250 mm.

4.2.6 Structure determination and refinement

Structure determination and refinement of importin-α (70-529) models presented in this chapter closely followed the protocols previously described in Section 3.2.7. Data reduction and refinement statistics are summarised in Table 4.1.

Table 4.1: Data reduction and refinement statistics for importin-α (70-529) structures.

	apo importin-α (70-529)	*Importin-α:SV40 NLS*	*Importin-α:S21*
Data Collection			
Source (λ)	APS BL 23 ID-B (0.98 Å)	AS BL PX-1, 3BM-1B (0.95 Å)	AS BL PX-1, 3BM-1B (0.95 Å)
Detector	MarCCD	ADSC Q210	ADSC Q210
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)			
a	78.9	78.1	78.7
b	89.8	89.5	89.6
c	100.3	99.5	100.2
Resolution (Å)	1.77 (1.82-1.77)	1.85 (1.95-1.85)	1.95 (2.06-1.95)
Total Observations	477,474	269,549	211,540
Unique Observations	69,418	56,931	51,843
Completeness (%)[a]	99.0 (94.1)	94.8 (71.2)	99.2 (98.6)
Mean I/σ[a]	19.5 (2.2)	16.1 (2.9)	13.5 (2.0)
R_{merge} (%)[a,c]	9.8 (39.7)	5.3 (51.9)	7.5 (59.8)
Wilson B-value ($Å^2$)	30.4	26.3	28.1
Refinement and Structure			
R-factor (R_{free})[c]	19.2 (21.6)	18.4 (21.4)	18.9 (22.5)
Number of atoms (Avg. B-value, $Å^2$)			
Importin-α	3,255 (35.5)	3,237 (30.8)	3,278 (32.7)
Peptide	-	114 (24.5) 46 (54.9)	59 (40.2)
Waters	348 (44.6)	448 (42.1)	381 (42.1)
Ramachandran plot (%)[b]			
Favoured region	99.5	98.8	98.6
Allowed region	0.5	1.2	1.4
Disallowed	0	0	0

[a] Outer shell statistics are shown in parenthesis.
[b] Calculated using MolProbity [159].
[c] The R_{merge} and R factor are defined in Section 2.2.7.1, p72.

4.2.7 Structural analysis

4.2.7.1 The data-data difference Fourier

Fourier analysis was used to compare importin-α:NLS peptide complexes (observed structure factors, F_o^{pep}) with *apo* importin-α (observed structure factors, F_o^{apo}). The structure factors from each dataset have been scaled, (represented by the subscript 'o-s', F_{o-s}^{pep} and F_{o-s}^{apo}), using the CCP4 program SCALEIT [167] prior to the Fourier calculation. We refer to the Fourier difference of these structure factors as a data-data difference Fourier, which is calculated using the electron density equation (Section 2.2.7.2) where,

$$\mathbf{F} = \left(\left| F_{o-s}^{pep} \right| - \left| F_{o-s}^{apo} \right| \right) \cdot e^{i\alpha_c^{apo}}$$

The phases for this Fourier are calculated using the *apo* importin-α (70-529) structure. In referring to the data-data difference map, we drop the use of the subscript 'o-s' replacing it simply with 'o'. The resulting electron density map, $F_o^{pep} - F_o^{apo}$, is useful for objectively determining the presence of a bound NLS peptide. It is important to note the distinction between the $F_o^{pep} - F_o^{apo}$ data-data difference map and the regular mF_o-DF_c difference map.

Figures showing electron density have been prepared using PyMOL [165] unless otherwise stated. The "carve" command from within PyMOL has sometimes been used to remove density contours at a given distance (~2.0 Å) from selected residues, allowing clearer images of relevant regions.

4.2.7.2 Normalised B-factor analysis

In order to analyse conformational dynamics of importin-α residues, we have used Debye Waller factors also known as atomic B-factors. Atomic B-factors have been calculated for each non-hydrogen atom during refinement and are proportional to the

mean square displacement of the atom about its equilibrium. The isotropic atomic B-factor is thus defined as,

$$B_j = 8\pi^2 \left\langle x_j{}^2 \right\rangle$$

where B_j is the B-factor of the j^{th} atom, and x_j is the isotropic displacement of the atom from its equilibrium position in ångströms.

There are a number of contributions to B-factors arising from local displacement (e.g. residue mobility) or from global displacements (e.g. lattice disorder and protein vibration). In the case of protein structures, B-factors are related to static disorder where thermal motion has been 'locked in' upon flash freezing the protein crystal. This static disorder may be due to protein mobility that has been 'locked in', or crystal lattice imperfections. One must take care when comparing atomic B-factors across multiple structures to ensure the comparison is meaningful.

In our analysis, we are comparing importin-α structures that have been solved in the $P2_12_12_1$ space group with almost identical unit cell dimensions and refinement procedures. We therefore feel confident that the local variation of B-factors across our different structures is meaningful and gives an indication of thermal mobility at biologically relevant temperatures. In order to emphasise local differences in B-factors, we have developed a short protocol to prepare structures for comparison, similar to what has been used in the literature previously [185].

A representation of each step in the protocol is shown in Figure 4.6, for comparisons between *apo* importin-α (70-529) and the importin-α:SV40 NLS peptide complex. The B-factors of atoms in structure *j* are first normalised according to the following equation,

$$B_z = \frac{B_i - \mu_j}{\sigma_j}$$

where B_z is the normalised B-factor, B_i is the B-factor of atom *i*, μ_j is the mean B-factor in structure *j* (excluding waters and peptide atoms), and σ_j is the standard deviation of B-factors in structure *j*. The normalised B-factors have a zero mean and unit variance. We first found all atoms that satisfy $B_z \geq 4$ and treated these as outliers (Figure 4.6A). However, since the C-terminus of importin-α has fluctuating B-factors, all atoms past the first outlier have been omitted, regardless of their normalised B-factor. In the *apo* importin-α structure, the first residue treated as an outlier is residue 431 (atom number 2717) and for simplicity this is applied to all structures.

After discounting outliers, B_z values were recalculated (Figure 4.6B). The normalised B-factors from structure *j* are then subtracted from structure *j+1*, and the difference has again been normalised using the equation above (see Figure 4.6C). In our case, *apo* importin-α (70-529) always corresponds to *j* and importin-α:NLS peptide complexes correspond to *j+1*. The final values are referred to as the B_z^{-apo} score in reference to the subtraction of *apo* importin-α B_z values.

The B_z^{-apo} score is averaged over residue side chain and main chain atoms using the CCP4 program PDBSET [167]. It was then coloured on a surface representation of importin-α using PyMOL [165] over a blue-white-magenta standard deviation spectrum from -3σ to +3σ. The B_z^{-apo} scores are used for analysing three complexes: importin-α:SV40 NLS, importin-α:S21 and importin-α:CLIC4 NLS.

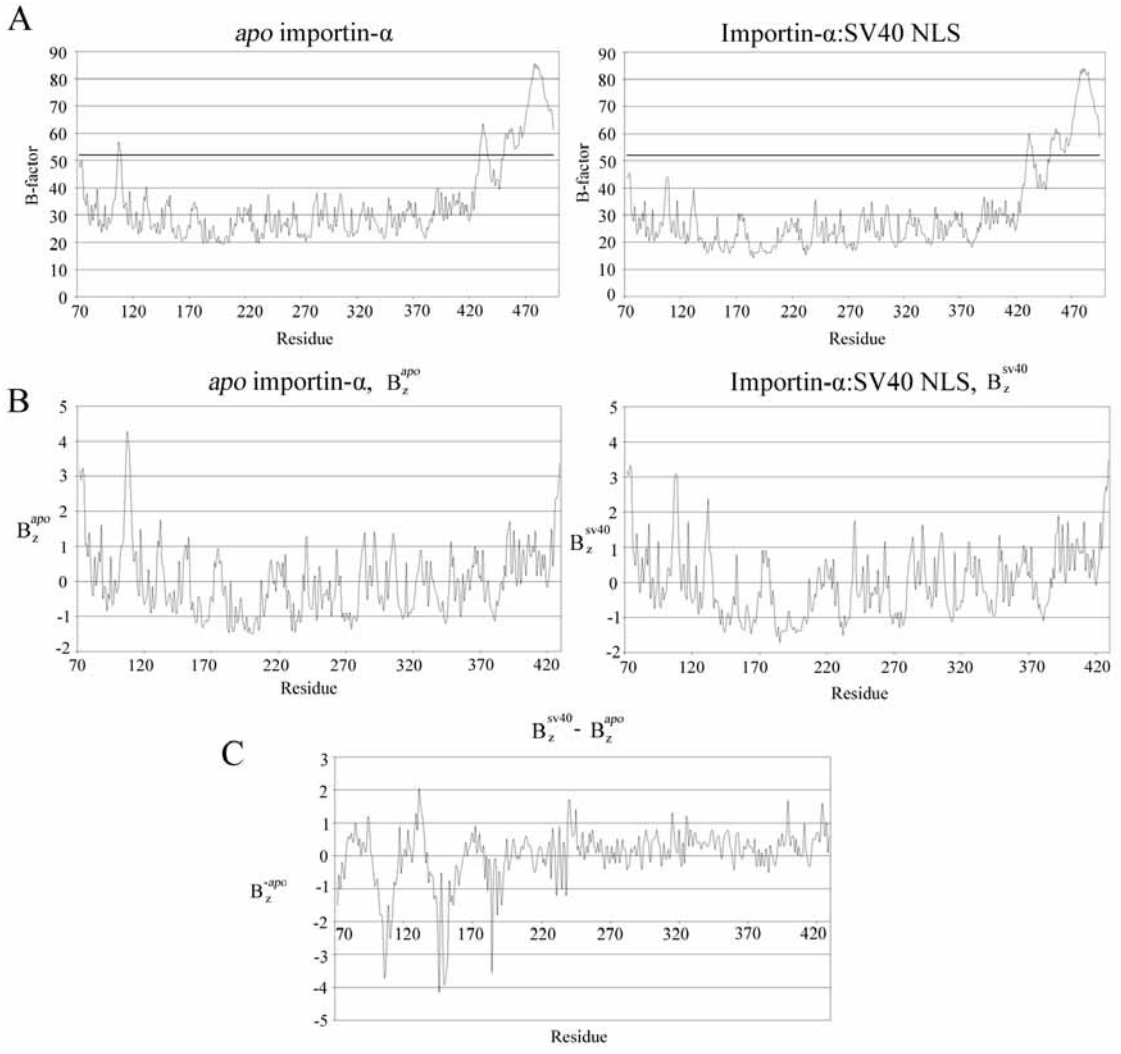

Figure 4.6: Raw and normalised atomic B-factors for *apo* importin-α and importin-α:SV40 NLS structures, with the calculated B_z^{-apo} score.

A) Experimentally observed atomic B-factors including outliers, for all importin-α atoms. The standardised score cut-off at Z = 4 is shown as a bold black line.
B) Normalised B-factors after outliers (Z ≥ 4) have been removed.
C) The B_z^{-apo} score is shown, representing the subtraction, $B_z^{sv40} - B_z^{apo}$, after normalisation.

4.2.7.3 Real space correlation coefficient

The real space correlation coefficient (RSCC) was determined with the CCP4 program OVERLAPMAP [167, 186] which compares two electron density maps and calculates an RSCC for each residue. We have used coordinates from importin-α:NLS structures to calculate a theoretical map (F_c), which is correlated with the experimental map (F_o), using refined phases.

The RSCC is defined as,

$$RSCC = \frac{\langle x \cdot y \rangle - \langle x \rangle \cdot \langle y \rangle}{\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \cdot \sqrt{\langle y^2 \rangle - \langle y \rangle^2}}$$

where x and y represent electron density values from the two maps to be compared.

4.3 Results

4.3.1 The *Mus musculus apo* importin-α (70-529) structure

In order to obtain an *apo* structure of mouse importin-α (70-529), the improved purification protocol was used as outlined in Section 4.2.1 - 4.2.3. The improved purification involved the additional use of ion exchange chromatography (IEC) and size exclusion chromatography (SEC). The *apo* form therefore corresponds to the (+SEC) importin-α structure, following on from the nomenclature used in Chapter 3. An IEC purification step appears necessary to prepare the importin-α (70-529) sample so that SEC can separate out endogenous *E. coli* proteins.

The final size exclusion chromatogram performed after IEC is shown in Figure 4.7. Protein corresponding to the major SEC peak 3 (*apo* importin-α (70-529)) was separated from material in peak 2, although the SEC shows this is only partial separation. Using the column calibration curve presented in Section 4.2.3, peak 1 corresponds to the void volume, peak 2 corresponds to a molecular weight of 159 kDa and peak 3 corresponds to a molecular weight of 68 kDa. Using SDS-PAGE, the protein in peak 3 is shown to correspond to high purity monomeric importin-α (70-529), with a theoretical molecular weight of 55.3 kDa. The SDS-PAGE results reveal there is a clear improvement of importin-α purity in the major peak, demonstrated by the absence of three bands at 64 kDa, 15 kDa and 10 kDa which were later identified as importin-α (70-529) + 30S ribosomal subunit S11, 30S ribosomal subunit S11 and 30S ribosomal subunit S21, respectively (refer to Table 4.3, p189). There is also a band at 110 kDa that corresponds to an importin-α (70-529) dimer. Possible reasons for the observation of a protein complex running on SDS-PAGE is discussed later in the thesis (refer to Section 4.3.5, p191).

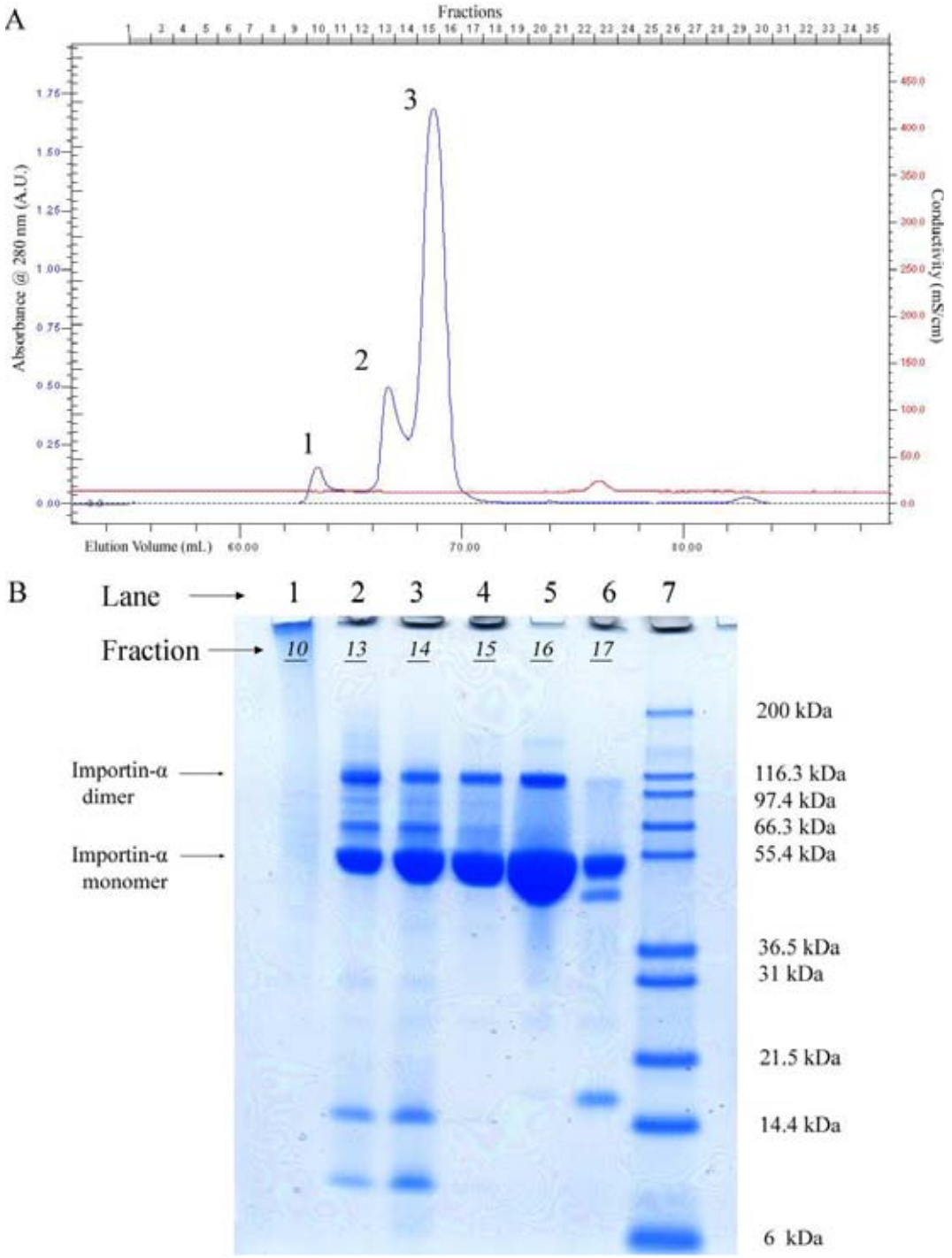

Figure 4.7: Size exclusion chromatography and SDS-PAGE analysis of importin-α (70-529).

A) The Superdex 200 10/300 GL SEC of importin-α run at a column flow rate of 0.25 mL/min. The blue curve corresponds to A_{280} and the red curve corresponds to conductivity (mS/cm). Three A_{280} peaks have been numbered where peak 1 corresponds to the void volume and peaks 2 and 3 correspond to partially separated protein species at 159 kDa and 68 kDa respectively (MW calculated using the column's calibration curve).

B) An SDS-PAGE gel of samples collected from SEC in panel A. Bands corresponding to monomeric and dimeric importin-α are labelled. Lane 1 corresponds to SEC peak 1, lanes 2 and 3 to SEC peak 2 and lanes 4 and 5 to SEC peak 3. Lane 7 is the Mark12^{TM} unstained protein standard (Invitrogen), and the MW of each band shown to the right. Bands corresponding to monomer and dimer are also shown.

After concentrating the protein sample corresponding to SEC peak 3, or gel lanes 4 and 5 (Figure 4.7), an *apo* form of importin-α (70-529) was successfully crystallized using the hanging drop vapour diffusion technique and a reservoir solution consisting of 0.7 M sodium citrate, 10 mM DTT and 70 mM HEPES at pH 7.0. Synchrotron radiation was used to obtain diffraction data to 1.77 Å resolution, a higher resolution than any importin-α structure currently available in the literature. Structure determination was performed using molecular replacement and our (no-SEC) importin-α (70-529) structure as the search model. Manual model refinement was performed in $2mF_o$-DF_c weighted electron density difference maps. The final *apo* model includes residues 72-496 of importin-α (70-529) and 348 waters, refined to an R-factor of 19.2% and an R_{free} of 21.6%.

The *apo* importin-α (70-529) core structure consists of 10 armadillo (ARM) repeats and closely resembles the C-terminal domain of full length importin-α (1-529) (PDB: 1IAL, RMSD of 0.25 Å across 423 C_α atoms). Although there are few differences in the C_α backbone compared to importin-α (1-529), the peptide bond between Lys240 and Asn241 shows evidence for a peptide bond plane-flip which has been included in the final model. There are also a number of side chains that display evidence for dual rotamer configurations including Ser149, Glu153, Ser194, Ser342, Asn350 and Val385. These residues have been modelled with dual side chain conformations at 50% occupancy each. While these rotamers have been observed in previous (no-SEC) importin-α (70-529) structures, the notable exception is the dual conformation of Ser149 that is only present in the *apo* structure. In importin-α:NLS complexes, Ser149 has van der Waals overlaps with the side chain of the NLS residue at P4 and the peptide bond between P3 and P4 residues. The average B-factor of importin-α atoms is 35.5 $Å^2$ and for water atoms it is 44.6 $Å^2$. Importin-α has two unique intermolecular interfaces in the crystal (symmetry mates: x-1/2, -y+1/2, -z and -x, y-1/2, -z-1/2), with a total buried surface area of 639.1 $Å^2$ and 392.1 $Å^2$, respectively.

4.3.1.1 Electron density analysis

By studying $m\mathrm{F_o}$-$D\mathrm{F_c}$ weighted difference maps it is clear that the rogue peptide observed in (no-SEC) importin-α (70-529) has been partially removed from the major binding site (see Figure 4.8). The residual electron density which does still exist in binding slots P1 - P6 appears to be mostly non-peptidic. However, a heterogeneous population of peptide species occupying the binding site in different configurations could also be possible. There are density fragments that more closely resemble peptide features, most notably in the P5 binding slot. It cannot be ruled out that the NLS-like sequence in the artificial *linker* domain, AKFERQ**,** does not contribute to the residual electron density.

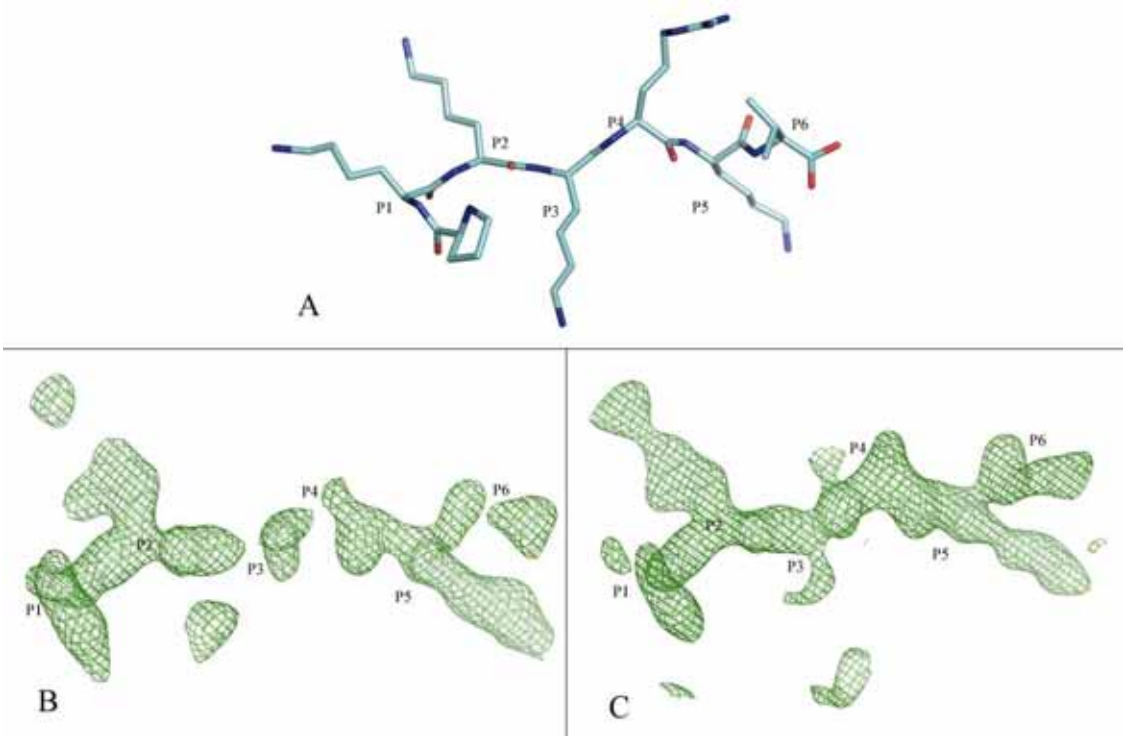

Figure 4.8: Difference maps near the major binding site in *apo* importin-α (70-529) and (no-SEC) importin-α (70-529).

A) The SV40 NLS peptide is shown to give the orientation of density maps in panels B and C. The structure of the importin-α:SV40 NLS peptide complex is described in Section 4.3.2. Binding slots P1 - P6 have been labelled.

B) The $m\mathrm{F_o}$-$D\mathrm{F_c}$ electron density difference map of *apo* importin-α (70-529). Green contours represent positive density at 2.8σ aligned to the SV40 NLS peptide in panel A.

C) The $m\mathrm{F_o}$-$D\mathrm{F_c}$ electron density difference map of (no-SEC) importin-α (70-529) (shown for comparison, structure described in Chapter 3). Green contours represent positive density at 2.8σ aligned to the SV40 NLS peptide in panel A.

The non-peptidic electron density could be explained by the presence of ordered waters and other small molecules in the binding site. Molecules that are present in the crystal solution include: DTT at 10 mM, Na^+ cations at 700 mM, Tris ($C_4H_{11}NO_3$) at 20 mM, Cl^- at 100 mM and the citrate ion $C_3H_5O(COO)_3^{3-}$ at 600 mM. The citrate ion is fully deprotonated in the crystal condition at pH 7.0. Modelling of these molecules in the binding site proves particularly difficult and it is likely that if some small molecules are indeed present then they are likely to be bound in different states and exist alongside a mixture of peptide sequences. What follows now is a more detailed description of each binding slot in the *apo* importin-α (70-529) mF$_o$-DF$_c$ difference map, analysed at a 2.8 σ-level as shown in Figure 4.9.

Electron density in the *apo* importin-α (70-529) major binding site

P1: There is no electron density corresponding to a peptide side chain or C_α atom at P1. There is density beyond P1 that may indicate this part of the peptide is flexible. Side chain density is observed at this position in the importin-α:SV40 NLS structure described in Section 4.3.2.

P2: This site contains reasonably strong non-peptidic density. The normal location of the lysine N^ζ side chain atom is now occupied by what is most likely a water molecule. While the lysine N^ζ atom coordinates 3 hydrogen bonds in importin-α:NLS complexes, the water present in the *apo* structure appears only capable of hydrogen bonding Asp192 and Thr155 with good geometry and bond distances of 2.96 Å and 2.86 Å, respectively. The third bonding partner is the carbonyl oxygen of Gly150, which is 3.44 Å from the water. The water fits the density well with an occupancy greater than 50%.

P3: Main chain and side chain density is very weak at the P3 C_α position and is non-existent until a σ-level of 2.5 in the difference map. The site could be partially occupied by a water molecule where an NLS side chain normally fits. The position of a P3 amide group contains the strongest difference density in the binding site with a peak at 7.8σ. Electron density returns where a P3 carbonyl group could be positioned to make a hydrogen bond to Trp142, but it is this P3 carbonyl group that creates van

der Waals contacts with an alternate conformation of Ser149 only present in *apo* importin-α.

P4: The density here is weak for both the main chain and side chain, corresponding to the second density break in the binding site.

P5: The density at P5 is stronger and resembles peptidic features in both the main chain and side chain. Noticeably P5 contains side chain density that is likely to correspond to a lysine or arginine residue.

P6: Similar to P1 at the N-terminus, the C-terminus contains weak density for the main chain and side chain. The position of the C_α at P6 contains no density until a σ-level of 1.8 in the difference map. The site still appears to be occupied, albeit at a very low level.

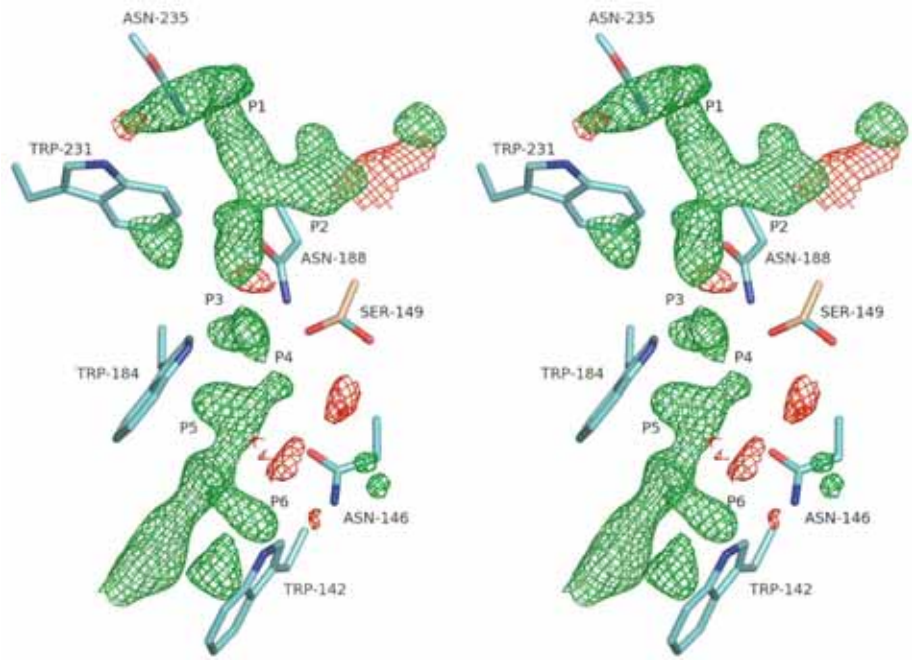

Figure 4.9: A stereo view of electron density in the *apo* importin-α (70-529) major binding site.

A) The mF_o-DF_c difference map is shown with nearby tryptophan and asparagine arrays in stick representation. Positive density is contoured at +2.8σ (green) and negative density is contoured at -2.8σ (red). Ser149 is shown with a secondary conformation coloured in light brown.

The residual electron density observed in the *apo* importin-α (70-529) binding site is difficult to interpret. A poly-alanine chain modelled and fit to the density suggests an occupancy of ~40%, although this could well be accounting for the presence of waters and other molecules. There are regions that appear weak and non-peptidic, while at P5

there is a fairly well defined peptide side chain similar to those observed in (no-SEC) structures. There are also breaks in the main chain density at P1, P3, P4 and P6 at the 2.8σ map level. There is a well defined water molecule at P2 and there is evidence of another near P3 that would create steric overlaps with the C_α atom or side chain of an NLS residue at P3. In the *apo* structure of yeast importin-α (PDB: 1BK5), binding slot P2 is occupied by a Co^{2+} cation, suggesting the site will always be occupied. Density in the P2 slot may suit a short side chain such as serine. Attempts to fit the citrate ion at this location were not successful.

4.3.1.2 Binding site residue flexibility

Another aspect of the *apo* importin-α (70-529) structure is the mobility of residue side chains in the major binding site. We have already mentioned the increased mobility of Ser149, which adopts two conformational states. There is also evidence for a second conformational state in the central tryptophan (Trp184) of the major binding site tryptophan array. The tryptophan array includes residues Trp142, Trp184 and Trp231, which form hydrophobic binding pockets at P3 and P5 that are normally occupied by the long aliphatic side chain portion of an NLS lysine or arginine residue at P5 (see Figure 4.13 in Section 4.3.2.2). Trp184 (and Trp142) adopts a rare rotamer conformation (rotamer score of ~3.5%, calculated using MolProbity which compares the side chain to a high quality reference dataset [159]) that facilitates the formation of the binding slot at P5. The other invariant tryptophans including Trp231, Trp273, Trp357 and Trp399, adopt more common rotamers (rotamer scores of 30-75%). It is feasible that the partial removal of the NLS residue at P5 would alleviate considerable constraints on Trp184 imposed by the side chain at P5, allowing a greater range of side chain motion. The same effect is not observed for Trp142, possibly due to a cation - π interaction with the Arg101 side chain on the opposite side of the P5 pocket.

In the difference electron density maps, an increased flexibility of Trp184 can be inferred from the strong negative density over the centre of the tryptophan indole group, concentrated on the $C^{\varepsilon 2}$ carbon atom at -7.7σ (see Figure 4.10). No significant negative electron density is observed at this location in other structures. This suggests an increased flexibility of Trp184 as it is now free to rotate in space that is normally

occupied by a bound peptide. Trp184 may switch between two rotamers which are related by an ~90° clockwise rotation in the χ_2 angle as seen in Figure 4.10. The second conformational state corresponds to the more common rotamer adopted by the other conserved tryptophans, however its occupancy is low and is estimated at 20%. Supporting evidence for the alternative conformation is the presence of a positive difference density peak at 3.5σ that corresponds to the expected position of the indole $N^{\varepsilon 1}$ atom in the alternative rotamer.

It is worth noting that the Trp184 side chain is normally hydrogen bonded through the $N^{\varepsilon 1}$ atom to the NLS peptide main chain at the carbonyl oxygen of the residue at P3, and Trp142 hydrogen bonds to the carbonyl oxygen of the residue at P5. The secondary conformation of Trp184 is not commensurate with an NLS residue at P3 since it will likely break this hydrogen bond and also sterically clash with NLS side chain and backbone atoms. An increased flexibility is similarly inferred for other residues in the major binding site, including Asn146 and Asn188, that also correspond to region of negative difference density (see Figure 4.10).

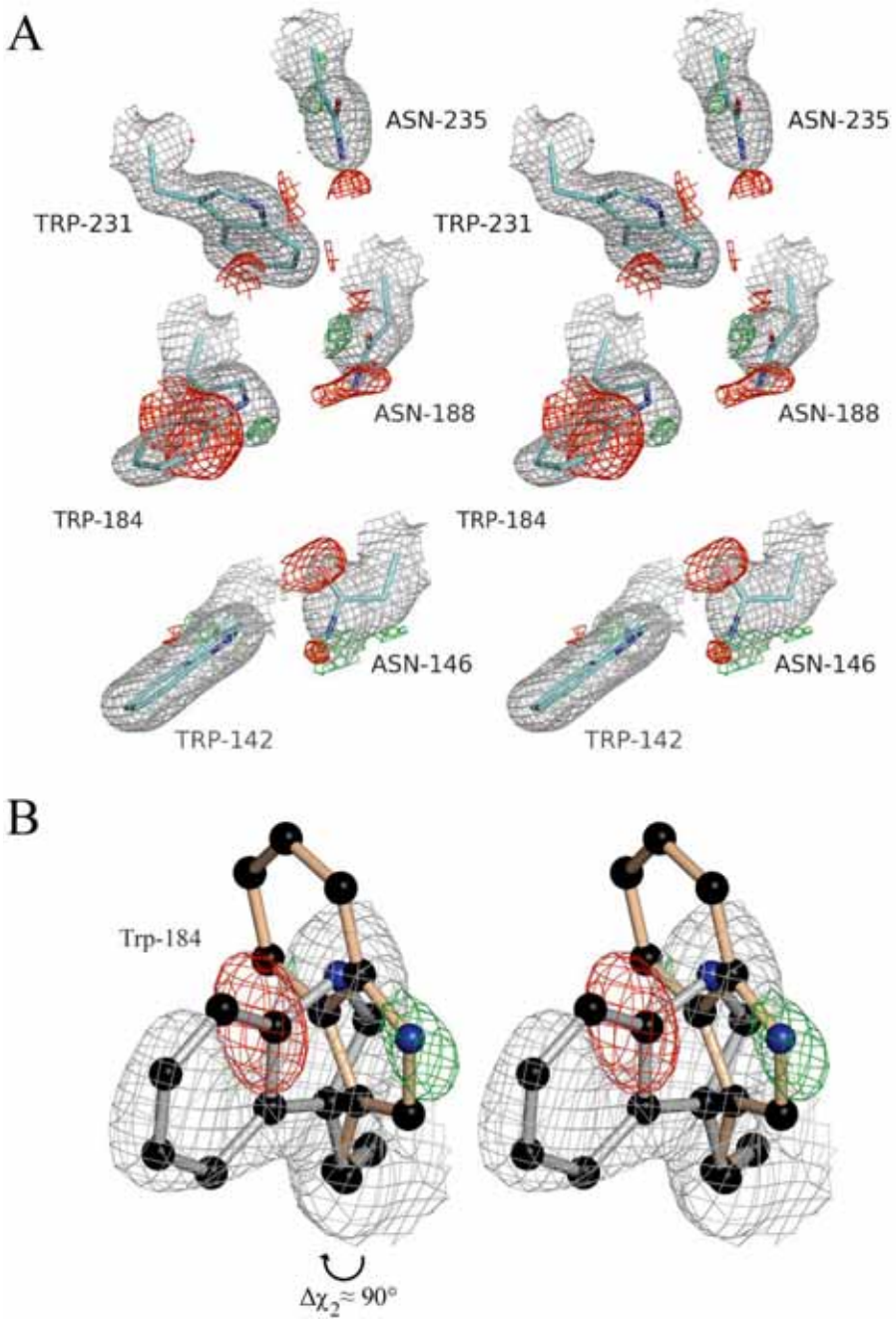

Figure 4.10: Stereo images of the tryptophan and asparagine arrays in the *apo* importin-α (70-529) major binding site.

A) Electron density has been 'carved' at 2.0 Å around tryptophan and asparagine residues which are shown in stick representation. The $2mF_o$-DF_c map is represented by grey contours at 1.5σ. The mF_o-DF_c difference map is represented by positive (green) contours at +2.8σ and negative (red) contours at -2.8σ.

B) The *apo* importin-α (70-529) Trp184 residue shows evidence for a dual side chain conformation. The more common rotamer is coloured in wheat, while the modelled rotamer is shown in grey. The mF_o-DF_c negative density is shown in red and contoured at 4σ, for clarity. Otherwise electron density is contoured as in panel A.

4.3.2 The structure of the importin-α:SV40 NLS peptide complex

The structures of importin-α:NLS complexes were solved after obtaining an *apo* form of importin-α (70-529), which has been described in the Section 4.3.1. By repeating the experiments described in Chapter 3, that resulted in ambiguous electron density for a bound peptide, co-crystallization of importin-α (70-529) and the SV40 NLS (^{126}PKKKRKV^{132}) was again used to successfully solve the structure of the importin-α:SV40 NLS complex. This structure serves as an important experimental control confirming that NLS peptides are now able to freely bind importin-α (70-529) without interference from rogue *E. coli* proteins.

Using the same importin-α (70-529) sample from which repeated *apo* structures were determined, we crystallized the importin-α:SV40 NLS peptide complex and determined its structure at 1.85 Å. This corresponds to the highest resolution structure of any importin-α complex currently deposited in the Protein Data Bank [187]. The core component of the importin-α (70-529) molecule includes residues 72-495 and is virtually identical to previously solved structures in this thesis (RMSD with *apo* importin-α (70-529), 0.22 Å over 424 C_α atoms in residues 72-495). The SV40 NLS peptide main chain is virtually identical to previous importin-α:SV40 NLS structures (PDB:1Q1S, RMSD 0.14 Å over six C_α atoms) [67], although there are various side chain conformational differences observed in our structure which are described in Section 4.3.2.2.

4.3.2.1 Electron density analysis

Due to the problems regarding a rogue peptide binding (no-SEC) importin-α (70-529), establishing beyond reasonable doubt the identity of the bound peptide in the importin-α:SV40 NLS complex via electron density analysis was an overriding concern. Comparing mF$_o$-DF$_c$ electron density with (no-SEC) importin-α structures, the position of the SV40 NLS peptide main chain is identical, as expected, but differences are evident in nearly all side chain binding slots P1 to P6 (see Figure 4.11). For example, while long side chain density does exist at the key binding site P2 in all

structures, it is only in the importin-α:SV40 NLS structure that this density defines the zigzag aliphatic portion of a lysine side chain. The water molecule observed in *apo* importin-α (70-529) is clearly no longer present. At P1 there is a stretch of side chain density where previously there was none. Binding slots P3 and P4 now contain strong side chain density compared with weak density in the *apo* structure at these locations. Notably there are no breaks in the peptide main chain density at the 2.8σ map level.

The SV40 NLS peptide model fits electron density well as shown in Figure 4.11. All SV40 NLS residues (^{126}PKKKRKV^{132}) are accounted for in the density, including the C-terminus carboxylate moiety and the N-terminus proline residue. The average peptide B-factor, 24.5 Å^2, is lower than the average importin-α (70-529) B-factor of 30.8 Å^2, highlighting the rigid binding of the SV40 NLS peptide in the major binding site. A lower average B-factor for the bound SV40 NLS peptide is not uncommon, (PDB:1Q1S and 1EJL) and indeed may be expected for a valid importin-α:SV40 NLS complex. The SV40 NLS peptide buries an accessible surface area of 547.8 Å^2 on importin-α (70-529), with the largest contribution from Trp142 (42.9 Å^2), Trp184 (72.1 Å^2), Trp231 (69.3 Å^2), Ser105 (34.0 Å^2), Ser149 (48.7 Å^2) and Arg238 (55.7 Å^2).

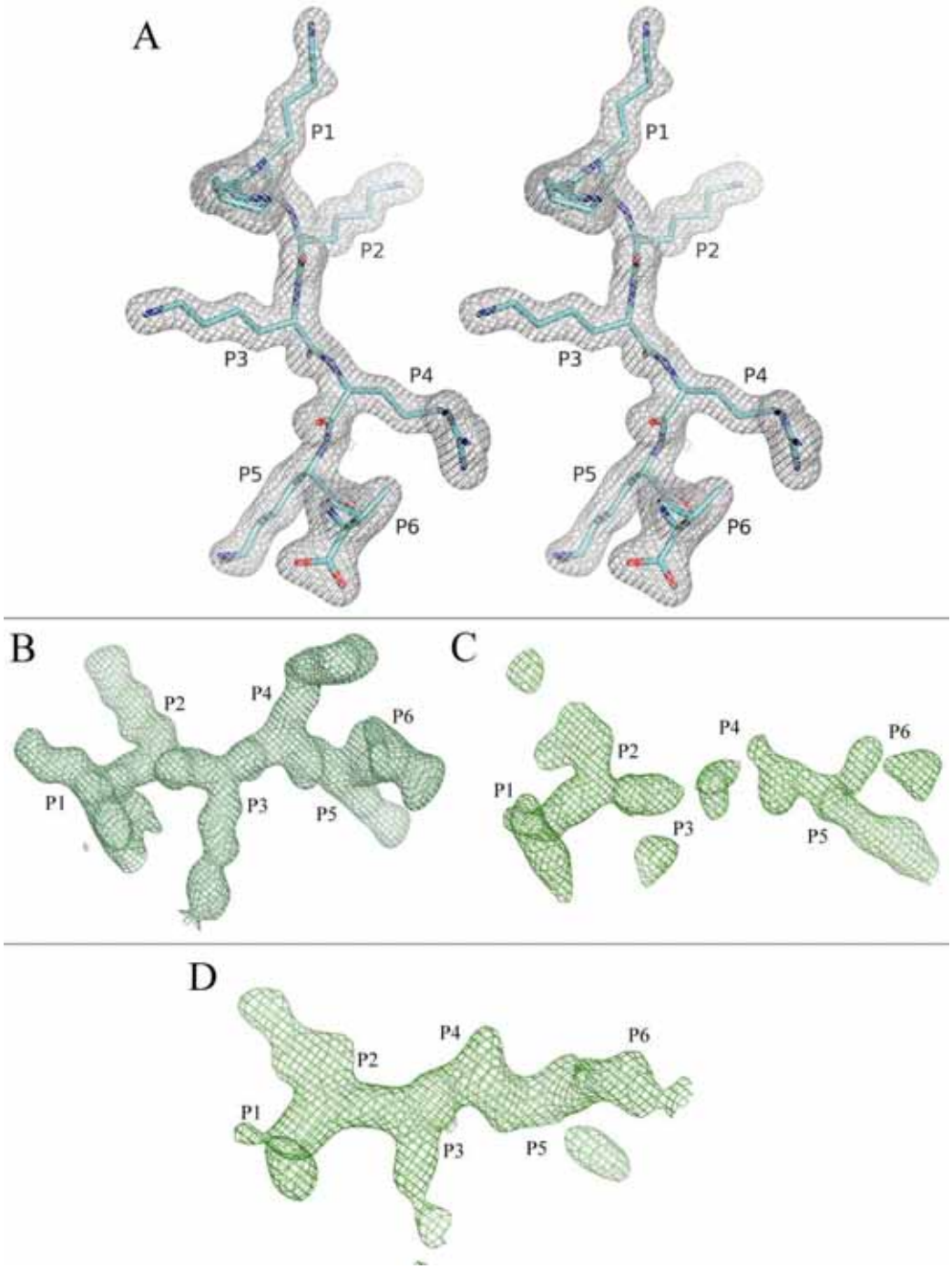

Figure 4.11: NLS peptide electron density in the importin-α:SV40 NLS complex.

A) A stereo image of the importin-α:SV40 NLS structure solved at 1.85 Å, showing the $2mF_o$-DF_c electron density map contoured at 1.5σ in grey. The SV40 NLS peptide is shown in stick representation and binding slots P1 - P6 have been labelled.

B) A simulated annealing omit map (coefficients mF_o-DF_c) of the importin-α:SV40 NLS complex starting from a temperature of 5000 K. Green contours represent positive density at 2.8σ aligned to the SV40 NLS peptide in panel A.

C) The mF_o-DF_c difference map of *apo* importin-α in the region of the major binding site. Green contours represent positive density at 2.8σ.

D) The mF_o-DF_c difference map of the (no-SEC) importin-α:SV40 structure described in Chapter 3. Green contours represent positive density at 2.8σ.

Comparing difference density to *apo* importin-α (70-529) and (no-SEC) importin-α:SV40, we see a strengthening of side chain density in all binding slots (Figure 4.11, panels B, C and D). The aliphatic portion of lysine side chains in P2, P3 and P5 are present with a characteristic zigzag extended conformation clearly defined in electron density maps ($\chi_2 \sim \chi_3 \sim \chi_4 \sim 180°$). Density at P1 corresponds to a lysine side chain up to the C^{δ} atom, after which the side chain appears mostly disordered. A valine at P6 fits the density well also. The P4 binding slot contains unambiguous density corresponding to an arginine residue (Arg130) at this location, a distinguishing feature of the SV40 NLS.

4.3.2.2 SV40 NLS rotamer conformations at P4 and P5

Although the P4 arginine side chain has a similar conformation up to the C^{δ} atom, as observed previously [67], our higher resolution maps reveal the guanidinium head group has a distinctly different orientation (see Figure 4.12). By changing the Arg130 side chain conformation, the residue adopts a more favourable rotamer with a score of 56.2% compared to 1.4% [159]. This improved rotamer also reveals an intricate network of well coordinated hydrogen bonds formed between the Arg130 side chain and the importin-α main chain carbonyl groups of Leu104, Arg106 and Glu107, which are part of the connecting loop between ARM1 and ARM2. Hydrophobic contacts with importin-α at Pro110, Ser149 and Asn146 are not significantly affected by changes in the arginine conformation.

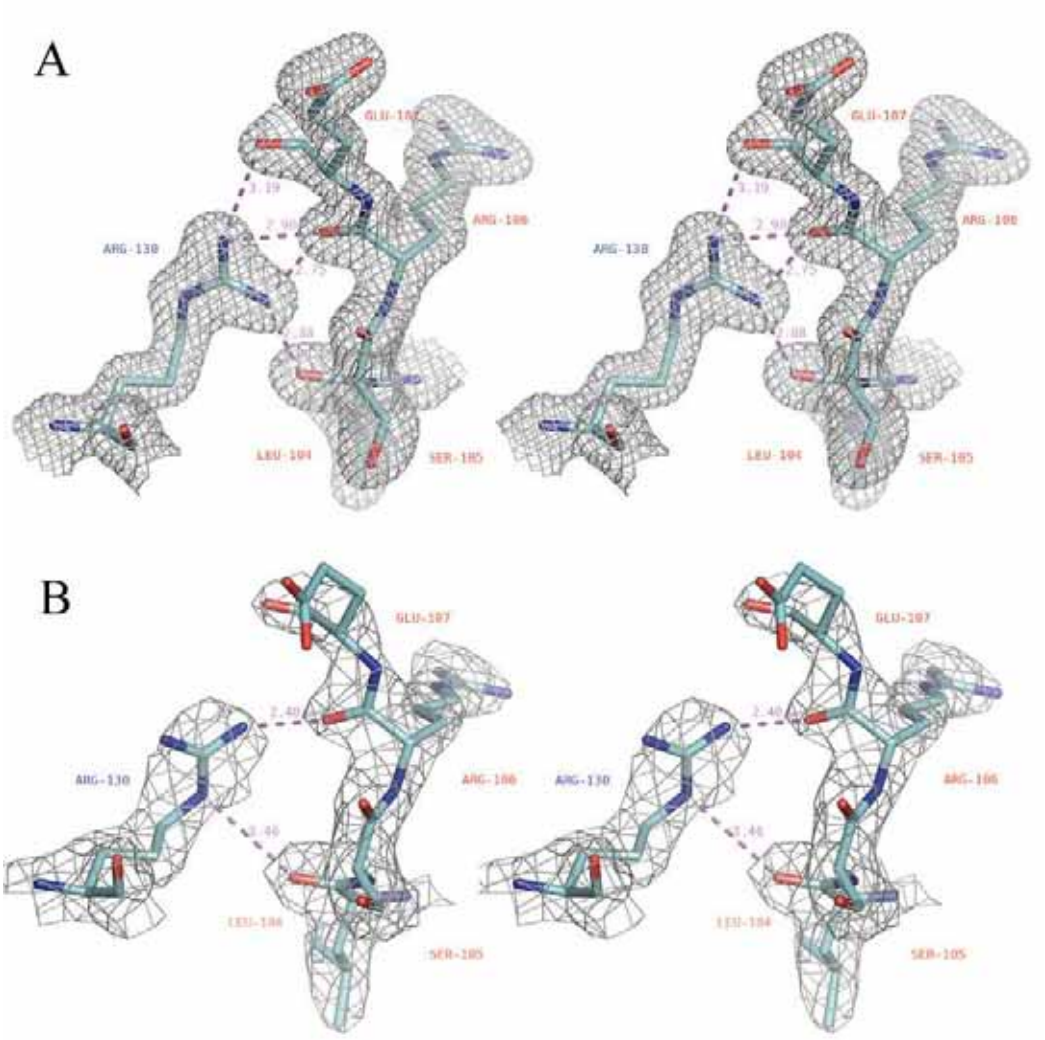

Figure 4.12: Stereo images of the SV40 NLS Arg130 side chain interacting with importin-α (70-529) in binding slot P4.

A) Shown here is the importin-α:SV40 NLS complex solved at 1.85 Å, described in this chapter. The $2mF_o$-DF_c map is contoured at 1.5σ in grey. Polar contacts are shown by dashed lines (magenta), with the corresponding bond distance in ångströms. Importin-α residues are labelled in red, SV40 NLS peptide residues are labelled in blue.

B) The highest resolution importin-α:SV40 NLS structure available in the PDB (PDB: 1Q1S, solved at 2.3 Å resolution [67]). The $2mF_o$-DF_c map has been reconstructed using structure factors deposited in the PDB and is contoured at 1.5σ in grey, at a lower grid sampling rate than the map in Panel A.

In binding slot P5, there is a clear distinction of dual conformations for the SV40 NLS residue Lys131, which have been included in the final model as dual rotamers with occupancies of 50% each (see Figure 4.13, panel A). An arginine side chain cannot substitute for side chain conformational states of Lys131. The two rotamers are related by a rotation in χ_4 of ~100°. The regular conformation, ^{a}Lys131, is a common rotamer with a score of 83%, while the alternate rotamer, ^{b}Lys131 has a score of 65% [159]. ^{a}Lys131 forms hydrogen bonds with $^{imp\text{-}\alpha}$Gln181 and two water molecules,

while [b]Lys131 only hydrogen bonds with two water molecules. The potential long range electrostatic interaction with [imp-α]Glu180 is decreased in [b]Lys131, with a slightly increased N^{ζ} - $O^{\varepsilon 1}$ distance of 6.5 Å compared with 6.1 Å for [a]Lys131. The van der Waals contacts between Lys131 and [imp-α]Trp142/Trp184 are extensive, as shown in Figure 4.13, panel B. Integrating over the van der Waals contact area [159, 188] gives a value of 26.0 $Å^2$, equally shared on the Trp142 and Trp184 van der Waals surface. The dual rotamer conformation of Lys131 has been noted previously [178], but it is not included in any published models.

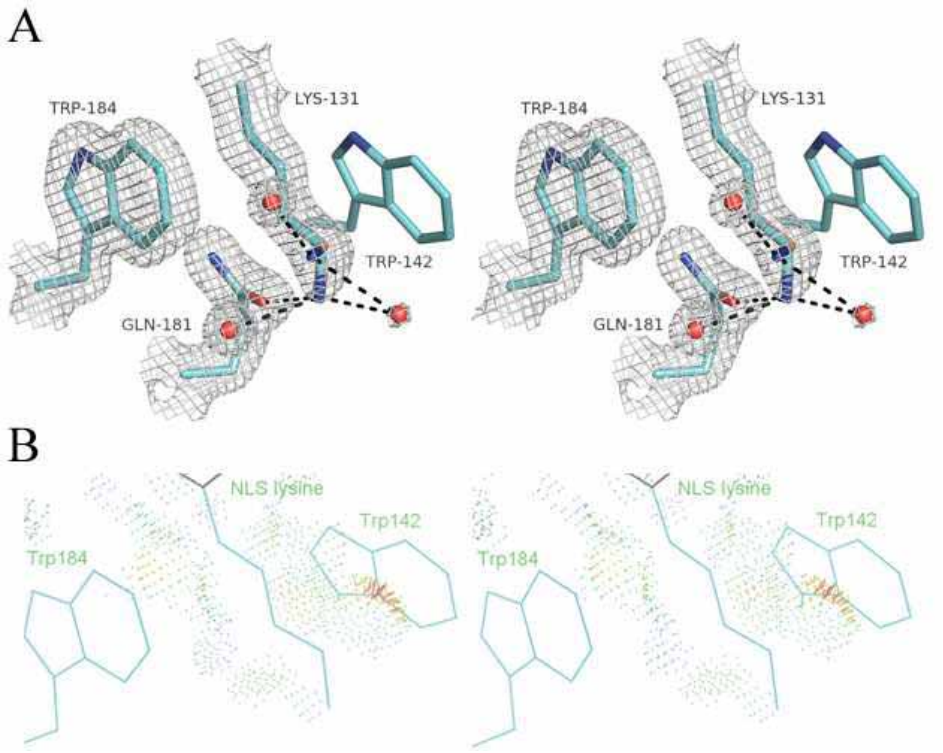

Figure 4.13: Stereo image of the P5 binding slot in the importin-α:SV40 NLS peptide complex.

A) The dual conformations of Lys131 in the SV40 NLS are shown, including nearby importin-α residues and ordered waters. Hydrogen bonds are represented by dashed lines. The $2mF_o$-DF_c map is shown in grey, contoured at 1.8σ. For clarity the map is not contoured around Trp142.

B) Van der Waals contacts between [SV40]Lys131 and [imp-α]Trp142/Trp184. Blue dots represent van der Waals separation at 0.5 Å, green dots at 0.25 Å and yellow and red are contact overlaps. Image created using kinemage [189] and MolProbity [159].

4.3.2.3 The SV40 NLS peptide in the minor binding site

The importin-α:SV40 NLS peptide complex also contains electron density for a peptidic molecule in the minor binding site, although it is markedly weaker than that seen in the major binding site (see Figure 4.14). An abridged SV40 NLS was built in the region with sequence ^{128}KKRKV^{132}. The peptide has an average B-factor of 54.9 Å^2 and buries an accessible surface area of 368.0 Å^2 on importin-α.

The minor binding site contains two key binding slots P1′ and P2′, which are most favourably occupied by a lysine and arginine respectively, as determined by mutagenesis studies of bipartite NLSs [61, 74]. In our monopartite SV40 NLS structure we have followed this consensus by modelling Lys129 at P1′ and Arg130 at P2′. This was also found to be the case in more recently solved importin-α:SV40 NLS complexes such as PDB:1Q1S [67]. However it differs from the first mouse and yeast importin-α:SV40 NLS structures (PDB: 1EJL and 1BK6) which placed Lys128 and Lys129 at P1′ and P2′ respectively [30, 178]. In our structure, ^{SV40}Lys129 forms an electrostatic interaction with $^{imp\text{-}\alpha}$Asp325, and ^{SV40}Arg130 forms an electrostatic interaction with $^{imp\text{-}\alpha}$Glu396.

We note that Arg130 in binding slot P2′ is parallel to the indole side chain of Trp399 and likely forms a cation - π bond. The bonding distance from the plane of the guanidinium group to the aromatic plane is in the range 3.6 - 4.0 Å, with an inter-planar angle of ~11°. This form of cation - π interaction, involving parallel arginine and tryptophan side chains, is commonly found in protein structures [190].

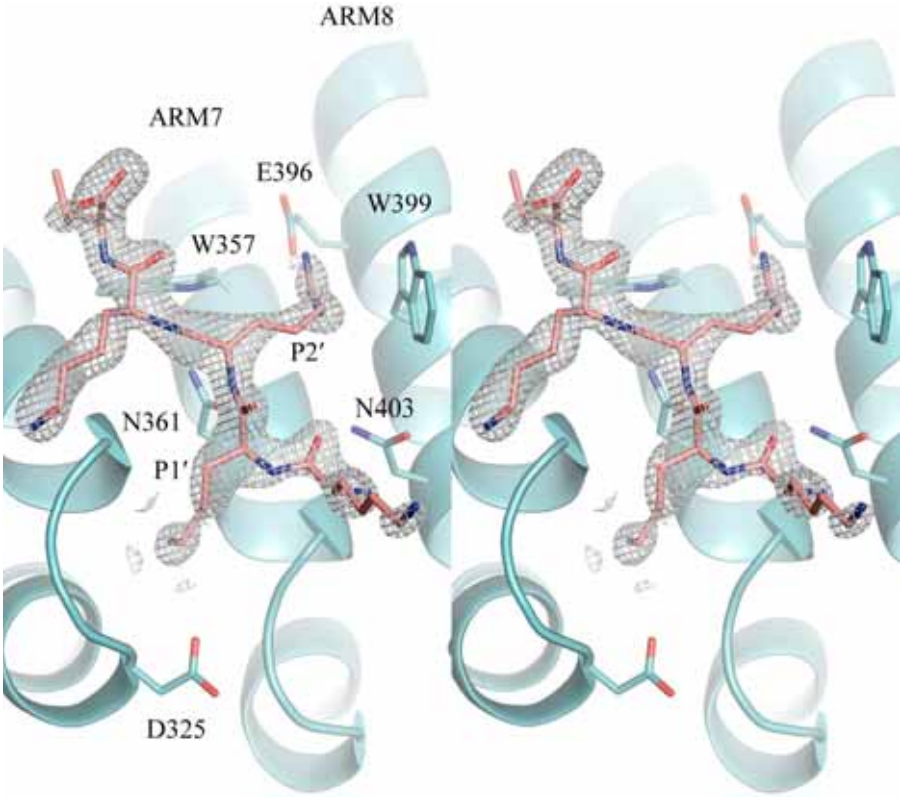

Figure 4.14: Stereo image of the SV40 NLS peptide in the importin-α (70-529) minor binding site.

The SV40 NLS peptide including residues KKRKV, is shown in stick representation (light red) bound to importin-α (70-529) shown in cartoon representation (cyan). Importin-α minor binding site residues are shown as sticks. The $2mF_o$-DF_c map is shown in grey, contoured at 1.2σ.

4.3.2.4 The data-data difference Fourier

To confirm the presence of the SV40 NLS peptide at the major binding site, we examined the data-data difference Fourier, $F_o^{sv40} - F_o^{apo}$, between the importin-α:SV40 NLS complex and *apo* importin-α (70-529). The calculation of this Fourier is explained in Chapter 2. The map shows strong residual positive density in the major binding site consistent with the SV40 NLS, accounting for the presence of all residues in the peptide, ^{126}PKKKRKV^{132} (see Figure 4.15A). There are no gaps in the density along the length of the main chain at a map level of 2.8σ, although there is a weak point near the amide nitrogen of Lys129 at binding slot P3. This position corresponds to a positive density peak in the *apo* importin-α (70-529) difference map. Side chain density is also well represented although it is not as clear compared with the $2mF_o$-DF_c map. Density at P5 is not well defined for a lysine residue and at P2 there is only weak

density near the lysine N^{ζ} atom. The weakening of density in the $F_o^{sv40} - F_o^{apo}$ map is to be expected since the *apo* reflection data that is being subtracted still contains information about the presence of small molecules in the binding site and at least one amino acid at P5. At P2, the *apo* structure contains an ordered water molecule in a position that corresponds to the lysine N^{ζ} atom.

The most identifiable characteristic in the $F_o^{sv40} - F_o^{apo}$ map is the strong density that corresponds to the SV40 NLS Arg130 residue at P4 (see Figure 4.15B). The density closely defines the arginine side chain conformation despite the subtraction of *apo* importin-α data. The $F_o^{sv40} - F_o^{apo}$ data-data difference map suggests the unoccupied P4 site in *apo* importin-α has become occupied by the SV40 NLS peptide added during co-crystallization. The unequivocal presence of an arginine at P4 provides strong supporting evidence that the SV40 NLS peptide is binding importin-α (70-529).

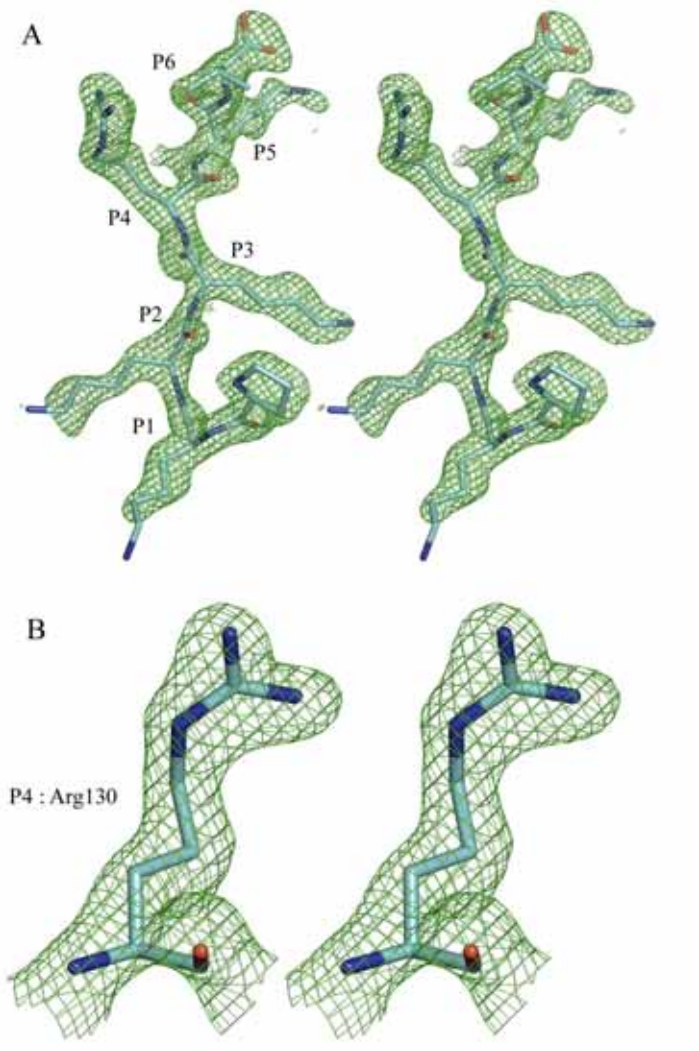

Figure 4.15: Stereo views of the $F_o^{sv40} - F_o^{apo}$ data-data difference Fourier map.

A) The importin-α:SV40 NLS peptide is overlaid on the $F_o^{sv40} - F_o^{apo}$ data-data difference map. Green map contours represent positive density at 2.8σ. The SV40 NLS peptide is shown in stick representation.

B) The $F_o^{sv40} - F_o^{apo}$ map zoomed in on the P4 binding slot occupied by the SV40 NLS Arg130 residue.

4.3.2.5 Normalised B-factor analysis

We have inspected the importin-α:SV40 NLS structure for changes in residue flexibility due to the increased constraints placed upon them by the presence of a bound SV40 NLS peptide. This has been examined through a comparison of atomic B-factors in the importin-α:SV40 NLS structure and the *apo* importin-α (70-529) structure. The flexible residues are often apparent in the electron density difference map, but it is simpler to quantify the effect by observing changes to the model's atomic B-factors which increase with higher atomic disorder. B-factor analysis can be a problematic subject since there are many variables that affect atomic disorder. We have tried to overcome these problems by analysing normalised B-factors, B_z, which are described in Section 4.2.7.1. To compare B_z values in the importin-α:SV40 NLS and *apo* importin-α (70-529) structure, we have subtracted the two to give the difference in normalised atomic B-factors, and then normalised these differences to give a value we refer to as the B_z^{-apo} score. The B_z^{-apo} score has a zero mean and a standard deviation of 1 (see Figure 4.16).

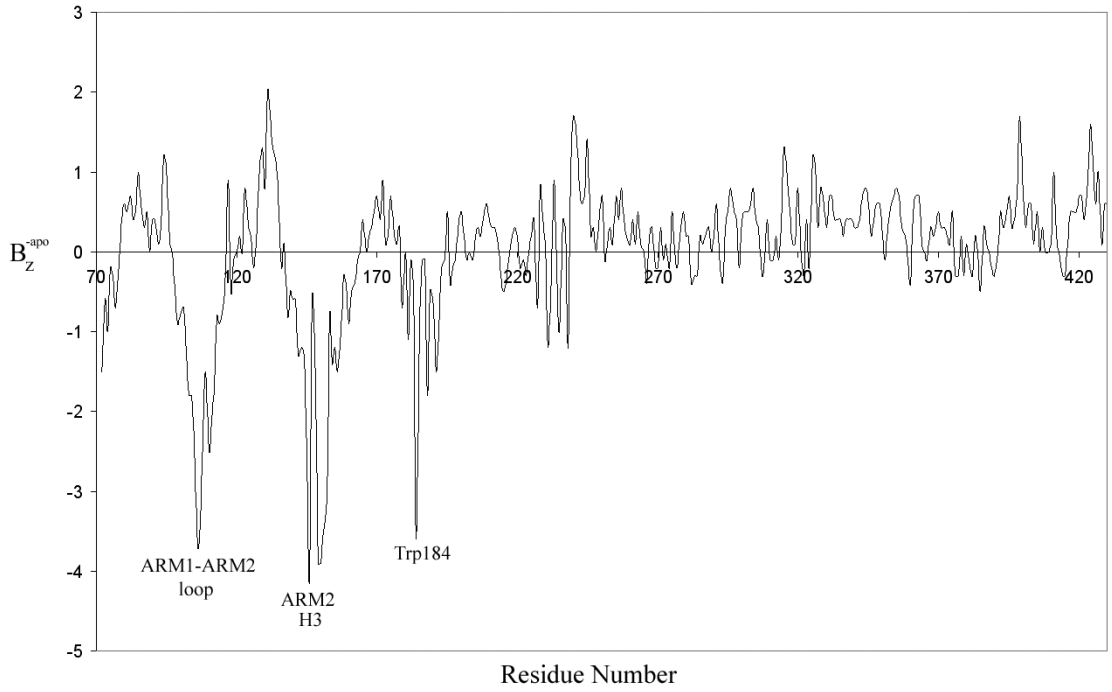

Figure 4.16: A plot of normalised B-factor scores for the importin-α:SV40 NLS complex.

Scores for residues 72-430 are plotted. The B_z^{-apo} value plotted here has been averaged over atoms in each residue. Three peaks have been labelled corresponding to their position within importin-α: the ARM1-ARM2 connecting loop, the ARM2:H3 helix and Trp184 in ARM3.

The B_z^{-apo} scores for the importin-α:SV40 NLS structure are also shown in Figure 4.17, coloured on the molecular surface of importin-α from -3 to +3 normalisation units or standard deviation units. The scores have been averaged for each importin-α residue over main chain (mc) and side chain (sc) atoms. Negative B_z^{-apo} scores are shown in shades of blue and high B_z^{-apo} scores are shown in shades of magenta. This excludes residues 431-496 that are considered outliers using this B-factor analysis, which have been coloured grey. The scores represent changes to the average isotropic displacement of an atom that should only arise due to the binding of the SV40 NLS peptide. For the most negative scores, this equates to a decrease of the atoms average isotropic displacement or x_j, by ~0.3 Å (refer to Section 4.2.7.1).

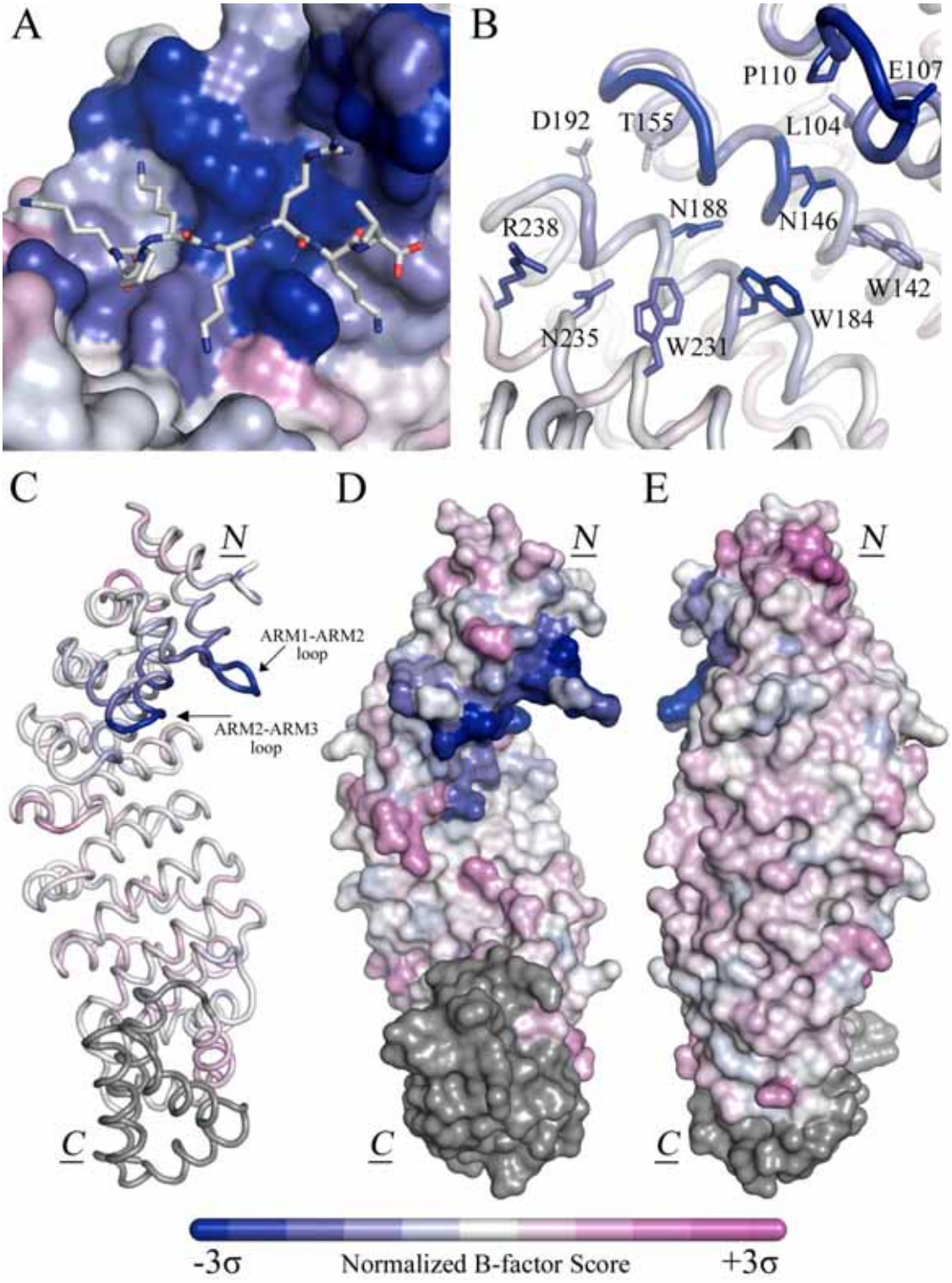

Figure 4.17: The importin-α:SV40 NLS peptide complex coloured by a normalised B-factor score, B_z^{-apo}.

The B_z^{-apo} colour spectrum is blue-white-magenta from -3σ to +3σ as shown in the figure. Outliers near the C-terminus are shown in grey.

A) The importin-α major binding site showing the bound SV40 NLS peptide. **B)** Important side chains are shown in the major binding site. **C)** Cartoon tube representation of the importin-α backbone. **D)** The importin-α inner concave surface. **E)** The importin-α outer convex surface.

We see from Figure 4.17 that atomic B-factors do indeed change upon binding the SV40 NLS peptide and the majority of significant negative scores are clustered near the importin-α major binding site (see Figure 4.17A). By comparison the minor binding site has no significant B_z^{-apo} scores. The major binding site tryptophan and asparagine arrays, which have negative difference density in the *apo* structure, have significantly lower B_z^{-apo} scores representing a decrease in residue mobility. These include residues whose side chains participate directly in ligand binding such as Asn146 (-5.2^{sc}), Asn188 (-2.8^{sc}), Asn235 (-2.0^{sc}), Trp142 (-1.4^{sc}), Trp184 (-4.4^{sc}) and Trp231 (-1.7^{sc}), which all have lower than expected B-factor values. The side chain of Arg238 (-2.1^{sc}) in the H3 helix of ARM4 also has a lower B_z^{-apo} score, presumably stabilised due to significant van der Waals contacts with the NLS peptide lysine side chain at P1.

While the presence of a critical lysine N^{ζ} atom at P2 does not significantly decrease B_z^{-apo} scores for side chain hydrogen bonding partners Thr155 (-0.9^{sc}) or Asp192 (-1.0^{sc}), there is a noticeable effect on the third hydrogen bonding partner, the carbonyl oxygen of Gly150 (-3.9^{mc}). This supports the *apo* importin-α model which includes a water molecule at the P2 site, since this water hydrogen bonds Thr155 and Asp192 leaving a non-bonded Gly150 with a higher B-factor.

Due to the creation of the lysine N^{ζ}:Gly150 hydrogen bond in the importin-α:SV40 structure the loop connecting ARM2 and ARM3 including residues Ala148 to Ser152, has reduced flexibility (see Figure 4.17C). A similar effect is noticed in binding slot P4 where the NLS peptide arginine (Arg130) side chain makes significant van der Waals contacts with Pro110 (-2.5^{res}) and is hydrogen bonded with Leu104 (-1.8^{res}), Arg106 (-3.7^{res}) and Glu107 (-3.4^{res}). Not only do these residues have a lower B_z^{-apo} score, but the entire loop connecting ARM1 and ARM2 in importin-α including residues Leu103 to Ile112.

This B-factor analysis appears capable of providing additional insight into the mode of NLS peptide binding and supports the identification of the SV40 NLS in our structure. In conjunction with the data-data difference Fourier which compares our importin-α:SV40 NLS structure with our *apo* importin-α structure, it is clear that we

have progressed from a relatively empty binding site to one which is occupied with our SV40 NLS peptide added during co-crystallization. Being able to unambiguously identify residue side chains marks an improvement in data quality and data interpretation, something that proves useful for the slightly weaker interaction posed between importin-α (70-529) and the CLIC4 NLS, described in Chapter 5.

4.3.3 The structure of the importin-α:30S ribosomal subunit S21 complex

After successfully separating importin-α (70-529) from *E. coli* contaminant proteins on SEC, we solved the structure of *apo* importin-α (70-529) as described in Section 4.3.1. The sample of importin-α (70-529) containing *E. coli* contaminants (shown in Figure 4.7 and Figure 4.24) was also crystallized in a buffer condition similar to that used in the crystallization of *apo* importin-α (70-529). Here we describe the 1.95 Å resolution structure of importin-α (70-529) bound to a rogue peptide identified as the *E. coli* 30S ribosomal subunit S21. The final model consists of residues 70-496 of importin-α (70-529) and 7 residues of S21, ^{44}ERKRAKA^{50}. Other endogenous *E. coli* proteins in the sample, such as the 30S ribosomal subunit S11, have also been identified using mass spectrometry and are described in Section 4.3.5.

In the importin-α:S21 complex, the ARM repeat module of importin-α is comparable to that seen in previous structures such as *apo* importin-α (70-529) (RMSD deviation of 0.20 Å across 425 C_α atoms in residues 72-496). The minor binding site contains very weak density that may correspond to a bound peptide, however since it is ambiguous no model is included in the final structure. Within the importin-α major binding site there is clearly defined density for a bound peptidic ligand (see Figure 4.18B). In comparison to the *apo* structure (Figure 4.18C) there is stronger $m\mathrm{F_o}$-$D\mathrm{F_c}$ electron density in all side chain binding slots and across the peptide main chain, which has no breaks in the density at 2.8σ.

4.3.3.1 Electron density analysis

Identification of contaminant *E. coli* proteins by mass spectrometry revealed a number of rogue peptide candidates (see Table 4.3, page 189). We notice the electron density in the major binding site closely matches an NLS-like motif, ^{45}RKRAKA^{50}, in one of these candidates: the 30S ribosomal subunit S21. The model of the S21 motif produces a good fit to the $2mF_o$-DF_c electron density (Figure 4.18A). The S21 peptide main chain aligns in the binding site similar to the SV40 NLS peptide (RMSD 0.10 Å over 6 C_α atoms in residues occupying P1 - P6). The S21 NLS-like motif has an average atomic B-factor of 40.2 Å^2 compared with the average atomic B-factor for importin-α of 32.7 Å^2. It buries 569.6 Å^2 of solvent accessible surface area on importin-α.

In our importin-α:*S21* structure we again see well defined density in the P2 binding slot corresponding to a lysine residue. There is no longer the presence of a water molecule at P2 that was present in the *apo* structure, and an unaccounted density globule in *apo*, near the P2 side chain, is no longer present. The P3 binding slot now contains strong main chain density and also displays density for a long side chain, which should correspond to an arginine residue in the S21 NLS-like motif. The S21 peptide model places the arginine side chain in an almost identical orientation at P3 compared with the arginine in the full length importin-α structure (PDB:1IAL), but it is dissimilar to the P3 arginine in the yeast importin-α:c-myc complex (PDB:1EE4) which has a more extended conformation. The P4 binding slot has density that closely defines an alanine residue, a distinguishing characteristic of the S21 NLS-like motif. Density at P5 is well accounted for with a lysine residue. Main chain density at P6 is strong and side chain density supports the presence of an alanine. At P1 we see an arginine side chain does not fit the electron density past the C_β atom, although similar disorder is observed in the lysine side chain at P1 in our importin-α:SV40 NLS peptide complex. A glutamic acid side chain at the N-terminus of the motif (Glu44) also shows signs of significant disorder.

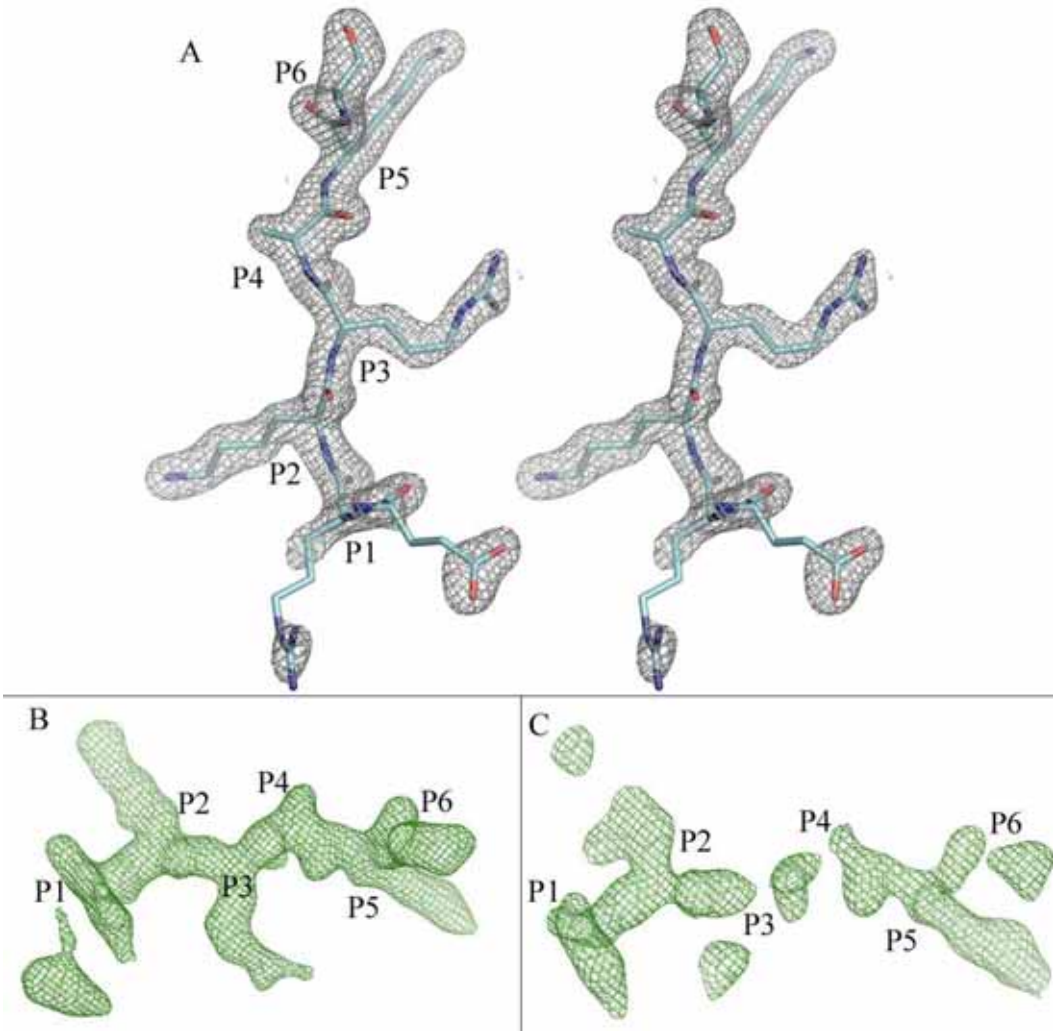

Figure 4.18: Electron density in the major binding site of the importin-α:S21 complex.

A) A stereo image of the S21 NLS bound to importin-α (70-529) in the major binding site. The $2mF_o$-DF_c weighted electron density map is shown in grey, contoured at 1.5σ.

B) The mF_o-DF_c difference electron density or 'omit' map of the importin-α:S21 major binding site, contoured at 2.8σ.

C) The mF_o-DF_c difference map in the *apo* importin-α (70-529) major binding site, contoured at 2.8σ.

4.3.3.2 A disulfide bond at Cys133 in importin-α

There is clear evidence of a disulfide bond at Cys133 in our importin-α:S21 complex, which is present in the connecting loop between H2 and H3 helices of ARM2. In the crystal, this region of the protein forms an interface with the outer convex surface of a symmetry related molecule near ARMs 5-6 (see Figure 4.19). S21 contains one cysteine residue, Cys23, which may facilitate the formation of the

disulfide bond. The C-terminus of the S21 NLS-like motif bound to the major binding site is in close proximity to Cys133. A long and flexible chain extending away from the peptide N-terminus of a neighbouring symmetry molecule could possibly reach the site, however it requires a chain longer than ^{S21}Cys23 - ^{S21}Glu44 in S21 and thus it is difficult to foresee how such a mechanism may work. Electron density corresponds to a sulfur atom at 80% occupancy with a refined B-factor of 41.8 Å^2. The disulfide bridge has an S^{γ}- S^{γ} bond length of 2.13 Å and a C^{β}- C^{β} distance of 3.95 Å. Further from the bonded sulfur atom electron density is not clear and cannot be used to unambiguously identify a cysteine residue. There is no density corresponding to an extended chain of residues linking the sulfur, therefore if it is present it is disordered in the crystal.

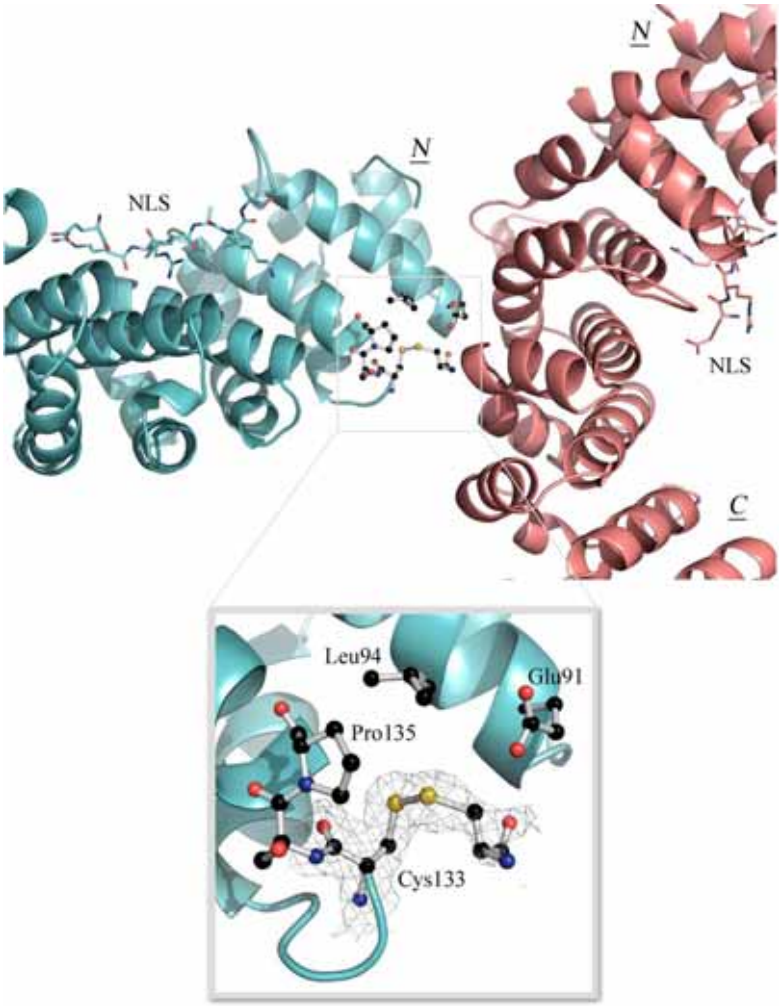

Figure 4.19: A disulfide bond at Cys133 in the importin-α:S21 crystal complex.

Importin-α (70-529) is shown in cartoon representation in cyan, with a symmetry molecule coloured light red. The S21 NLS-like motif bound in the major binding site is shown in stick representation.

Inset: The 2*m*F_o-*D*F_c electron density map is shown in grey, contoured at 1.5σ. A free cysteine residue has been modelled forming a covalent disulfide bond with Cys133.

4.3.3.3 The data-data difference Fourier

To confirm the presence of the S21 peptide, we examined the $F_o^{S21} - F_o^{apo}$ data-data difference Fourier, as described in Section 4.2.7.1. The $F_o^{S21} - F_o^{apo}$ difference map shows residual density for a peptide in the binding site (see Figure 4.20), although the density is not particularly strong and cannot unambiguously define S21 side chains. This is in contrast to the importin-α:SV40 NLS $F_o^{sv40} - F_o^{apo}$ difference Fourier map, which shows well defined side chains particularly in binding slots P3 and P4. This discrepancy could be the result of S21 binding importin-α with a slightly lower occupancy compared to the SV40 NLS peptide. Although we were unable to unambiguously model the S21 sequence with an occupancy less than 1, weak negative electron density centred on the Trp184 side chain indicates the site is not fully occupied. Since we have only crystallized importin-α (70-529) with a mixture of proteins that includes S21, it may not be unexpected that the occupancy of the binding site is lower and hence the data-data difference electron density weaker than in the case of the importin-α:SV40 NLS complex.

At a map contour level of 2.0σ the $F_o^{S21} - F_o^{apo}$ map clearly defines a peptide main chain (Figure 4.20). P3 density corresponds to a long side chain which matches the orientation of the S21 residue, Arg47. There is a break in the main chain density at P4 and the Ala48 side chain does not unambiguously correspond to weak density here. Binding slots P2 and P5 understandably do not contain strong density due to the presence of density at these locations in the *apo* importin-α (70-529) structure.

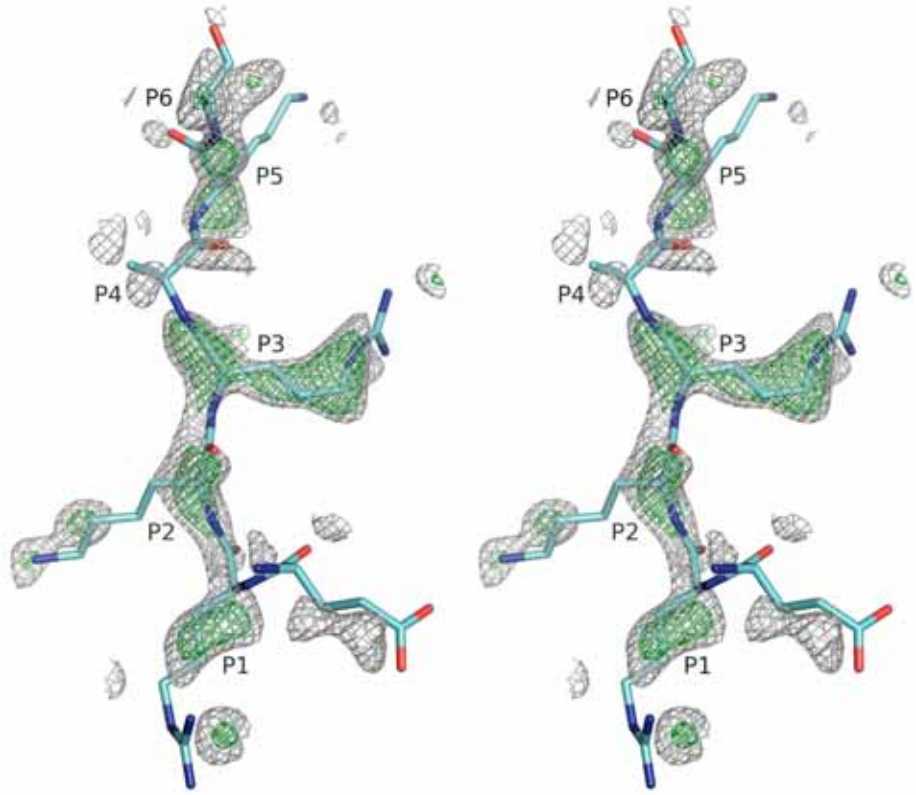

Figure 4.20: Stereo image of the $F_o^{S21} - F_o^{apo}$ data-data difference Fourier map.

Residues in the S21 NLS-like motif are overlaid on the $F_o^{S21} - F_o^{apo}$ difference map in the importin-α major binding site. Green contours represent positive difference density at 2.8σ, and grey contours represent positive difference density at 2.0σ.

4.3.3.4 Normalised B-factor analysis

Similar to the effect observed in the importin-α:SV40 NLS complex, the binding of S21 places tight constraints on residues in the major binding site. This was analysed by comparing normalised atomic B-factors from importin-α:S21 and *apo* importin-α (70-529), as described in Section 4.2.7.1. From these calculations we obtain a value we refer to as the B_z^{-apo} score, which has a zero mean and standard deviation of 1 (see Figure 4.21).

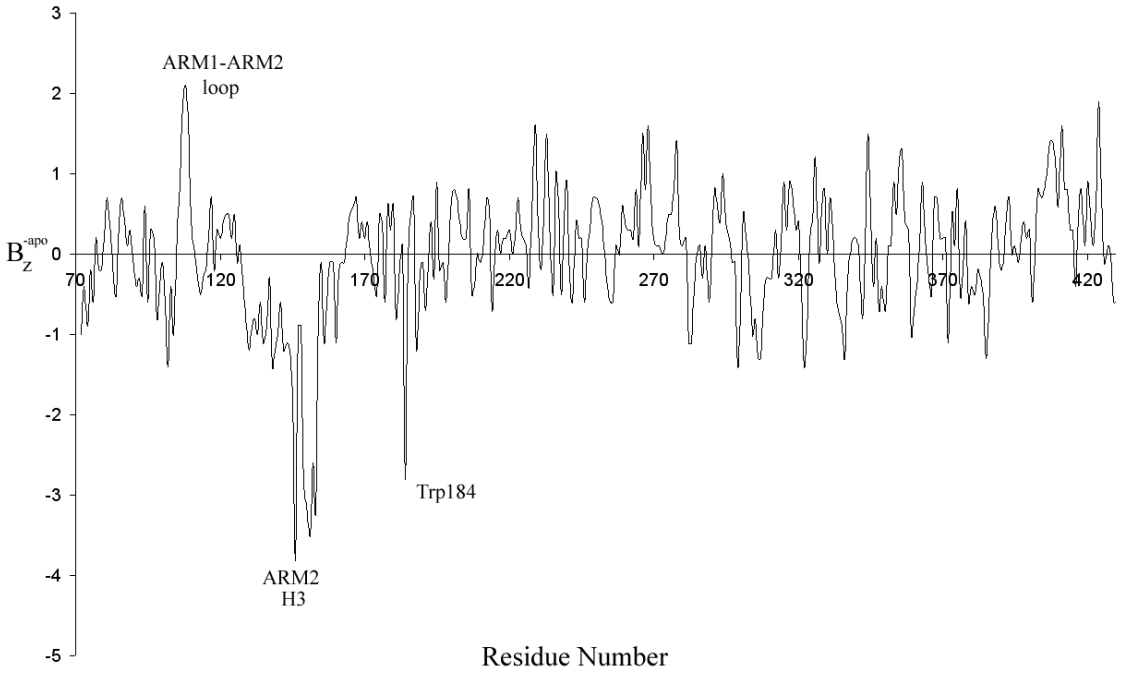

Figure 4.21: A plot of normalised B-factor scores for the importin-α:S21 complex.

The B_z^{-apo} score has been averaged over atoms in residues 72-430. Three peaks have been labelled: the ARM1-ARM2 connecting loop, the ARM2:H3 helix and Trp184 in ARM3.

The B_z^{-apo} scores for the importin-α:S21 complex are also shown in Figure 4.22, coloured on the molecular surface representation of importin-α from -3 to +3 normalisation units or standard deviation units. Residues 431-496 are considered outliers and are therefore coloured grey. The scores represent changes to residue flexibility that should only arise due to the binding of S21. There is a clustering of low B_z^{-apo} scores in the major binding site, but not the minor binding site.

In the major binding site, residues with side chains participating in peptide binding such as Asn146 (-5.6^{sc}), Asn188 (-2.4^{sc}), Asn235 (-1.5^{sc}), Trp142 (-1.2^{sc}), Trp184 (-3.4^{sc}) and Trp231 (-0.4^{sc}) all have low B_z^{-apo} scores reflecting their decreased flexibility. Arg238 (-1.1^{sc}) also has a low score. The key importin-α hydrogen bonding partners in the P2 binding slot display a similar B_z^{-apo} pattern as observed for the SV40 NLS complex, with Thr155 (-0.1^{res}) and Asp192 (-0.1^{res}) showing no significant changes, while Gly150 (-3.1^{res}) does have a significantly low B_z^{-apo} score. It seems quite likely a water molecule fills the P2 lysine N^{ζ} position in the *apo* structure, forming stabilising hydrogen bonds with Thr155 and Asp192 but not Gly150. Again

the reduced flexibility of Gly150 affects the adjacent residues from Ala148 to Ser152 in the ARM2-ARM3 connecting loop.

Differences in the way S21 and the SV40 NLS peptide bind importin-α (70-529) are apparent near the binding slot position P4. In importin-α:S21, an alanine residue occupies P4 making no interactions with the ARM1-ARM2 loop, while in the importin-α:SV40 NLS structure an arginine at P4 makes substantial interactions with the loop region including a network of four hydrogen bonds. In two of the residues that hydrogen bond ^{SV40}Arg130, Arg106 (+0.9^{res}) and Glu107 (+1.9^{res}), we actually see an increased B_z^{-apo} score in the importin-α:S21 structure. However a third hydrogen bonding partner, Leu104 (-1.0^{res}), does have decreased flexibility that is mainly evident on the side chain, not the main chain that hydrogen bonds ^{SV40}Arg130. This decrease in B_z^{-apo} is could be due to side chain hydrophobic contacts involving Leu104 and the significantly tightened residue, Asn146. Since there are no interactions at P4 with the ARM1-ARM2 connecting loop which includes residues Leu103 to Ile112, it is not unexpected that the chain has not decreased in flexibility. We observe an increase in B_z^{-apo} scores over this region and the reason for this is not immediately apparent. Electron density maps show that residue side chains of Arg106 and Glu107 have particularly weak corresponding density in importin-α:S21, which may facilitate the increase in B_z^{-apo} score.

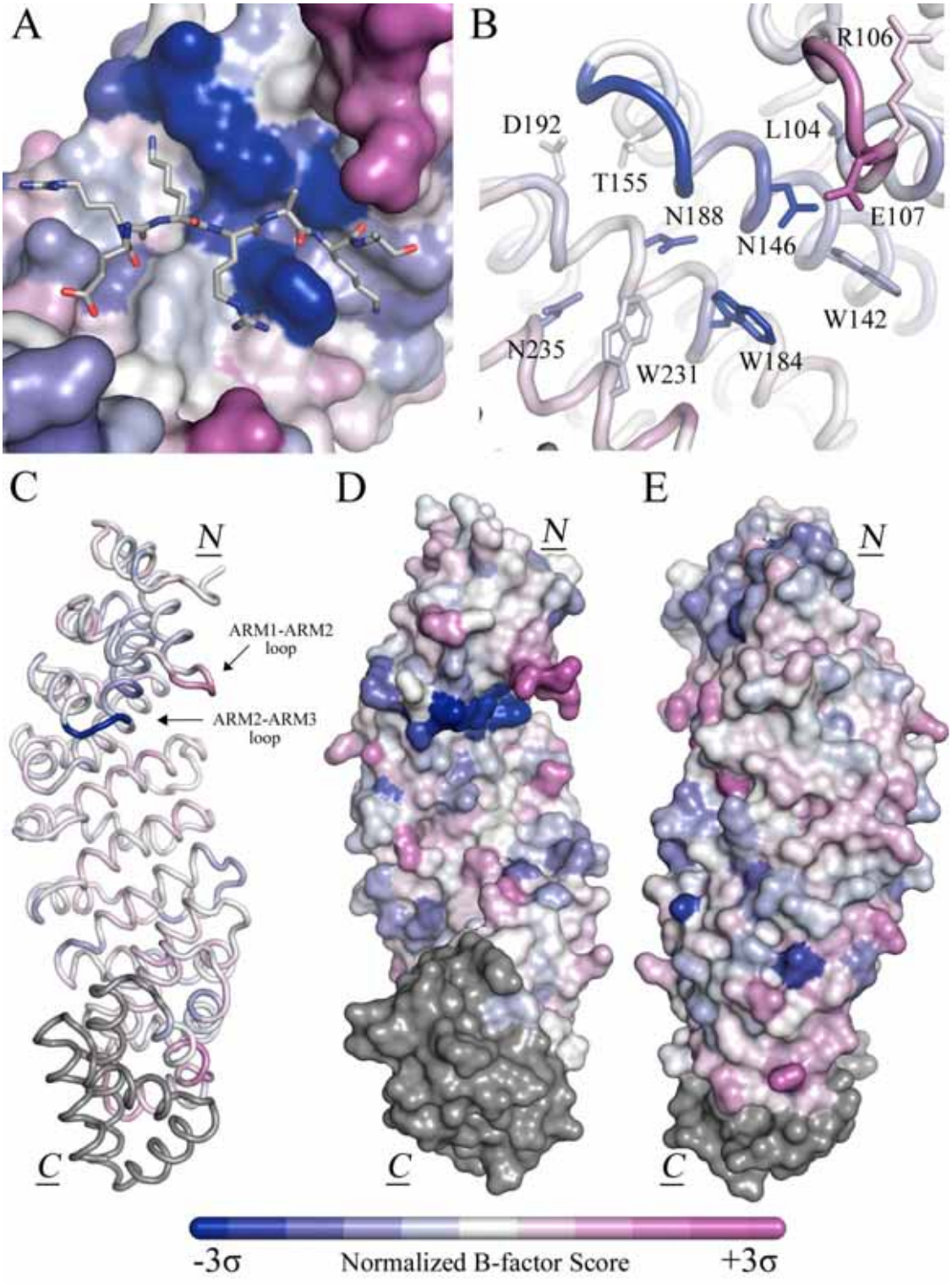

Figure 4.22: The importin-α:S21 complex coloured by a normalised B-factor score, B_z^{-apo}.

The B_z^{-apo} colour spectrum is blue-white-magenta from -3σ to +3σ as shown in the figure. Outliers near the C-terminus are shown in grey.

A) The importin-α major binding site showing the bound S21 NLS-like motif. **B)** Important side chains are shown in the major binding site. **C)** Cartoon tube representation of the importin-α backbone. **D)** The importin-α inner concave surface. **E)** The importin-α outer convex surface.

The normalised B-factor analysis provides additional insight into NLS binding by supporting the identification of specific S21 side chains as a result of the way they affect the flexibility of interacting importin-α (70-529) residues. Thus it is a useful method for supporting the S21 peptide model in conjunction with electron density analysis and data-data difference Fourier maps.

4.3.4 The real space correlation coefficient

From the (no-SEC) importin-α (70-529) structure, we suggested the rogue peptide density corresponds to a sequence motif, ^{P1}AKAA(K/R)A^{P6}, where A may correspond an alanine or longer side chain. Electron density in the importin-α:S21 major binding site corresponds to a sequence, ^{P1}AK(K/R)AKA^{P6}. Thus, there are two important revisions: at P3 side chain density corresponds to a lysine or arginine and at P5 density corresponds to a lysine. As described in Section 4.3.5, it is likely that the S11:S21 heterodimer is able to bind importin-α (70-529) through two sequence motifs that have NLS-like characteristics. To investigate which motif is the best fit to the importin-α major binding site electron density, we have used the Real Space Correlation Coefficient (RSCC) to provide an unbiased judgement of model quality. This has been applied to all residues that occupy binding slots P1 - P6, and the average over main chain atoms and side chain atoms for each residue is presented in Table 4.2 . We have also compared the results of other importin-α (70-529) structures presented in this thesis, including the importin-α:CLIC4 NLS complex presented Chapter 5. We also show RSCC values for the structure of a PLSCR1 NLS peptide bound to importin-α (70-529) [14], described in Section 4.4.5.

No distinguishable features are revealed by the peptide main chain fit to density, with an RSCC of 0.96 in all structures except the PLSCR1 peptide which has a significantly lower RSCC of 0.87 (Table 4.2). Side chain density fitting does vary however and we see in the SV40 NLS peptide good correlation to electron density with an RSCC of 0.92, the highest value obtained over the analysed structures. A slightly lower RSCC of 0.89 is calculated for CLIC4 NLS. For the S21 NLS-like peptide, we observed the model fit is also reasonable with an RSCC of 0.85, but it is equally

reasonable for the S11 NLS-like motif when the binding configuration places Lys125 at P2: ^{P1}PKKRRV^{P6}. The configuration where the S11 peptide N-terminus mimics the SV40 peptide by binding a lysine at P1 (^{P1}KKRRV^{P5}) leads to a lower RSCC of 0.77.

To distinguish between S11 and S21 peptide models we need to ignore the termini P1 and P6 binding slots where side chain density can be weak and ambiguous. In particular this applies to the P1 slot, where long side chains are commonly disordered in importin-α:NLS crystal structures. The RSCC value over binding slots P2 - P5 generally increases when compared to the value taken over P1 - P6 (Table 4.2). The values of the SV40 NLS and CLIC4 NLS increase to 0.94, comparable to their main chain RSCC values. The S21 model shows an increase in RSCC to 0.88, while the most accurate configuration of the S11 model decreases to 0.83. This difference arises due to the better correlation in density at P4 and P5 for the S21 motif.

Table 4.2: The real space correlation coefficient of NLS peptides.

NLS source	Major binding site	RSCC		
		Main chain	Side chains	
	1 2 3 4 5 6	P1 - P6	P1 - P6	P2 - P5
Model	A K K/R A K A	-	-	-
S21	E R K R A K A	0.96	0.85	0.88
S11	P P K K R R V –C	0.96	0.85	0.83
	P P K K R R V -C	0.96	0.77	0.80
SV40	P K K K R K V -C	0.96	0.92	0.94
CLIC4	V A K K Y R N -C	0.96	0.89	0.94
PLSCR1	G K I S K H	0.87	0.83	0.85

4.3.5 Identification of endogenous *E. coli* proteins

Prior to the identification of *E. coli* ribosomal proteins binding importin-α (70-529), we investigated the possibility that the rogue ligand observed in (no-SEC) importin-α (70-529) structures (Chapter 3) actually corresponds to a short peptide. A peptide appeared to be the logical candidate for the rogue ligand as we could only observe a short stretch of residues in the electron density difference maps. We performed analytical mass spectrometry methods to search for a low molecular weight

species present in the solution used to crystallize (no-SEC) importin-α (70-529). Scanning was limited to within a 10 kDa MW range that allowed for detection of longer peptide chains and small proteins.

The importin-α sample was found to contain an abundance of ions with a mass to charge ratio (m/z) of 576.78 and an experimentally determined atomic mass of 1152.56 amu. This peptide was identified as the 11 residue C-terminal tail of importin-α, ^{519}VQDGAPGTFNF^{529}, with a theoretical isotopically averaged atomic mass of 1152.23 amu. It is worth noting that the sequence occupies a region in importin-α that is disordered in crystal structures, presumably due to the flexibility of the C-terminus. We did not see evidence of proteolysis during protein preparation, as analysed with SDS-PAGE, which could explain the presence of the C-terminus peptide. As assessed by mass spectrometry, it is sufficiently abundant to account for the high occupancy of the rogue peptide observed in our crystal structures. Being 11 residues long it is also close to the expected length. However it does not elucidate the characteristic features of the electron density observed in the (no-SEC) importin-α major binding site. Most importantly there are no basic residues present in the sequence, particularly lysine, which are required to bind in the importin-α binding site. Since the density is a poor match to the sequence, we rule out the possibility that the C-terminal peptide is bound in the major binding site.

Further experiments were unable to find other small peptides present in the (no-SEC) importin-α (70-529) sample. However, an 8,368.5 Da protein was detected (see Figure 4.23). This was later identified as the 30 S ribosomal subunit S21, which has a theoretical MW of 8,368.7 Da calculated excluding the first methionine.

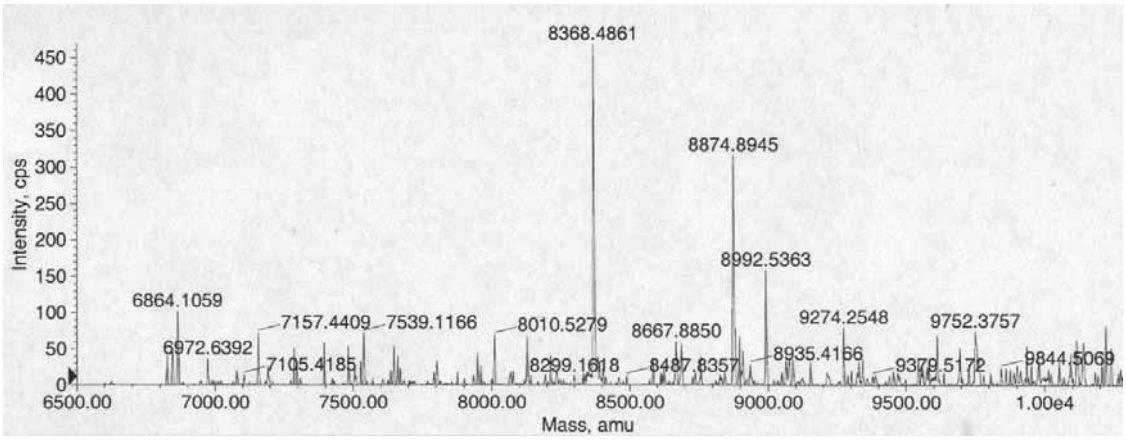

Figure 4.23: ESI-TOF-MS mass spectra of the 30S ribosomal subunit S21.

Shown is the ESI-TOF-MS mass reconstruction corresponding to the intact 30S ribosomal subunit S21 in the (no-SEC) importin-α (70-529) sample. The peak value of 8,368.5 amu corresponds closely to the S21 theoretical MW of 8,368.7 Da.

After developing an improved purification protocol that extended the standard protocol outlined in the literature, we were able to isolate a sample of importin-α (70-529) that contained contaminant *E. coli* proteins (refer to Figure 4.7). It should be noted that this sample was separated from *apo* importin-α (70-529) in a higher MW peak, and thus low MW protein species are likely to constitute a complex in solution. The sample was loaded on SDS-PAGE (Figure 4.24) and mass spectrometry was used to identify the proteins in each band. The identified proteins and relevant properties are listed in Table 4.3.

Other than importin-α (70-529), the proteins identified by mass spectrometry are endogenous to *E. coli* and therefore have no functional role in nuclear import and hence no established NLS. Manual sequence analysis was performed to identify NLS-like sequence motifs that contain characteristic stretches of basic residues, in particular looking for sequences that match the rogue peptide electron density. Most of the identified proteins contain clusters of basic residues that form a reasonable NLS-like sequence, so we have investigated these motifs in the corresponding published structure.

Table 4.3: Mass spectrometry identification of endogenous *E. coli* proteins.

Band	Identified protein	Mascot Score[a]	MW on gel (kDa)	Actual MW (kDa)	Dimensions[b] (Å)	NLS-like sequence[c]
1	Importin-α	338	110	110.6[d]	-	
	IF2	233		97.3	120 x 60 x 30	^{647}E**K**A**R**E^{652}
2	Importin-α	568	90	55.3	90 x 30 x 30	-
	ClpA	157		84.1	130 x 55 x 40	^{415}S**KR**K**KT**^{420}
3	Importin-α	431	64	69.1[d]	-	-
	Ribosomal S11	339				
4	Importin-α	1878	54	55.3	90 x 30 x 30	-
5	Ribosomal S11	186	15	13.8	35 x 30 x 25	^{123}PP**KKRRV**^{129} - COOH
6	Ribosomal S21	165	10	8.5	-	^{45}**RKR**A**K**A^{50}

[a] The total Mascot score used to rank protein hits.
[b] Dimensions of the folded protein module, excluding flexible regions. IF2 (PDB: 1ZO1), ClpA (PDB: 1KSF), S11:S21 (PDB: 2AVY).
[c] Bold residues are basic, underlined residues are solvent exposed.
[d] Calculated MW corresponds to the importin-α (70-529) dimer or importin-α:S11 complex.

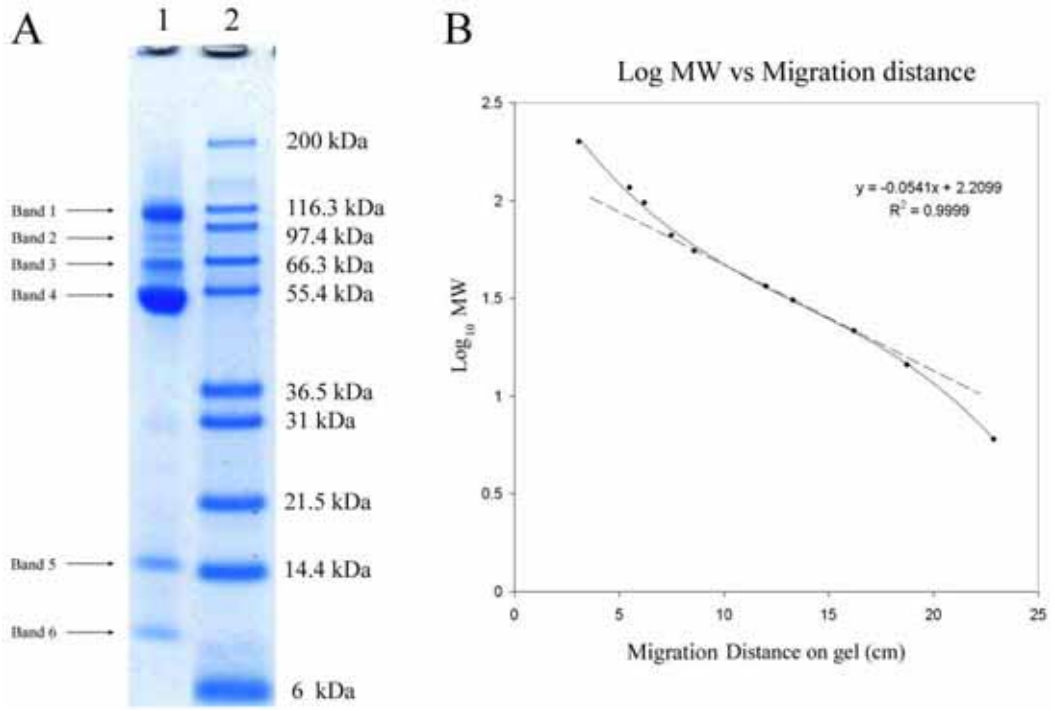

Figure 4.24: SDS-PAGE analysis of the importin-α (70-529) and *E. coli* protein sample.

A) SDS-PAGE of importin-α (70-529) and *E. coli* contaminants. Lane 1 contains the importin-α:*E. coli* solution and lane 2 contains the Mark12^{TM} Unstained protein Standard (Invitrogen). Standard molecular weights are shown on the right. Band numbering in lane 1 correspond to the first column in Table 4.3.

B) A plot of the logarithm of molecular weight (base 10) against protein migration distance (cm) for protein standards in panel A. The linear segment of the curve is designated by a dashed line and the corresponding straight line equation is shown with an R^2 correlation coefficient.

In the highest molecular weight band at 110 kDa, importin-α (70-529) and the translation initiation factor 2 (IF2), were identified. IF2 is involved in ribosomal initiation where it appears to associate with the 30S and 50S ribosomal subunits and also 30S subunits including S11 [191-193], which was identified in band 5. There is an NLS-like motif in IF2, ^{647}EKKARE^{652}, which is located near the protein surface within a loop region (PDB: 1ZO1). Despite this the side chains of Lys648 and Arg651 are partially buried suggesting the region would need to undergo a structural transformation to allow the motif to bind.

It is likely that importin-α (70-529) is dimeric in band 1 as the theoretical molecular weight of the dimer (110.6 kDa) matches closely to the molecular weight of band 1 on the gel (110 kDa). The dimensions of IF2 (120 Å x 60 Å x 30 Å) estimated from the published structure appear too large to fit within the solvent space of the importin-α (70-529) crystal and so it is unlikely the IF2 NLS-like motif corresponds to the rogue ligand. However it cannot be ruled out that IF2 and importin-α interact in solution.

Band 2 contains *E. coli* ClpA, an 84.1 kDa member of the Hsp100/Clp family of ATPases [194]. ClpA forms the substrate binding unit of the ClpAP degradation pathway [195]. An NLS-like sequence in ClpA includes residues ^{415}SKRKKT^{420}, which are within a loop region of the protein (PDB:1KSF). Basic residues in the motif are solvent exposed except for Lys418 which is buried. The low concentration of the protein, as judged from SDS-PAGE, suggests ClpA does not correspond to the rogue ligand.

Band 3 contains importin-α (70-529) and the 129 residue long 30S ribosomal subunit S11. This protein is also identified in band 5 at an experimentally determined molecular weight of 15 kDa, corresponding closely to its monomeric theoretical molecular weight of 13.8 kDa. The X-ray crystal structure of the 30S ribosome has been solved at 3.5 Å (PDB:2AVY), showing subunits S1-S21 bound to 16S ribosomal RNA [196]. In the structure, S11 exists as a globular protein with flexible N-terminal and C-terminal tail regions.

Within the S11 sequence there are two major clusters of negatively charged residue: the N-terminal sequence ^{9}RKRVRK^{14} and C-terminal sequence ^{123}PPKKRRV^{129}-COOH. Both of these motifs appear capable of binding importin-α in an NLS-like mode. The N-terminal sequence is disordered in the crystal structure suggesting a degree of flexibility, however reconciling the motif with electron density in the importin-α:S21 'omit' difference map is difficult since core residues Val12 and Arg13 do not fit unambiguously. The C-terminal sequence, which represents the last 6 residues of S11, bears notable similarities to the SV40 NLS sequence, PPKKKRKV, but does not match rogue peptide electron density in all key binding slots regardless of whether Lys125 or Lys126 occupies P2. Binding with a proline prior to P1, like the configuration of the SV40 NLS, leaves P5 unfavourably occupied by a valine. Binding with a proline in P1 is different to the configuration of the SV40 NLS, but it allows the core binding region (P2 - P5) to be occupied by four basic residues. In this case electron density for Pro124 at P1 and Arg127 at P4 is notably absent. Considering the quality of electron density for an arginine at P5 in our own importin-α:SV40 NLS structure, this is a conspicuous absence.

The presence of S11 in band 3 (64 kDa), in company with importin-α (70-529), suggests the two may form a complex which would have a theoretical molecular weight of 69.1 kDa. Since the complex is not separated under denaturing conditions, it may be that S11 and importin-α (70-529) have formed a disulfide bond that is strong enough to withstand the reduced environment during SDS-PAGE (2 mM DTT). S11 contains two cysteine residues that in the folded structure correspond to a buried Cys70 and a solvent exposed Cys121. The latter is close to the C-terminal NLS-like sequence (^{121}CRPPKKRRV^{129}). While it cannot be ruled out that the disulfide bond formed with importin-α occurs after SDS-PAGE denaturation, we do note the presence of a disulfide bond at residue $^{imp\text{-}\alpha}$Cys133, observed only in the importin-α:S21 structure. The sulfur atom covalently bound to importin-α (70-529) appears to be present with an occupancy ~80%. If the solvent exposed ^{S11}Cys121 is forming the disulfide bond to $^{imp\text{-}\alpha}$Cys133 in the crystal structure, then the constraints placed on S11 due to this bond mean the C-terminal NLS-like motif cannot bind.

Band 6 contains another 30S ribosomal subunit, S21. This is a 71 residue long protein that runs at 10 kDa on SDS-PAGE, close to its theoretical MW of 8.5 kDa. A search through the S21 sequence reveals a single NLS-like motif, ^{45}RKRAKA^{50}, which matches the rogue peptide electron density better than any other sequence that has been considered. If this motif was to bind in the importin-α (70-529) major binding site, it would require Lys46 to occupy the P2 site which in turn aligns Ala48 in the P4 site and Lys49 at P5. This is the only motif that places an alanine at P4: the required residue to correspond to electron density in this binding slot. The lysine at P5 also matches the rogue peptide electron density well, providing a more favourable fit than an arginine.

In the 30S ribosome, S21 exists in an extended conformation with 2 regions of helical secondary structure where it forms an interface with S11 or 16S rRNA (Figure 4.25). The last 17 residues, just after the NLS-like motif, are not modelled in the structure and are likely to be quite flexible. Looking at the structure of the 30S ribosomal subunit we note that S21 is positioned in contact with S11 and the two make no significant interactions with the surrounding S7 and S18 subunits. The contact made between S11 and S21 is extensive, burying a solvent accessible surface area of 944.3 Å^2 on S21 (17.1% of total area) and 933.7 Å^2 on S11 (12.1% of total area). There are 5 hydrogen bonds between S11 and S21, mostly clustered towards the S11 C-terminus. These presumably help to stabilise the S11:S21 interaction in the ribosome. It is surprising that the only two ribosomal subunits identified are closely interacting on the ribosome, perhaps suggesting that an S11:S21 heterodimer is stable independent of 16S rRNA.

The NLS-like motif near the S21 C-terminus fits into a 16S rRNA surface cavity and possesses helical structure. It is directed away from the core module of S11:S21 leaving it highly exposed and readily available to interact with importin-α (70-529) if it were free from the ribosome. There is a single cysteine in S21 at Cys23 in the centre of the protein, which forms a hydrogen bond with the carbonyl oxygen of Lys19 on the opposite side of the interface between S11 and S21. Although Cys23 is likely to be surface exposed when free of the ribosome, S21 was not identified in other gel bands like S11. Because an NLS binds anti-parallel to importin-α and Cys23 is towards the

N-terminus, it is difficult to reconcile the formation of a disulfide bond between ^{S21}Cys23 and $^{imp\text{-}\alpha}$Cys133, with a major binding site interaction involving the NLS-like motif.

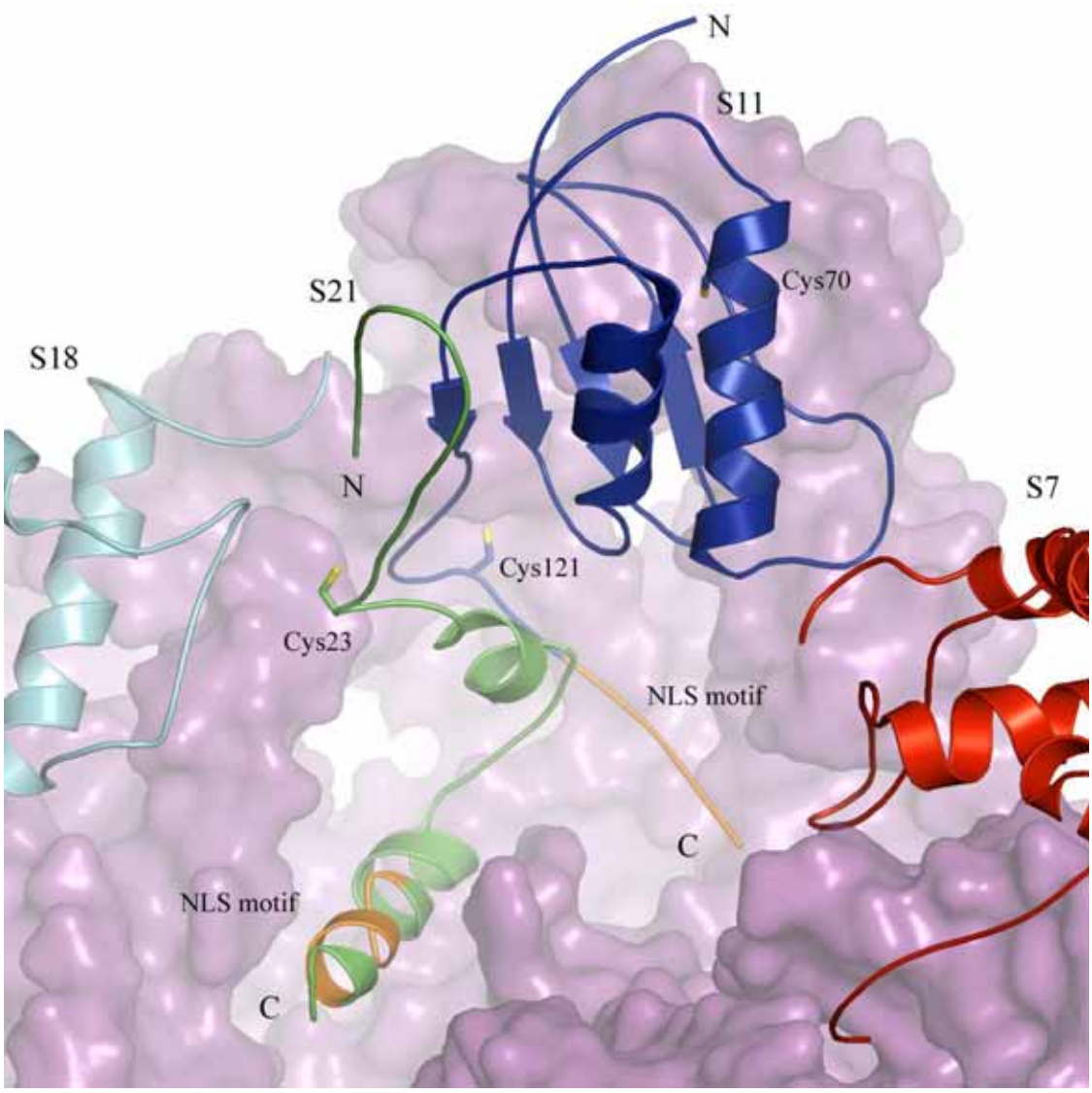

Figure 4.25: Crystal structure of *E. coli* 30S ribosomal subunits.

The 30S ribosomal subunits S7 (red), S11 (blue), S18 (cyan) and S21 (green) are shown in cartoon representation. The 16S rRNA molecular surface is coloured magenta. Cysteine residues are shown as sticks with sulfur atoms coloured yellow. NLS-like regions in S11 and S21 are coloured orange and the amino (N) and carboxyl (C) termini are marked. The structure is reproduced from PDB: 2AVY [196]. Figure created using PyMOL [165].

Estimating the abundance of *E. coli* proteins from SDS-PAGE, S11 and S21 are likely to be the only proteins with a molar concentration high enough to account for the approximately 1:1 molar ratio observed between importin-α (70-529) and the rogue peptide. The two 30S ribosomal subunits S11 and S21 are also the only proteins

identified that can be easily accommodated in the solvent space of the importin-α (70-529) crystal (Figure 4.26). The solvent space above the minor binding site forms an alcove with dimensions of approximately 100 Å x 45 Å x 30 Å, which is just sufficient for an S11:S21 heterodimer to fit under the shortest axis. The S11:S21 NLS-like motif capable of binding importin-α (70-529) in the major binding site is likely to be at the flexible C-terminus of either S11 (^{123}PPKKRRV^{129}-COOH) or S21 (^{45}RKRAKA^{50}). Therefore the S11:S21 folded module should have a wide range of movement allowing it to adopt a favourable orientation that fits into the crystal alcove.

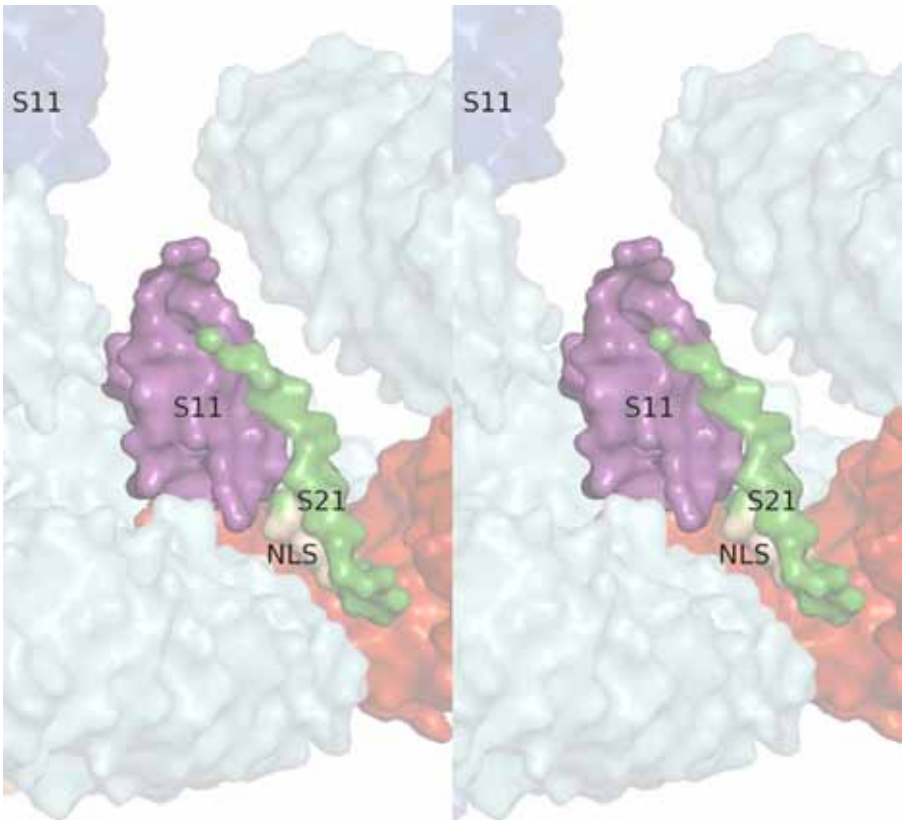

Figure 4.26: Stereo image of a hypothetical complex formed between importin-α (70-529) and the S11:S21 heterodimer.

The figure shows the superposition of importin-α:S21 (Section 4.3.3) and S11:S21 from PDB: 2AVY, with the S21 NLS-like motif (light brown) binding to the importin-α (70-529) major binding site in the correct orientation. The molecular surface is shown for S11 (magenta), S21 (green), importin-α (red), importin-α symmetry molecules (cyan) and S11 symmetry molecules (blue). S21 symmetry molecules do not appear in the image due to the viewing angle only.

4.4 Discussion

4.4.1 Separation of *apo* importin-α (70-529) from endogenous *E. coli* proteins

A significant advancement in our understanding of the results from Chapter 3 was achieved when an improved purification protocol that included ion exchange chromatography (IEC) and size exclusion chromatography (SEC), in addition to the standard protocol, was able to separate importin-α (70-529) from contaminant molecular species. This yielded two important samples: *apo* importin-α (70-529) and importin-α (70-529) bound to *E. coli* proteins. The structure of the former was solved at 1.77 Å, the highest resolution importin-α structure currently available and the first *Mus musculus* importin-α solved with an empty binding site.

The only structure of importin-α with an empty binding site currently reported in the literature is that of *Saccharomyces cerevisiae* importin-α (88-530) (PDB:1BK6), solved at 2.2 Å. We note that preparation of this construct also involved overexpression in BL21(DE3) cells and purification using IEC without SEC. Although it is in the same space group as our structure ($P2_12_12_1$) it contains two importin-α molecules in the asymmetric unit forming a dimer with an extensive interface and it was noted that yeast importin-α is homodimeric in solution [30]. Despite this, it seems probable the dimer observed in the structure is a result of a different crystal packing arrangement. Although the subunits do not directly overlap the major binding site, with the ARM10 helices of one subunit interacting with the central H3 helices of the other subunit, they will still sterically interfere with the N-terminal flanking region of an NLS. This will likely lower the NLS affinity for the binding site and make the yeast *apo* structure easier to achieve even in the presence of *E. coli* contaminants. We were curious to analyse the veracity of the stated unliganded importin-α structure, but there are no deposited structure factors and little discussion in the paper to elucidate queries about residual electron density [30].

In our *apo* structure we note the presence of weak peptidic side chain density with what could be solvent and other ions filling the site. In the mF_o-DF_c difference map at a 2.8σ contour level there are clear breaks in the main chain density accompanied by

empty side chains in binding site positions P1, P3, P4 and P6. Despite the weak density towards the centre of the binding site, P5 contains relatively strong density for a lysine or arginine. It may be that with the removal of endogenous *E. coli* proteins, the N-terminal *linker* domain is able to occupy P5 leaving other binding slots mostly empty. Since the *linker* domain is disordered in the crystal structure we are unable to confirm such an interaction.

Binding slot P2 has an electron density stub with a well defined water molecule that fits in the critical lysine N^{ζ} atom location, maintaining two of the three hydrogen bonds left vacant by the absence of a lysine. In the *apo* yeast importin-α structure a cobalt ion present in the crystallization buffer is modelled at P2 near the site of the lysine N^{ζ} atom [30]. It is clear the P2 site favours the establishment of hydrogen bonds and is willing to accept available buffer and solvent molecules in the absence of an NLS lysine. This is consistent with the known binding strength of the site and its importance in nuclear import.

We also note an increase in the flexibility of binding site residues such as Ser149, which can only be clearly modelled with an alternative rotamer in the *apo* structure. We expect this is the case since the alternate Ser149 conformation creates steric overlaps with the main chain of an NLS peptide at P2 and P3. There is also evidence that Trp184 has extra flexibility in the *apo* form, with strong negative density encompassing the indole side chain which may have dual rotamer conformations: one is a rare tryptophan conformation that facilitates the formation of the P5 binding pocket, and the other is more common conformation that eliminates the well formed P5 pocket. Interestingly, the equivalent tryptophan in yeast *apo* importin-α (PDB:1BK6), Trp195, is seen to adopt this more common rotamer where it has been modelled with a χ_2 rotation of 90º compared to the conformation it adopts in the accompanying yeast importin-α:SV40 NLS structure (PDB:1BK5) [30]. In our structure Trp184 shows evidence that it adopts the same common rotamer seen in yeast *apo* importin-α, however the occupancy of the rotamer is low perhaps suggesting the side chain switches between the two conformational states.

4.4.2 The importin-α:SV40 NLS complex

The *apo* importin-α (70-529) sample assisted with the successful determination of the importin-α:SV40 NLS complex at 1.85 Å resolution, in which the SV40 NLS peptide, ^{126}PKKKRKV^{132}, is unambiguously modelled in the major and minor binding sites. In comparison, the (no-SEC) importin-α:SV40 NLS structure contains only weak electron density that corresponds to a bound peptide, thus it is clear the removal of competing *E. coli* ligands has made a significant difference.

We have analysed the $F_o^{sv40} - F_o^{apo}$ data-data difference Fourier which represents a comparison of scaled structure factors in the importin-α:SV40 NLS complex and *apo* importin-α (70-529). This approach has been used in all importin-α:NLS complexes presented in this thesis where it provides a rigorous and unbiased method to test the validity of peptide assignment in the importin-α binding site. Strong positive electron density remains in the $F_o^{sv40} - F_o^{apo}$ map near the major binding site, clearly indicating the presence of the SV40 NLS peptide. In contrast, the minor binding site contains only weak density that does not clearly identify the SV40 NLS peptide. Despite the fact our *apo* structure contains bound molecules in the major binding site, the difference Fourier map at 2.8σ defines unbroken main chain and side chain density for residues in binding slots P1 to P6. At P1 there is strong main chain and partial side chain density for a lysine. The lysine residue at P2 contains corresponding density but since the site is occupied by water in the *apo* structure, density is weak near the lysine N^{ζ} atom. P3 contains main chain density as well as unambiguous density for a lysine side chain. P5 also contains density for a lysine residue in the $F_o^{sv40} - F_o^{apo}$ map, despite density for what may correspond to a lysine in *apo* importin-α (70-529). This probably suggests the P5 lysine has a higher occupancy than the P5 residue in *apo*. At P6 there is density for a valine and the carboxyl terminator group is clearly observed at the point where density ceases.

The P4 binding slot has been identified as the least well defined of the core binding slots that include P2 - P5 [62], however we demonstrate that it does provide one of the key indications that a particular peptide sequence has bound to importin-α.

Unlike the other core binding slots, a degree of residue diversity is accepted at P4. Therefore it is more common for non-basic residues to occupy the site and present a defining characteristic that can be assessed. In our work, this is emphasised by the weak main chain and side chain density in *apo* at this location, which means the data-data difference Fourier will reveal clear density features that only correspond to the importin-α:NLS complex. Also, the endogenous *E. coli* 30S ribosomal subunit S21, has an NLS-like sequence that fits the major binding site with an alanine residue at P4, which is not easily confused with a longer side chain. The SV40 NLS peptide in our importin-α:SV40 NLS complex contains an arginine at P4 (Arg130) that is clearly modelled in the $F_o^{sv40} - F_o^{apo}$ difference Fourier map. We have shown that ^{SV40}Arg130 adopts a conformation that forms an extensive network of hydrogen bonds between its N^{H1} and N^{H2} atoms, to importin-α main chain carbonyl oxygen atoms in the loop linking ARM1 and ARM2. The modelled conformation of ^{SV40}Arg130 is more favourable than what is likely to be a misinterpreted conformation in previously solved importin-α:SV40 NLS structures [6, 8, 12].

A case has been made that NLS peptide binding to importin-α (70-529) decreases the flexibility of residues in and around the importin-α major binding site by either direct or indirect interaction. The normalised atomic B-factor score, B_z^{-apo}, was used to show differences in residue mobility by comparing importin-α:SV40 NLS and importin-α:S21 structures with the *apo* structure. This type of analysis has previously been used to estimate regions of flexibility in proteins [185], and to investigate active sites [197], but has apparently not been used to analyse binding interactions by comparing proteins in the *apo* and complexed forms. The normalised B-factor analysis we perform therefore represents a novel investigative tool which we apply to the study of importin-α flexibility. We have shown that normalised B-factors accurately describe changes to importin-α (70-529) after binding substrates.

In the case of the importin-α:SV40 complex, the results show that importin-α (70-529) residues participating in hydrogen bonding or van der Waals contacts with the SV40 NLS peptide generally have negative B_z^{-apo} scores. A negative score corresponds to a decrease in normalised B-factors and suggests that residue flexibility has also decreased. Some of the most significant B_z^{-apo} scores correspond to residues in

the tryptophan and asparagine arrays of the major binding site, particularly Trp184 and Asn146, which interact with NLS peptide residues in the core P3-P5 binding slots. There is also reduced flexibility near the importin-α loop connecting ARM1 and ARM2. Extra constraints on the loop will mainly arise due to the interaction introduced with ^{SV40}Arg130 at P4. A similar effect is seen near P2 where the loop connecting ARM2 and ARM3 is also tightened due to the hydrogen bond involving $^{imp\text{-}\alpha}$Gly150, introduced by the presence of ^{SV40}Lys128.

The normalised B-factor analysis with the data-data difference Fourier make for a convincing argument that the SV40 NLS peptide has unambiguously bound importin-α (70-529) in the major binding site. Binding of the SV40 NLS in the minor binding site cannot be confirmed using these methods, something that might be expected due to the weaker interaction. The findings are supportive evidence that the *apo* structure will not interfere with NLS peptide binding in importin-α:NLS co-crystal structures. It also provides additional insight into the mechanism of NLS binding and serves as a useful experimental control in readiness for an importin-α:CLIC4 NLS structure.

4.4.3 The importin-α:S21 complex

After separating *apo* importin-α (70-529) using an improved purification protocol, we obtained a sample of importin-α (70-529) including *E. coli* contaminants some of which may correspond to the rogue ligand observed in (no-SEC) structures. Since the protein contaminants are endogenous to *E. coli*, a prokaryotic cell lacking a nucleus, it may simply be a coincidence that they can bind the nuclear import receptor, importin-α, and thus survive through purification. By crystallizing the importin-α:*E. coli* sample, we managed to obtain a 1.95 Å structure that includes strong main chain and side chain density in the major binding site corresponding to a bound NLS-like motif in the 30S ribosomal subunit S21, ^{45}RKRAKA^{50}. The sequence also corresponds closely to the rogue peptide electron density observed in the (no-SEC) importin-α (70-529) structures presented in Chapter 3.

The importin-α:S21 structure shows that Lys46 fits into the critical P2 slot, aligning Ala48 at P4 commensurate with a density stub observed here. A lysine at P5

(Lys49) is also an unambiguous fit to electron density. It is still difficult to assign a residue unambiguously in the P3 binding slot, where density corresponds to Arg47 but could equally correspond to a lysine. Similar to other importin-α:NLS structures, the side chain at P1 is mostly disordered and the P6 carbonyl group marks the end of the ordered peptide chain. The $F_o^{S21} - F_o^{apo}$ data-data difference Fourier confirms the presence of the S21 NLS-like motif, with weak density for main chain atoms and the side chain of Arg47 at P3. Other side chains are not unambiguously defined, which may reflect the low occupancy of S21 in the major binding site.

Analysis of a normalised B-factor score, B_z^{-apo}, was performed on the importin-α:S21 structure similar to that performed on the importin-α:SV40 NLS complex. A pattern of low B_z^{-apo} scores in the major binding site is consistent with the reduced residue flexibility expected when a peptide binds. This is particularly notable over the conserved tryptophan and asparagine arrays, as well as the ARM2-ARM3 connecting loop. A notable difference in the importin-α:S21 complex is the increased flexibility of the ARM1-ARM2 connecting loop, which normally interacts with long NLS side chains in the P4 binding slot. In the importin-α:SV40 NLS or importin-α:CLIC4 NLS, the P4 residue corresponds to an arginine and tyrosine respectively. Normalised B-factor analysis of these structures shows the ARM1-ARM2 connecting loop experiences a significant decrease in B_z^{-apo} scores. However in the S21 NLS-like motif, the P4 residue corresponds to an alanine that does not interact with the loop region. Therefore the mobility of the ARM1-ARM2 loop is likely governed by the NLS P4 side chain interaction with importin-α.

4.4.4 Interaction between importin-α and 30S ribosomal subunits

Mass spectrometry was used to identify protein species present in the importin-α:*E. coli* sample via peptide mass fingerprinting. The largest protein identified is the 97.3 kDa *E. coli* translation initiation factor, IF2 (isoform 1), which is one of three proteins known to initiate mRNA translation in prokaryotes, the other 2 being IF1 and IF3 (see review paper [198]). The basic role of IF2 is in ribosomal initiation where it appears to associate with the 30S ribosomal subunit and then

facilitate formylmethionine-tRNA (fMet-tRNA) binding to the ribosomal P (peptidyl) site [199]. The N-terminal domain of IF2 has been shown to interact with both the 30S and 50S ribosomal subunits [191, 192]. Furthermore, IF2 has been crosslinked to a number of 30S ribosomal proteins including S11 [193]. Other roles for IF2 could indicate it acts as a chaperone, forming stable complexes with unfolded proteins [200]. IF2 contains an NLS-like motif, ^{648}KKAR^{651}, which is partially solvent exposed near a core structured domain. While the motif matches the rogue peptide electron density, it will be difficult for this NLS-like motif to bind importin-α without an unfolding event to expose the local region including flanking residues. Furthermore, the crystal packing of importin-α (70-529) leaves a pocket of space approximately 100 Å x 45 Å x 30 Å above the minor binding site, not enough for the large folded core domain of the IF2 molecule to be present in the crystal. There may still be an interaction with importin-α (70-529) that is stable in solution, possibly facilitated through 30S ribosomal subunits S11 and S21. This is a possible explanation for the presence of IF2 in the importin-α (70-529) sample after purification.

The only identified protein not associated with translation is ClpA, an 84.1 kDa component of the ATP-dependent ClpAP protease and a member of the AAA$^+$ superfamily (ATPases associated with various cellular activities) [201]. ClpA is further classified as a member of the Hsp100/Clp family of ATPases with ATP-dependent chaperone activity [194]. It has been found that ClpA forms the substrate binding unit of the ClpAP degradation pathway [195]. The X-ray crystal structure of ClpA has been solved, revealing a compact structure consisting of two AAA$^+$ modules with very few loop regions [202]. The most likely NLS sequence from ClpA includes residues ^{416}KRKK^{419}, which are in a partially solvent exposed loop region linking two core domain helices. Again, this protein would need to undergo a degree of local unfolding to expose the NLS-like motif and flanking regions. Like IF2, the ClpA protein is also too large to fit in the solvent space of the importin-α (70-529) crystal and therefore it seems unlikely that we are observing an interaction between ClpA and importin-α in our crystal structure.

The identified proteins that are capable of fitting in the solvent space of the importin-α (70-529) crystal are the S11 (13.8 kDa) and S21 (8.5 kDa) ribosomal

subunits. Their separation on SEC at a MW higher than monomeric importin-α (70-529) suggests that these proteins are part of a larger complex in solution. The subunits are seen to be interacting in the X-ray crystal structure of the *E. coli* ribosome [196]. With an extensive heterodimeric interface in the 30S ribosome structure, it is possible S11:S21 form a stable complex outside of the ribosome. Furthermore, mass spectrometry identified importin-α (70-529) and S11 together in a gel band at 65 kDa, close to the 69.1 kDa theoretical MW of an importin-α:S11 complex. Since SDS-PAGE would be expected to unfold complexes, the most likely scenario is that the two have formed a disulfide bond that is stable in DTT and SDS. It may be no coincidence that a feature of the importin-α:S21 complex is evidence for a disulfide bond at Cys133, a residue which is relatively close to the major binding site.

The S11 and S21 protein sequences have numerous clusters of basic residues, however only two sequences from each protein are particularly noteworthy because they are solvent exposed and extended at the flexible C-terminus tail. In the *E. coli* ribosome crystal structure these sequences are nearby one another in a similar orientation. In S11, the sequence ^{123}PPKKRRV^{129}-COOH, is reminiscent of the classical NLS sequence from SV40 (^{125}PPKKKRKV^{132}) with one less basic residue. The SV40 NLS is not located at the C-terminus of SV40, but Val129 is the last residue in S11. If the S11 NLS-like peptide were to bind to importin-α (70-529) in an SV40 NLS-like manner, this would place Lys126 in the key binding slot P2 with arginine residues in P3 and P4, and Val129 in P5. Such a configuration would not be favourable for peptide binding as seen by a low real space correlation coefficient (RSCC) matching electron density in the importin-α:S21 binding site. A shift of one residue towards the C-terminus puts the core KKRR motif of S11 in a more favourable position, (P2 - P5), and increases the RSCC score. However, this places Arg127 at P4, a binding slot at which we would expect to see side chain density similar to that seen for the arginine at P4 in the importin-α:SV40 NLS structure. The importin-α:S21 structure has clear and strong density that corresponds to an alanine side chain at P4.

In S21 the sequence ^{45}RKRAKA^{50} corresponds to an NLS-like motif located near the C-terminus. In the 30S ribosome this NLS-like motif has helical structure stabilised by interactions with 16S rRNA. The sequence corresponds closely to the rogue peptide

electron density seen in the (no-SEC) importin-α (K25E, R28G) mutant structure and also the importin-α:S21 structure. This sequence is proposed to be the rogue peptide seen throughout our importin-α (70-529) structural studies, fitting the electron density more appropriately than any other NLS, as confirmed by a higher RSCC. With the S21 C-terminal sequence binding importin-α (70-529) it is possible that the entire S21 protein is intact when bound, making it a rare importin-α:cargo protein structure solved where the cargo is full length, albeit artificial.

In the assembly pathway of the 30S ribosomal subunits, it is seen that the S11 and S21 subunits are assembled along the S15 assembly branch, in which S15 first interacts with 16S rRNA allowing further interactions with S18 and S6, and finally S11 and S21 [203]. While the S18 and S6 subunits are known to form a functional heterodimer that interacts with S15, they are not interacting in the crystal structure. There is currently no information about the possibility of an S11 and S21 heterodimer despite the close interaction observed in the ribosomal crystal structure and their assembly in the ribosome along the same pathway. Since IF2 associates with S11, a possible explanation for its presence in the sample may be that it indirectly associates with importin-α through S11 and S21 (IF2:S11:S21:importin-α). The breakdown of this complex may have been observed on SDS-PAGE in which only an S11:importin-α covalent interaction remains. The molar ratio of the complex constituents, as estimated from SDS-PAGE, would suggest such a quaternary complex has a lower concentration in solution compared to the trimeric complex of S11:S21:importin-α.

In a complexed state, it may be that S11 is also making interactions with importin-α after forming an S11:S21 heterodimer that binds to importin-α through the S21 NLS-like interaction. A covalent disulfide bond between S11 and importin-α could explain how only these proteins associate on SDS-PAGE and furthermore increase the affinity of S21 for importin-α. The disulfide bond at $^{imp\text{-}\alpha}$Cys133 in the importin-α:S21 structure has been maintained despite the presence of reducing agent in the crystallization condition. Since the crystals of importin-α:S21 took over one month to develop, it is possible the crystal was no longer in a strong reducing environment when the disulfide formed. An S11:S21 complex should be small enough to diffuse through the importin-α crystal, although it seems more plausible the interaction first

takes place before crystallization. There are two cysteine residues in S11 that could be forming the disulfide bond with $^{imp\text{-}\alpha}$Cys133: Cys70 and Cys121. Since Cys70 is buried in the folded structure, Cys121 in a flexible region of S11 is the more likely candidate. Although it is possible that the disulfide bond is formed with a neighbouring symmetry S11:S21 dimer, it is difficult to reconcile this with an NLS-like interaction involving S21 due to spatial considerations. An alternative scenario may be the presence of both S11 and S21 interacting with importin-α independently.

The S21 subunit in the ribosome structure is fixed into position via its extended C-terminal tail filling an RNA surface cavity and it is noted that the 3′ end of the 16S rRNA is positioned closely to the S11:S21 sub-units. While we have modelled the S21 RKRAKA peptide fragment in the importin-α:S21 structure, we hypothesise that we are observing an interaction between importin-α (70-529) and an intact S11:S21 heterodimer binding to the major binding site via the C-terminal tail of S21, or a combination of S11 and S21. There is enough solvent space in the crystal to allow this S11:S21 dimer to be present and disordered in the structure when anchored via an interaction with the importin-α major binding site.

We see no clear evidence that the binding site is occupied homogeneously in electron density maps and therefore if it is a combination of S11 and S21 interacting with importin-α, the occupancy of S11 at the binding site must be very low. Hence it is unlikely S11 binds importin-α by itself, but possibile that S21 does. In order to clarify this point, future experiments using small angle X-ray scattering (SAXS) could be performed to identify the presence of S11:S21 bound to importin-α (70-529) in an analogous mode to importin-α:cargo interactions. SAXS experiments have been performed previously on karyopherins to show intrinsic flexibility and structural changes that arise due to cargo binding [19, 36], however no such experiments have involved importin-α.

4.4.5 Examination of the importin-α:PLSCR1 NLS peptide complex

The presence of *E. coli* ribosomal proteins in purified importin-α (70-529) samples used for crystallization and X-ray diffraction experiments has not currently

been considered in the literature. Because of this we feel certain that the complicating effects of rogue ligand binding have occurred before, but went unnoticed. We draw reference to a particular study by Chen *et al.* 2005 [14], which presents the structure of *Mus musculus* importin-α (70-529) bound to a non-classical NLS sequence from phospholipid scramblase 1 (PLSCR1, PDB:1Y2A). The importin-α construct used in co-crystallization is identical to our own (mouse importin-α isoform 2, residues 70-529) in the same pET-30a vector and purification followed standard protocols in the literature which, as stated, do not utilise ion exchange chromatography nor size exclusion chromatography. The non-classical NLS studied was a 10 residue long sequence from PLSCR1, ^{257}G**K**IS**K**HWTGI^{266}, a weakly basic motif comparable to the NLS-like motif, ^{24}A**K**FE**R**Q^{29}, found in the N-terminal *linker* region of the truncated importin-α construct used in the experiment.

It was reported that the structure of the importin-α:PLSCR1 NLS complex was determined at 2.2 Å in the space group $P2_12_12_1$, and structure factors were deposited to the protein database. While describing their structure Chen *et al.* mention that,

> "After several rounds of manual building and refinement all PLSCR1 NLS residues were clearly modeled in the electron density." [14],

but as we now demonstrate, the consideration necessary to make such a statement has not been duly applied. We used the deposited structure factors to analyse $2mF_o$-DF_c and mF_o-DF_c weighted electron density maps and noted many similarities with our own (no-SEC) importin-α (70-529) structure. While there is main chain density that corresponds to a peptidic molecule, side chain density is weak except for the same two binding slots, P2 and P5, which accept the only two lysine residues in the PLSCR1 NLS (Figure 4.27B). A serine at P4 and histidine at P6 poorly fit available density. A low RSCC score for the PLSCR1 peptide as presented in Section 4.3.4 confirms a general ambiguous fit of residues over binding slots P1 - P6. Beyond P6 there is indeed no unambiguous density to suggest the peptide chain continues, but despite this the model includes C-terminal residues made up of an unusual NLS C-terminal sequence, WTGI, which is indicated to have extensive interactions with importin-α (70-529).

To show that the PLSCR1 NLS has been presumptuously modelled in the importin-α major binding site, we have replaced it with model of the *E. coli* ribosomal protein S21 and run several rounds of refinement. The results are shown in Figure 4.27C. Although the binding site may be occupied by a mixture of peptides, we observe that the S21 model fits the density just as well as the PLSCR1 NLS. To confirm that the peptide is not present at the site, we calculated the $F_o^{plscr1} - F_o^{apo}$ data-data difference Fourier. Similar data-data difference Fouriers have been described for the importin-α:SV40 NLS complex and importin-α:S21 complex in Section 4.3.2 and 4.3.3, respectively. The $F_o^{plscr1} - F_o^{apo}$ map shows there is only ambiguous density for a bound PLSCR1 NLS peptide (Figure 4.27A).

The data-data difference Fourier is strong evidence that the importin-α:PLSCR1 NLS structure has only managed to replicate our early findings where unclear electron density results from an incomplete importin-α (70-529) purification. Even at a low Fourier map contour level of 2.0σ, there is no main chain peptide density and certainly no side chain density that can distinguish this peptide from any other. It therefore appears that the structure of the importin-α:PLSCR1 NLS peptide complex is the result of crystallography work lacking proper controls, leading to a premature or false understanding of the interaction between the PLSCR1 NLS peptide and importin-α.

To determine if this problem is plaguing other experiments involving importin-α complexes is difficult since early structures do not have deposited structure factors to allow close examination of the structure. In these cases it seems we are to assume a sceptical researcher would have noticed inconsistencies if they had arisen. We note that other importin-α:NLS structures accompanied by deposited structure factors do display clear density for the modelled peptide bound to importin-α [53, 67].

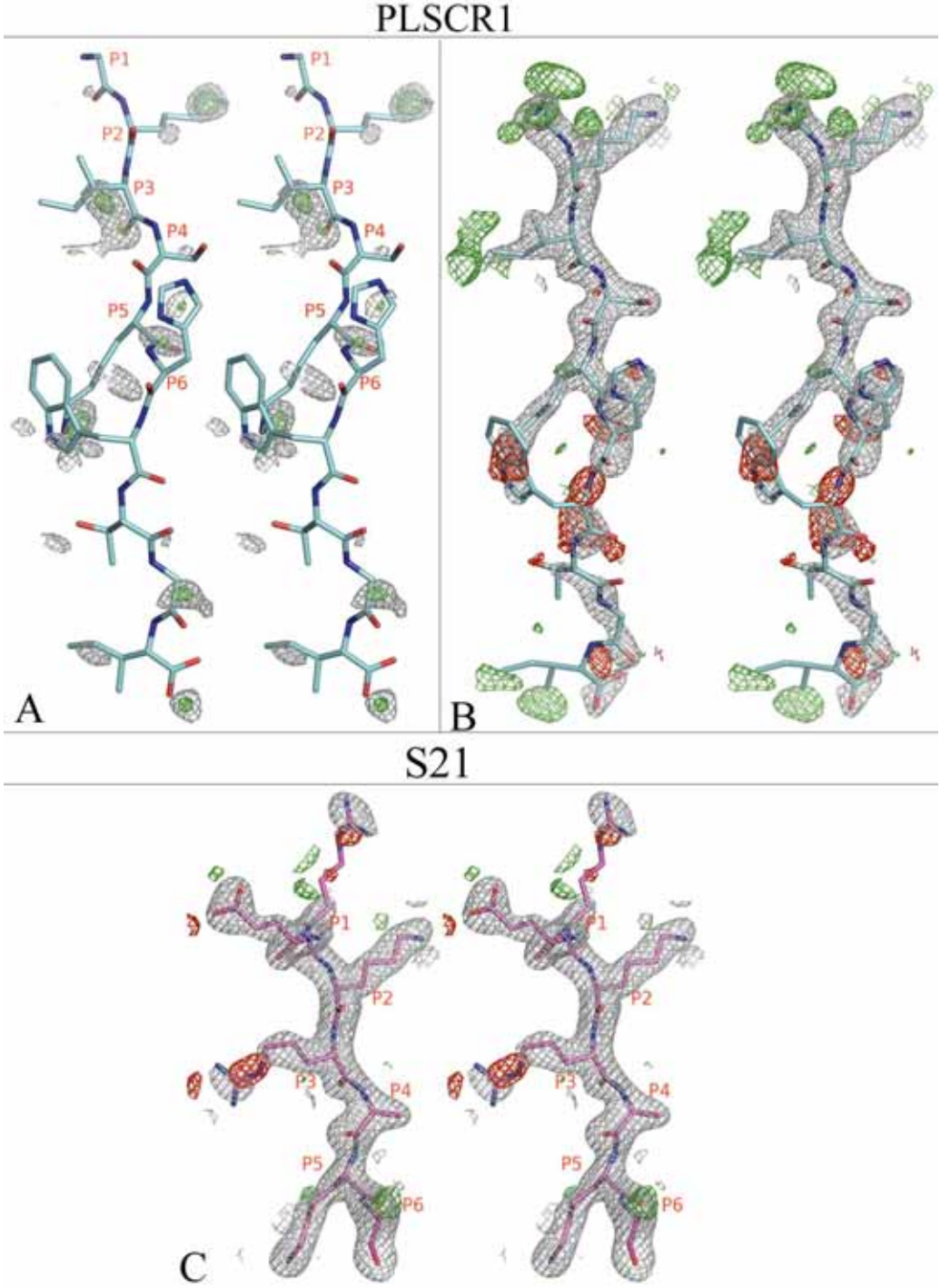

Figure 4.27: Stereo views of the published importin-α:PLSCR1 NLS complex with accompanying electron density.

All maps shown have been calculated using structure factors deposited with the importin-α:PLSCR1 structure (PDB:1Y2A, [14]). The figure was created using PyMOL [165].

A) The $F_o^{plscr1} - F_o^{apo}$ data-data difference Fourier map. Positive density contours are displayed at 2.8σ and 2.0σ (green and grey, respectively).
B) The importin-α:PLSCR1 $2mF_o$-DF_c (grey, 1.5σ) and mF_o-DF_c (green +2.8σ, red -2.8σ) weighted maps.
C) The S21 peptide refined into the major binding site using importin-α:PLSCR1 data. Map contours are the same as in panel B.

4.5 Conclusion

The NLS binding specificity of importin-α is a complex issue that has received a significant amount of attention in the literature. A well known binding partner of importin-α is the SV40 large T-antigen, which contains a classical NLS represented by the sequence motif ^{126}PKKKRKV^{132}, forming the basis on which all other potential NLS sequences are compared. The long stretch of basic residues provides an ideal complimentary surface charge to negatively charged pockets on the importin-α binding surface. The alignment of the peptide to the binding region is also important, since the presence of a lysine at binding slot P2 is critical for nuclear import [60, 64, 70] and all importin-α:NLS structures currently solved are modelled with a lysine here. Other residues within the NLS sequence are found to affect nuclear import to a lesser degree, particularly P3, P4 and P5 [62]. Studies have attempted to identify the required NLS sequence for binding importin-α and the results suggest a short 4 residue motif binding across P2 - P5 forms the most significant interaction. An optimal NLS sequence has been suggested through both mutational and structural analysis giving a similar consensus motif: ^{P2}K(K/R)x(K/R)^{P5} [62] or ^{P2}KRRK^{P5} [50].

We have discovered that endogenous *E. coli* proteins are capable of binding importin-α in an NLS-like manner, somewhat unexpectedly since there is unlikely to be a functional role for an NLS in prokaryotes. However, through a combination of mass spectrometry and structural modelling considerations, an NLS-like motif in the 30S ribosomal subunit S21, ^{P1}RKRAKA^{P6}, was shown to be binding to the importin-α (70-529) major binding site, commensurate with the consensus NLS sequence. It may be that the 30S ribosomal subunit S11 also binds importin-α (70-529), possibly as an S11:S21 heterodimer which is disordered in the crystal structure. The only ordered regions of the S11:S21 complex could correspond to where S21 interacts with the major binding site and at Cys133, where S11 may form a disulfide bond with importin-α (70-529). Such a possibility opens up an interesting avenue of further study, as there are few importin-α complexes solved with full length substrates. The use of small angle X-ray scattering (SAXS) to probe for an S11:S21 dimer bound to importin-

α may be a suitable technique in this instance since S11:S21 is likely to be quite mobile.

As we established in Chapter 3, the binding of S21 to importin-α (70-529) creates difficulties in obtaining structures of the importin-α:SV40 NLS complex and importin-α:CLIC4 NLS complex, by complicating interpretation of electron density in the binding site. This problem may be affecting other investigations of importin-α:NLS interactions, with at least one published example of an inaccurate structure. The work presented here should aid future structural studies of importin-α complexes.

By developing a purification protocol that removed *E. coli* contaminants, we were able to solve a high resolution *apo* importin-α (70-529) structure with an empty binding site. The success in removing bound *E. coli* proteins from *apo* importin-α (70-529) was confirmed with the importin-α:SV40 NLS peptide complex solved at 1.85 Å. A thorough analysis of the SV40 NLS peptide was performed using a range of novel methods including the $F_o^{sv40} - F_o^{apo}$ data-data difference Fourier and the normalised B-factor score, B_z^{-apo}. These methods relied on information obtained from the *apo* importin-α (70-529) structure and thus the *apo* form serves as a useful and necessary control in establishing the authenticity of our importin-α:NLS structures.

Numerous importin-α:NLS peptide crystal structures have been presented in the literature, each demonstrating a mode of NLS binding that overlaps in an almost identical fashion. Indeed, the similarity of NLS docking to importin-α belies its ability to recognise diverse target sequences and function as a receptor for multiple proteins. However there is always the necessity for a strong core of basic residues to be present in a specific sequence that aligns these structurally similar long basic side chains in identical binding slots. This emphasises the increased duty of care required of the crystallographer to identify NLS residues based on the observed electron density, rather than an assumed outcome.

Chapter 5

The importin-α:CLIC4 NLS peptide complex

5.1 Introduction

The transport of proteins to the nucleus via the importin-α:β nuclear import pathway is one of the best understood nuclear trafficking systems in the cell. The pathway operates through the importin-α receptor that recognises and directly binds to cargo protein in the cytoplasm. The importin-α:cargo complex then binds to importin-β, which is primarily responsible for negotiating passage through the Nuclear Pore Complex (NPC). This transport process is dependent on the ability of importin-α to recognise specific NLS sequences presented by the cargo protein in preparation for nuclear import. The acidic environment of the importin-α binding sites confers a high affinity to clusters of basic residues in the NLS sequence. Monopartite NLS sequences consist of a single cluster of basic amino acids approximately six residues long, which generally interact with the major binding site in importin-α. Structural studies have shown that an NLS binds importin-α in an extended conformation suggesting functional NLSs found within structured regions of a cargo protein first need to be unfolded and thus made flexible in preparation for binding. A specific interaction between importin-α and CLIC4 was proposed by Suh *et al*. 2004 [123].

A number of CLIC4 binding partners have previously been identified including: the protein scaffolding proteins AKAP350 and 14-3-3 isoforms, the neurotransmitter mediator histamine H3R and dynamin I, which is involved in endocytosis. The structure of a soluble form of CLIC4 has been solved which shows it adopts the canonical GST fold with an N-terminal thioredoxin-like domain and all α-helical C-terminal domain [86]. Although the function of CLIC4 is not well known, it can form poorly selective anion channels that are sensitive to redox conditions. Like other CLIC family members, it is hypothesised that CLIC4 undergoes a structural transition from the soluble form to an integral membrane form. CLIC4 is functionally important in the cell and has recently been implicated in angiogenesis with the observation that

suppressed CLIC4 expression leads to the disruption of tubular morphogenesis [129]. Within renal, ovarian and breast cancer cells CLIC4 expression is reduced, while in tumour stroma CLIC4 is up-regulated. It is suggested that the regulation of CLIC4 by cancer pathogenic factors indicates it may be a suitable target for cancer therapy [204].

Upon cellular stress, CLIC4 translocates to the nucleus where it is involved in apoptosis [123]. Due to the role of importin-α as an adaptor molecule in facilitating access to the nucleus, it was suggested that the stress induced CLIC4 nuclear transport pathway involves a direct interaction with the importin-α receptor. Mutagenesis of a cluster of basic residues in the CLIC4 putative NLS site (^{199}KVVAKKYR^{206} to ^{199}TVVAITYG^{206}) was sufficient to prevent nuclear translocation, suggesting this monopartite NLS-like sequence plays an active role in nuclear transport and supporting an importin-α mediated CLIC4 import pathway. The full length CLIC4 soluble structure shows that the putative NLS has a helical conformation packed tight against the core of the protein that would need to undergo some structural rearrangement in order to bind importin-α with an extended conformation.

In this chapter we present the X-ray crystal structure of importin-α (70-529) bound to a peptide corresponding to the CLIC4 NLS. The importin-α (70-529) construct lacks the N-terminal autoinhibitory domain so that it does not compete with peptide binding and complicate observations of any interaction with the CLIC4 NLS peptide. The crystal structure shows that the CLIC4 NLS peptide binds to the importin-α major binding site in an extended conformation consistent with a classical importin-α:NLS complex. There is no clear interaction between the CLIC4 NLS peptide and the minor binding site of importin-α.

In the major binding site, electron density clearly defines peptide residues ^{201}VAKKYRN^{207}, which have been included in the final model with Lys203 occupying the critical P2 binding slot. Our results reveal that Lys199 at the putative CLIC4 NLS N-terminus is disordered in the crystal and is therefore not necessary for peptide binding. The core binding slots P2 - P5 are occupied by residues KKYR, a rather atypical NLS sequence due to the presence of a bulky aromatic tyrosine in the P4 binding slot. Surprisingly, the Tyr205 side chain is favourably placed at P4 forming

hydrogen bonds with the importin-α main chain. An analysis of normalised B-factors demonstrates a localised reduction in atomic flexibility experienced by importin-α residues due to the binding of the CLIC4 NLS peptide. The changes to residue flexibility are equivalent to those measured in the importin-α:SV40 NLS complex.

The importin-α:CLIC4 NLS structure presented in this chapter adds to the growing body of knowledge on the structural mechanisms that govern the classical nuclear import model. It also clarifies that the CLIC4 NLS sequence can indeed directly bind to importin-α, on condition that it can unfold into an extended conformation, commensurate with the proposed CLIC4 nuclear import mechanism that occurs in response to cellular stress.

5.2 Methods

5.2.1 Expression and purification

The expression and purification of *Mus musculus* importin-α (70-529) isoform 2, has been described previously in Sections 4.2.1-4.2.3. All importin-α samples used for experiments described in this chapter are *apo* importin-α (70-529), which has been separated from *E. coli* contaminants using the improved purification protocol. Full length CLIC1 and CLIC4 proteins used in the blue native-PAGE binding assay were prepared using similar protocols described in Section 2.2.2.

Two CLIC4 NLS peptides were synthesised (Sigma-Genosys) and used in crystallizing importin-α:CLIC4 peptide complexes. One comprised NH_2-^{198}VKVVAKKYRN^{207}-COOH and is termed the CLIC4 NLS peptide, and the second comprised NH_2-^{58}DLKRKPAD^{65}-COOH and is called the CLIC4$^{58\text{-}65}$ peptide. While the proposed CLIC4 NLS motif includes residues 199-206 (^{199}KVVAKKYR^{206}) [123], the CLIC4 NLS synthetic peptide used for crystallization included one residue extensions at both termini (NH_2-^{198}V<u>KVVAKKYR</u>N^{207}-COOH). The synthesis of this extended NLS peptide was to ensure that importin-α binding slots were fully occupied and that possible effects from flanking residues could be taken into account.

The CLIC4$^{58\text{-}65}$ peptide was designed primarily to act as an experimental control, assuming a lack of binding would occur between this peptide and importin-α. The CLIC4 sequence contains a second cluster of basic residues, ^{60}KRK^{62}, which is represented in the synthetic peptide, NH_2-^{58}DL<u>KRK</u>PAD^{65}-COOH. Properties of the two synthetic CLIC4 peptides are shown in Section 3.2.2.

5.2.2 Blue native-PAGE binding assay

The methods used for running blue native-PAGE (BN-PAGE) have been described previously in Section 3.2.5. Isolated samples of *apo* importin-α (70-529), CLIC1 and CLIC4 were run on the gel separately, at a concentration of 10 μM in each

case. To test for a binding interaction between the CLICs and importin-α (70-529), samples were prepared by incubating 1:1 molar ratio mixtures of importin-α (70-529) with CLIC1 or CLIC4, at room temperature for 10 minutes before loading on the gel. Alternatively, mixtures were incubated at 37 °C for 10 minutes before loading.

5.2.3 Crystallization and data collection

Importin-α (70-529):CLIC4 NLS (198-207)

The importin-α (70-529):CLIC4 NLS crystals were obtained by co-crystallization using hanging drop vapour diffusion. The reservoir consisted of 500 µL of 0.7 M sodium citrate, 10 mM DTT and 70 mM Hepes at pH 7.4. Drops consisted of 0.75 µL of importin-α (70-529) at 18 mg/ml, 0.75 µL of CLIC4 NLS peptide at 3 mg/ml and 1 µL of reservoir. These conditions result in a 7.6 x molar excess of the CLIC4 NLS peptide over importin-α (70-529). There was no incubation of peptide and protein other than the immediate mixing in the crystallization drop. Crystals grew to maturity in approximately two weeks at 20 °C, with the largest attaining final dimensions of 0.5 x 0.1 x 0.1 mm.

Importin-α (70-529) + CLIC4 peptide (58-65)

Co-crystallization trials of the $CLIC4^{58-65}$ peptide and importin-α (70-529) using similar conditions described above were not successful: therefore a peptide soaking technique was used. Crystals of *apo* importin-α (70-529) were first grown in 1 µL drops as described above. Then 1 µL of $CLIC4^{58-65}$ peptide at 4 mg/ml was added to the drop, giving a 6.5 x molar excess of peptide over importin-α. The peptide was allowed to soak for 3 days at 20 °C. Longer periods of soaking led to visible crystal deterioration and weaker X-ray diffraction.

Crystal handling and Data Collection

Since importin-α crystals are prone to cracking they were gradually transferred into a cryoprotectant solution that consisted of the reservoir solution supplemented

with 20% glycerol. The transfer was performed over three steps. Firstly the crystal was moved into a fresh reservoir solution and allowed to stand for one minute. The crystal was then moved to a 50:50 mixture of reservoir and cryoprotectant solutions, and allowed to stand for one minute before being moved into a 100% cryoprotectant solution.

The importin-α:CLIC4 NLS crystals were flash cooled in liquid nitrogen and stored in a cryogenic dry shipping dewar. They were then subsequently mounted at 100 K in a nitrogen cryostream on beamline PX-1 (3BM-1B) at the Australian synchrotron and data collection was performed using an ADSC Quantum 210 (Q210) detector. The importin-α:CLIC4$^{58\text{-}65}$ crystals were flash cooled in a 100 K nitrogen cryostream on our laboratory X-ray source (or home source, HS), which is a Nonius rotating anode generator using Cu K_α radiation and osmic confocal mirror optics with a Mar345 image plate detector.

A number of initial images were screened at orthogonal φ-angles to gauge crystal quality before auto-indexing and determining the optimum rotation range using STRATEGY from within MOSFLM [180]. Crystals were found to have symmetry in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of approximately, a = 79 Å, b = 90 Å and c = 100 Å (see Table 5.1). The final diffraction dataset was obtained over 180 images with a 1° oscillation angle (φ) and 5 second exposure time per image (see Figure 5.1). The crystal to detector distance was set at 220 mm and the beam width set to 200 x 200 μm.

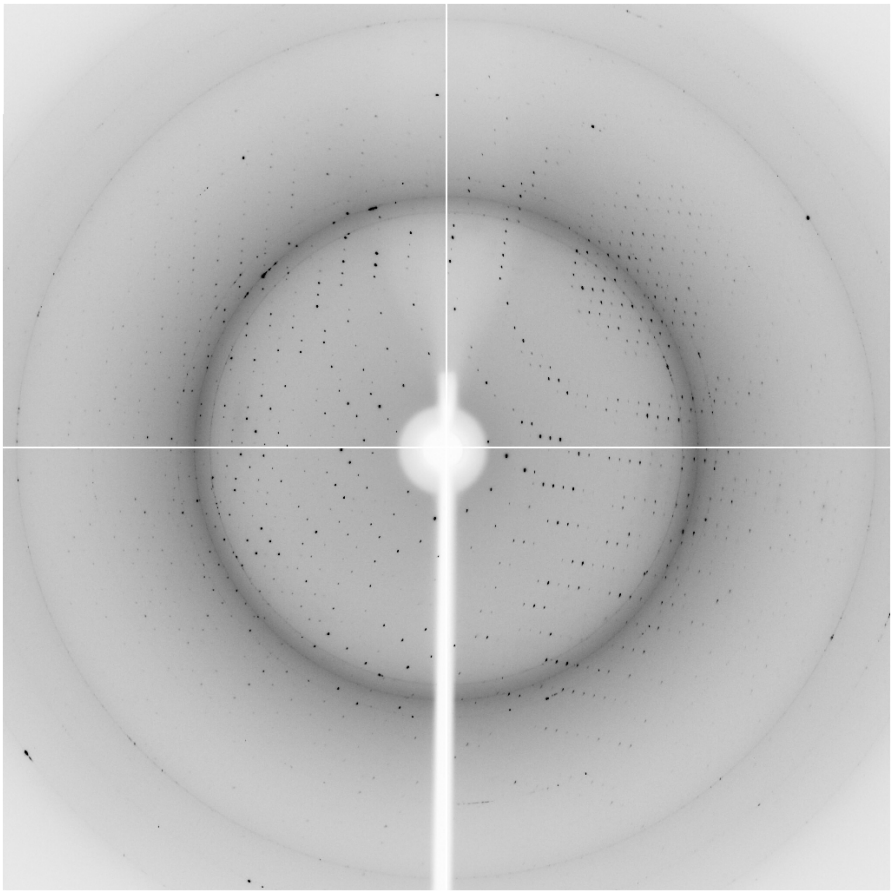

Figure 5.1: Synchrotron X-ray diffraction pattern of the importin-α:CLIC4 NLS crystal.

Shown is a 1° oscillation image taken on an ADSC Q210 detector at the Australian Synchrotron in Melbourne. The crystal to detector distance was set at 220 mm. The edge of the image corresponds to 2.2 Å.

5.2.4 Structure determination and refinement

Structure determination and refinement of importin-α:CLIC4 peptide complexes closely followed the protocols described in Section 3.2.7. Data reduction and refinement statistics are summarised in Table 5.1.

Table 5.1: Data reduction and refinement statistics for structures of importin-α:CLIC4 peptide complexes.

	Importin-α:CLIC4 NLS (199-206)	Importin-α:CLIC4 (58-65)
Data Collection		
Source (λ)	AS BL PX-1, 3BM-1B (0.95 Å)	HS (1.54 Å)
Detector	ADSC Q210	Mar345
Space Group	$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)		
a	78.6	78.5
b	89.6	89.8
c	100.1	99.1
Resolution (Å)[a]	2.0 (2.11-2.00)	2.35 (2.48-2.35)
Total observations	318,168	128,569
Unique observations	46,758	29,157
Completeness (%)[a]	96.7 (80.3)	98.0 (96.6)
Mean I/σ[a]	12.2 (1.8)	12.7 (2.0)
R_{merge} (%)[a,c]	9.6 (77.6)	11.2 (63.4)
Wilson B-value ($Å^2$)	30.4	41.5
Refinement and Structure		
R-factor (%)[c]	20.0	20.8
R_{free} (%)[c]	23.6	24.9
Number of atoms (Avg. B-value, $Å^2$)		
Importin-α	3244 (33.8)	3258 (39.3)
Peptide	62 (38.7)	-
Waters	382 (46.2)	82 (38.0)
Matthews Coefficient (solvent)	3.2 $Å^3$/Da (61.4%)	3.2 $Å^3$/Da (61.3%)
Ramachandran plot (%)[b]		
Favoured region	99.1	98.4
Allowed region	0.9	1.6
Disallowed	0	0

[a] Outer shell statistics are shown in parenthesis.
[b] Calculated using MolProbity [159].
[c] The R_{merge} and R factor are defined in Section 2.2.7.1, p72.

5.3 Results

5.3.1 Binding assay of full length CLIC1/CLIC4 and importin-α (70-529)

We first sought to establish if it were feasible to obtain a full length CLIC4:importin-α (70-529) complex by examining the interaction of the two proteins under native conditions. This was performed using blue native-PAGE (BN-PAGE) to assess the interaction and determine whether resulting complexes were suitable for further study involving protein crystallization and structure determination. In this experiment, the pH (7.5) is close to physiological conditions and the ions $BisTris^{+}$ and $Tricine^{-}$ are present in the protein environment. Mixtures of importin-α (70-529)/CLIC1 and importin-α (70-529)/CLIC4, incubated at room temperature or at 37 °C before being analysed on the gel, are shown in Figure 5.2.

For comparison, isolated samples of importin-α (70-529), CLIC1 and CLIC4 were run on the BN-PAGE as shown in Figure 5.2. The importin-α (70-529) sample shows signs of distinctive high molecular weight (MW) aggregates on BN-PAGE, with the monomeric species running ~50 kDa higher than the theoretical MW of 55.3 kDa. Similar behaviour is observed for CLIC1 and CLIC4 that run ~15 kDa and ~50 kDa higher than their theoretical MW (~30 kDa), respectively. The discrepancy between the BN-PAGE MW and theoretical MW should not confuse the interpretation of complex formation. Although CLIC1 runs as a single and well defined band, we note that the CLIC4 monomer corresponds to a smeared band with evidence of higher MW aggregates.

The incubated mixtures do not appear to reveal any new additional bands on the BN-PAGE corresponding to complex formation and incubation of protein samples at 37 °C resulted in no discernable differences compared to room temperature incubation. High MW importin-α (70-529) aggregates are apparently not present in the mixtures suggesting they have dissociated, most likely due to a non-specific interaction with the CLIC proteins, although an explanation for this is unclear. CLIC1 and CLIC4 both appear unmodified in their behaviour when incubated with importin-α (70-529).

The BN-PAGE binding assay suggests that it is unlikely a distinctive full length importin-α:CLIC4 complex forms under native conditions. In order to ultimately obtain the crystal structure importin-α:CLIC4, appropriate environmental conditions and structural modifications to CLIC4 necessary to bind importin-α need to be first determined. This is a demanding task that we decided could be pursued after first obtaining confirmation that the CLIC4 NLS does indeed bind to importin-α, as predicted. Therefore our efforts concentrated on obtaining the structure of the importin-α:CLIC4 NLS peptide complex.

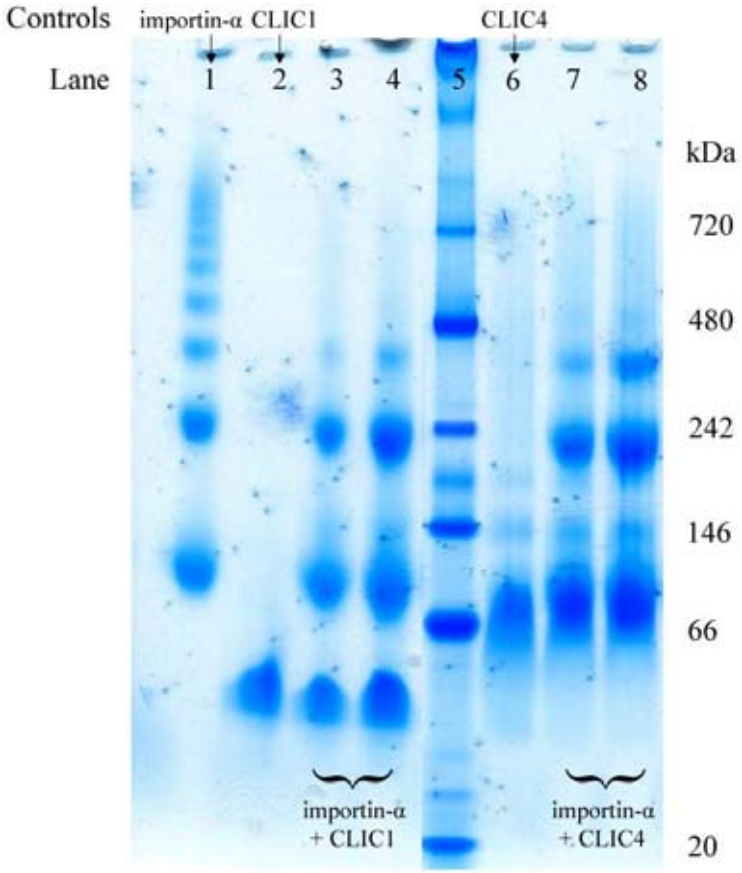

Figure 5.2: BN-PAGE binding assay for importin-α (70-529), with full length CLIC1 and CLIC4.

Isolated importin-α (70-529), CLIC1 and CLIC4 are run in lanes 1, 2 and 6, respectively. Lane 3 contains the room temperature incubation of importin-α and CLIC1, lane 4 contains the 37 °C incubation. Lane 7 contains the room temperature incubation of importin-α and CLIC4, lane 8 contains the 37 °C incubation. NativeMark™ protein standards are in lane 5, with corresponding molecular weights on the right.

5.3.2 The structure of importin-α (70-529) soaked with the CLIC4 (58-65) peptide

To determine if NLS-like sequences in CLIC4 can bind importin-α other than the putative NLS (^{198}VKVVAKKYRN^{207}), we looked at an N-terminal domain motif (^{58}DLKRKPAD^{65}) that is present in the CLIC4 connecting loop between β-strand 2 and α-helix 2. This motif compares with the putative CLIC4 NLS motif that is present in the C-terminal domain and which adopts helical conformation. The CLIC4$^{58-65}$ sequence is a loosely conforming NLS motif with three consecutive basic residues, meaning the core binding slot of P5 will be occupied by a non-basic residue. However the sequence is similar to the importin-α auto-inhibitory NLS-like motif (^{48}LKRRNV^{53}) which has been shown to bind to the major binding site [37, 59]. If the CLIC4 NLS-like motif (58-65) interacts with importin-α, it must preferentially bind with Lys60 in the P2 binding slot meaning binding slots P1 - P4 would be occupied by matching residues compared to the auto-inhibitory NLS-like motif.

A structure of importin-α has been solved by crystallizing *apo* importin-α (70-529) and then soaking in the CLIC4$^{58-65}$ peptide. This is a similar method to the one used in obtaining the yeast importin-α:SV40 NLS structure at 2.8 Å [30]. Our structure has been solved at a lower resolution (2.35 Å) than other *apo* importin-α:NLS structures mainly due to crystal degradation upon soaking the CLIC4$^{58-65}$ peptide. The structure includes importin-α residues 70-496 clearly defined in the $2F_o$-F_c electron density map. The average atomic B-factor for importin-α atoms is 39.3 Å^2 and for waters it is 38.0 Å^2.

Inspecting the F_o-F_c difference map in the region of the major binding site shows there is no indication of a bound CLIC4$^{58-65}$ peptide (Figure 5.3A). This is also the case in the region of the minor binding site. The weak density in the major binding site corresponds to that seen in the *apo* structure, including density for a water molecule at P2 and main chain density breaks over the P3 - P4 binding slots. There is also negative density over residues in the tryptophan and asparagine arrays, suggesting a degree of atomic flexibility that has been previously noted in the *apo* importin-α (70-529) structure (Section 4.3.1).

To confirm there is no density corresponding to a CLIC4$^{58-65}$ peptide, we calculated the $F_o^{clic4(58-65)} - F_o^{apo}$ data-data difference Fourier (Figure 5.3B). The resulting difference electron density also shows no signs of a bound peptide down to the background level of noise in the map.

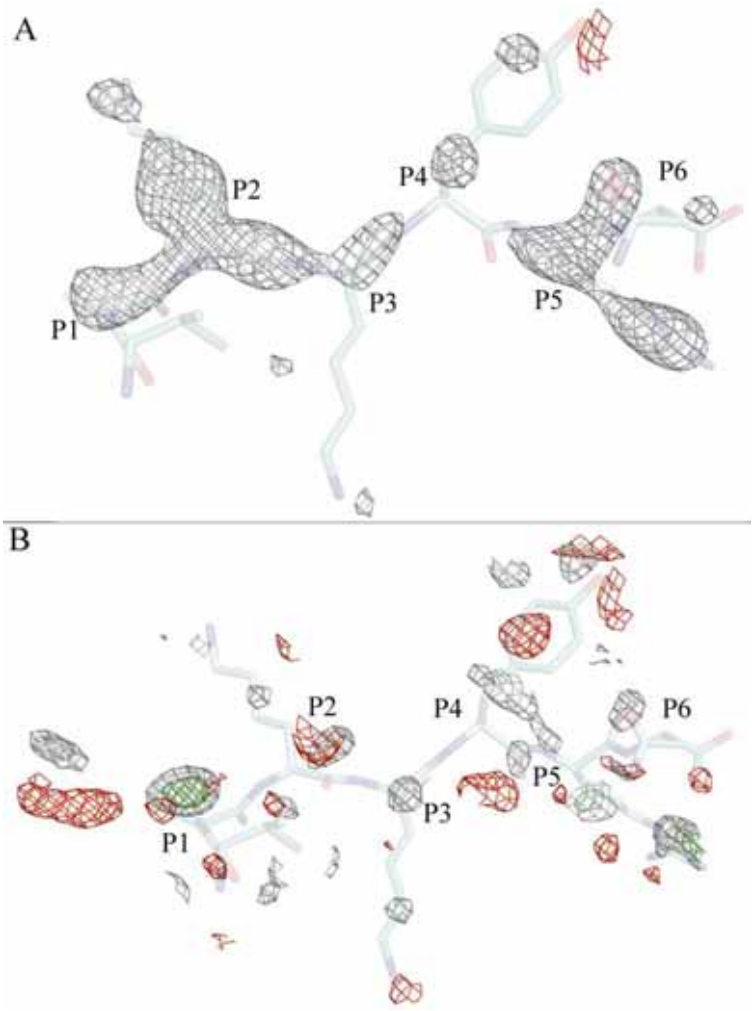

Figure 5.3: Electron density in the importin-α:CLIC4$^{58-65}$ major binding site.

A) The F_o-F_c difference electron density map in the region of the importin-α major binding site. The CLIC4 NLS has been superimposed for orientation. Positive contours are coloured in grey (+2.8σ), negative contours are coloured red (-2.0σ).

B) The $F_o^{clic4(58-65)} - F_o^{apo}$ data-data difference Fourier map in the importin-α major binding site. Contours are coloured as green +2.8σ, grey +2.0σ and red -2.0σ. The CLIC4 NLS has been shown to provide orientation.

5.3.3 The importin-α:CLIC4 NLS peptide complex

The importin-α (70-529) construct used to obtain the importin-α:CLIC4 NLS complex lacks the first 69 residues that correspond to the flexible importin-β binding domain (IBB). This domain is known to have an auto-inhibitory function whereby an NLS-like sequence competes for the importin-α binding site, reducing binding affinity for cargo proteins and also helping to facilitate release of cargo within the nucleus [37, 38, 58]. The removal of the auto-inhibitory domain to create a truncated importin-α avoids possible competition for the binding site between this internal NLS and the CLIC4 NLS peptide.

The preparation of importin-α (70-529), which followed standard protocols in the literature, left *E. coli* ribosomal proteins capable of interacting with importin-α (70-529) in the major binding site (Chapter 3). In order to remove this interaction, additional purification steps are necessary to separate out these endogenous *E. coli* proteins. The importin-α (70-529) sample containing an empty binding site has been termed '*apo* importin-α'. In order to obtain a complex through co-crystallization of importin-α (70-529) and the CLIC4 NLS, this *apo* importin-α sample was used.

The structure of importin-α (70-529) bound to the CLIC4 NLS peptide (^{198}VKVVAKKYRN^{207}) was solved at 2.0 Å using synchrotron radiation. Electron density corresponding to residues 201-207 (VAKKYRN) of the CLIC4 NLS peptide was unambiguously identified in the importin-α major binding site between ARMs 2-4. The F_o-F_c map constructed by omitting the peptide from model phases is shown in Figure 5.4. The importin-α minor binding site contains no electron density that unambiguously corresponds to the CLIC4 NLS and therefore no peptide was modelled at this site.

The model of importin-α in the CLIC4 NLS peptide complex includes residues 72-496 and closely resembles the full length importin-α structure that incorporates the N-terminal auto-inhibitory domain (PDB: 1IAL, RMSD of 0.20 Å across 425 C_α atoms in residues 72-496), as well as our *apo* importin-α (70-529) structure (RMSD of 0.15 Å across 425 C_α atoms in residues 72-496). The major backbone differences between the

apo importin-α (70-529) and the CLIC4 NLS peptide bound structure are restricted to the less well defined N-terminal and C-terminal regions. The major binding site spanning ARMs 2-4 has a similar conformation compared to the equivalent region in *apo* importin-α (70-529), with an RMSD of 0.16 Å (46 C_α atoms) across the inner H3 helices, suggesting there are minimal backbone conformational changes due to peptide binding.

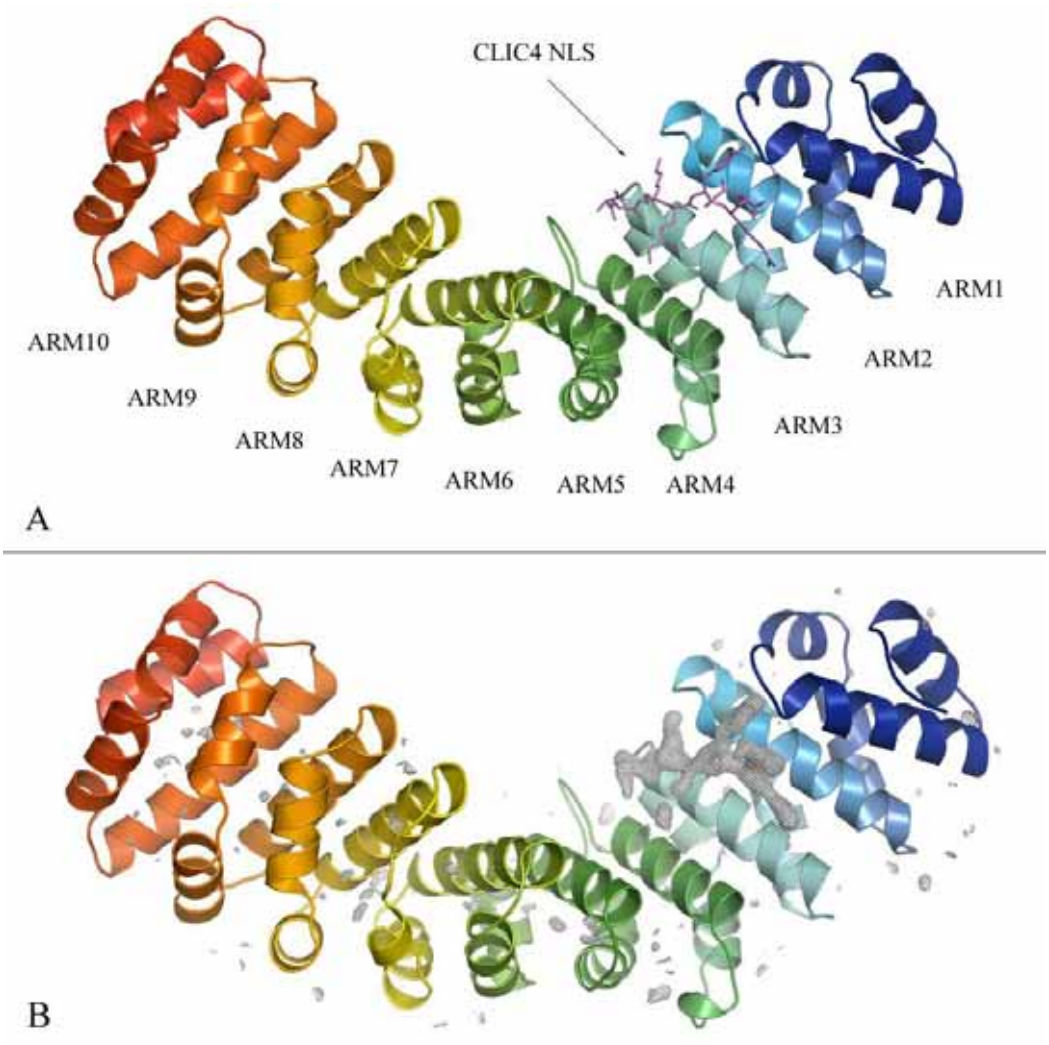

Figure 5.4: A global view of the importin-α:CLIC4 NLS peptide complex.

A) The importin-α molecule is shown with C-terminal domain ARM repeats labelled and depicted by a spectrum of colours from blue to red. The CLIC4 NLS is shown in stick representation (magenta).

B) The F_o-F_c difference electron density or 'omit' map over all atoms in importin-α. Positive contours are shown at 2.8σ in grey. Density corresponding to the bound CLIC4 NLS peptide is clearly visible in the major binding site.

The CLIC4 NLS binds in an extended conformation and is aligned in a similar orientation to that seen in previous importin-α:NLS structures, including our own importin-α:SV40 NLS structure (RMSD of 0.15 Å across six C_α atoms in residues occupying P1 - P6). The average atomic B-factor for importin-α in the structure is 32.1 $Å^2$ for main chain atoms, 35.7 $Å^2$ for side chain atoms and 33.8 $Å^2$ overall (3244 atoms). For the peptide, B-factors are slightly higher than those for importin-α: 36.9 $Å^2$ for main chain atoms, 40.2 $Å^2$ for side chain atoms, and 38.7 $Å^2$ overall (62 total atoms). This is in contrast to our importin-α:SV40 NLS peptide structure in which the SV40 NLS in the major binding site has a lower average B-factor (24.5 $Å^2$) than importin-α (30.8 $Å^2$).

5.3.3.1 Electron density analysis

The CLIC4 NLS peptide fits electron density in the major binding site well over both the main chain and side chain atoms (see Figure 5.5). Peptide residues 202-207 are built into the key binding slot positions P1 - P6, with the critical P2 slot occupied by Lys203. This means that the core basic motif, ^{203}KKYR^{206}, fills the central binding slots P2 - P5 where the majority of peptide side chain interactions take place with importin-α. This demonstrates that the CLIC4 NLS satisfies the accepted consensus sequence for an optimal NLS, ^{P2}K(K/R)x(K/R)^{P5} [62]. The residue at P4, which has been shown to energetically contribute the least to peptide binding out of the four main binding slots [62], is unambiguously occupied by Tyr205 as defined by $2F_o$-F_c electron density. This is the first importin-α:NLS structure that contains an aromatic peptide residue anywhere within the major binding site.

The N-terminal peptide residues, ^{198}VKV, are disordered in the crystal and are therefore likely to be highly flexible. This is particularly noteworthy since the basic residue, Lys199, defined as part of the putative CLIC4 NLS (KVVAKKYR) [123], does not contribute to peptide binding. If the N-terminal flanking region does increase CLIC4 NLS affinity for importin-α, it is unclear how it does so from the crystal structure. The terminating carboxyl group of the peptide at Asn207 is well defined by the $2F_o$-F_c electron density map and makes no interaction with importin-α.

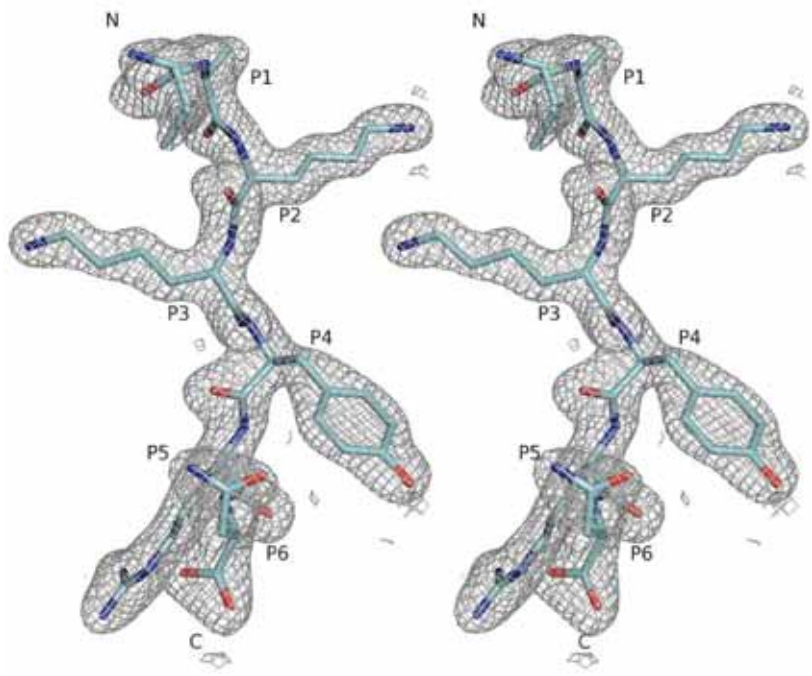

Figure 5.5: A stereo image of the CLIC4 NLS peptide and $2F_o–F_c$ map.

The CLIC4 NLS peptide bound to the importin-α major binding site is shown in stick representation. Electron density is contoured at 1.5σ in grey. Binding slot positions P1 - P6, and the N- and C-terminus are labelled.

We have analysed the veracity of the CLIC4 NLS model built in the major binding site by inspecting the $F_o^{clic4nls} - F_o^{apo}$ data-data difference Fourier, where $F_o^{clic4nls}$ refers to importin-α:CLIC4 NLS structure factors and F_o^{apo} refers to *apo* importin-α (70-529) structure factors, as described in Section 4.2.7.1. The results are shown in Figure 5.6A. As expected, the electron density is strong along the peptide main chain with well defined carbonyl and amide backbone groups. The one exception to this is in the location of the amide group of Lys204 at P3, where there is a break in the main chain density at a 2.8σ map level. The corresponding position in the *apo* structure has particularly strong density at this point. The fact there is weak density here in the $F_o^{clic4nls} - F_o^{apo}$ difference Fourier is probably not relevant to deficiencies in the CLIC4 NLS peptide model.

Peptide side chain density is also well defined in the $F_o^{clic4nls} - F_o^{apo}$ difference Fourier. At P1 density weakly corresponds to Ala202. The lysine residue at P2

(Lys203) is well defined despite the presence of a partially occupied water in this location in the *apo* structure. P3 is equally well defined for Lys204, a position that lacks any side chain density in *apo* importin-α. P5 represents Arg206 sufficiently well even though the *apo* structure contains density corresponding to a long side chain in this binding slot. P6 loosely fits Asn207.

Perhaps the most definitive characteristic of the $F_o^{clic4nls} - F_o^{apo}$ difference Fourier is the strong and unambiguous electron density corresponding to Tyr205 at P4 (Figure 5.6B). There is a strong positive difference density peak near the Tyr205 O-C^{ζ} bond at the 9.1σ map level, the strongest density peak in the $F_o^{clic4nls} - F_o^{apo}$ difference map. Another strong positive density peak is present over Tyr205 C^{γ}-C^{δ} phenyl bonds at the 7.9σ level. Also, there is weak and unaccounted for density ~4 Å from the aromatic plane of Tyr205 in a position that may correspond to a cation - π interaction. The bonding distance is similar to a Na^+ - π interaction involving a tryptophan side chain which was also established to be ~4 Å from the aromatic plane [183]. We rule out the possibility of a water molecule corresponding to the density as there are no hydrogen bonding partners within 3.5 Å. The cation in this case is likely to be a Na^+ ion from the crystallization buffer, with a low occupancy (< 50%).

The strong $F_o^{clic4nls} - F_o^{apo}$ density over the tyrosine side chain at P4 is undoubtedly aided by the presence of only weak density in this binding slot in the *apo* structure. In the importin-α:S21 structure, an alanine residue fits at P4, making it an ideal location to conclusively differentiate between it and the importin-α:CLIC4 NLS complex. It is also the case that electron density for a bulky aromatic side chain is not easily confused with the more common lysine and arginine NLS residues. The presence of Tyr205 is definitive evidence that the CLIC4 NLS peptide binds importin-α (70-529).

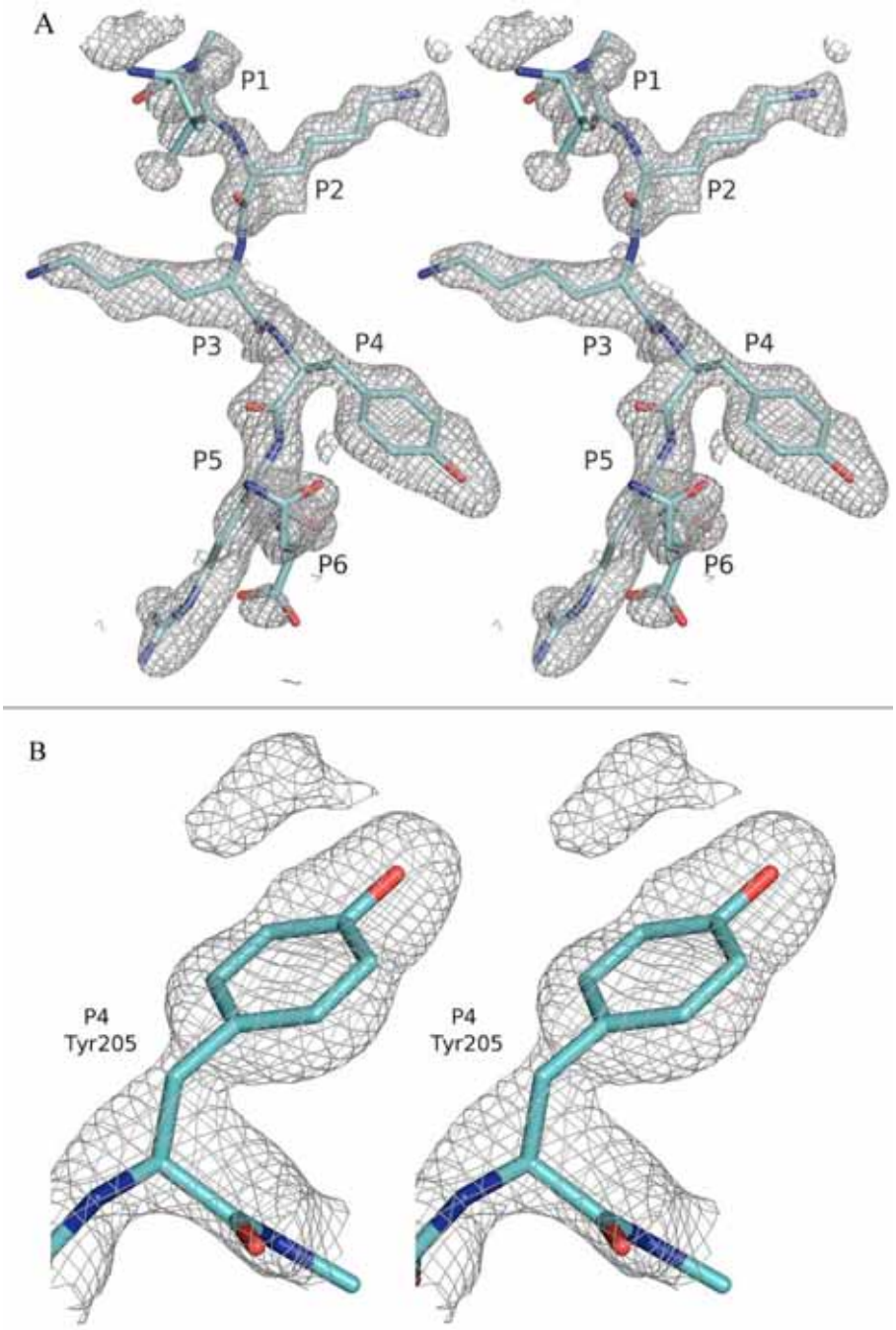

Figure 5.6: Stereo images of the CLIC4 NLS bound to importin-α, showing the $F_o^{clic4nls} - F_o^{apo}$ data-data difference Fourier map.

A) The data-data difference Fourier map is shown near the CLIC4 NLS peptide. Grey contours represent positive difference density at 2.8σ.

B) The same map from panel A is shown here in the vicinity of Tyr205 in the P4 major binding slot. Electron density above the phenyl ring may correspond to a metal cation - π interaction.

5.3.3.2 CLIC4 NLS interactions with importin-α

The CLIC4 NLS forms an intricate network of interactions with importin-α through both main chain and side chain atoms, similar to previous importin-α:NLS structures [30, 37, 50, 53]. Hydrogen bonding by the main chain of the CLIC4 NLS peptide involves importin-α side chains in the conserved WxxxN motifs of ARMs 2 - 4, which includes residues Trp142, Trp184, Asn146, Asn188 and Asn235 (see Figure 5.7). In the CLIC4 NLS peptide, this corresponds to hydrogen bonded carbonyl and amide groups from each second residue in the major binding site: P1 (Ala202), P3 (Lys204) and P5 (Arg206). The peptide side chain in binding slots P2 (Lys203), P4 (Tyr205) and P5 (Arg206), form hydrogen bonds to the importin-α main chain and side chains. Additionally, ^{CLIC4}Lys203 forms a critical salt bridge with $^{imp\text{-}\alpha}$Asp192 (bond length 2.80 Å, N^{ζ} - O^{δ}) at P2, the most energetically significant interaction involved in importin-α recognition of NLSs [60, 64].

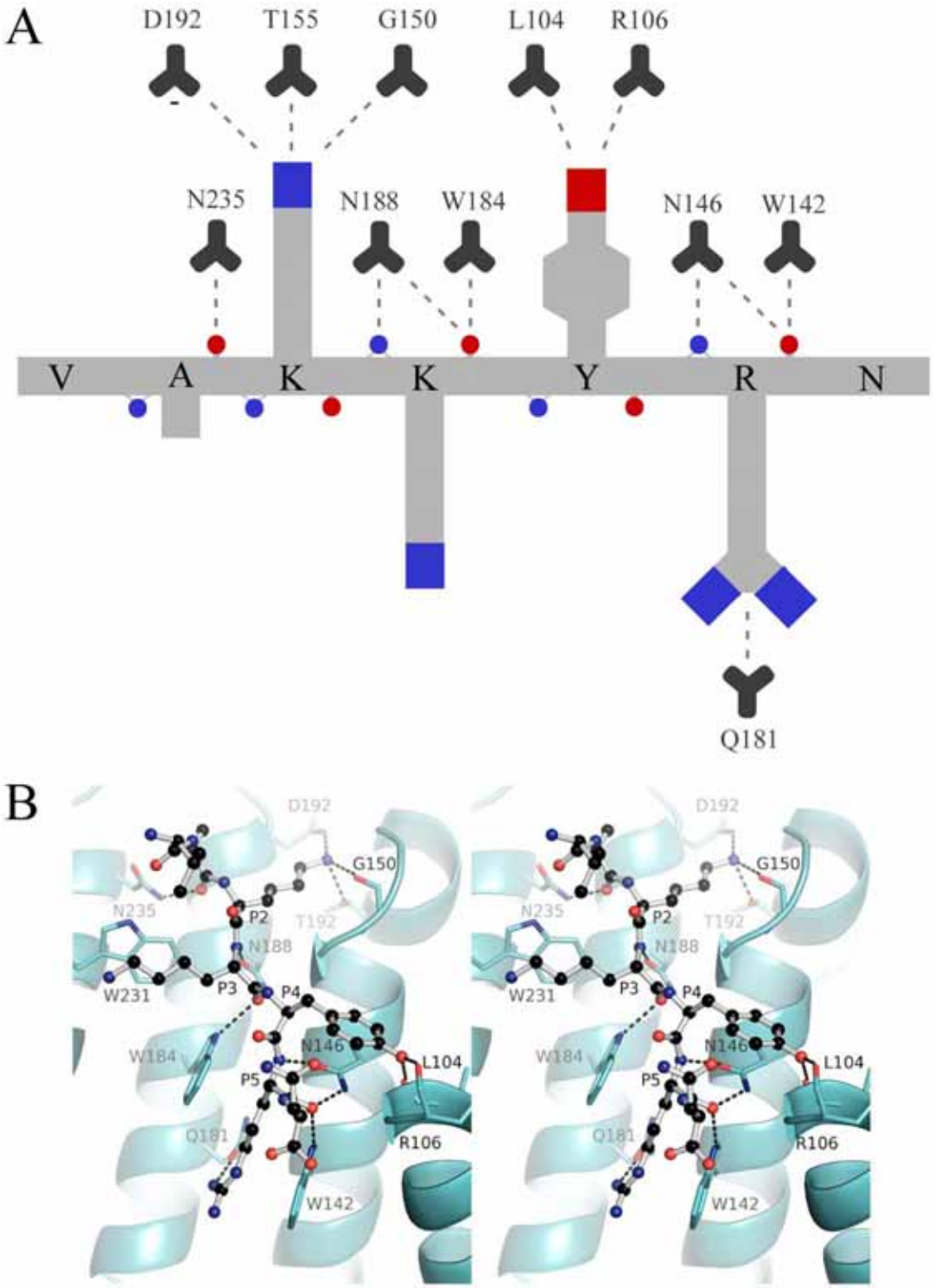

Figure 5.7: Hydrogen bonding between importin-α and the CLIC4 NLS.

A) A schematic representation of hydrogen bonds (dashed lines, < 3.5 Å) between importin-α and the CLIC4 NLS peptide, ^{P1}AKKYRN^{P6}. Backbone carbonyl oxygens and amide nitrogens are shown as red and blue spheres, respectively. Nitrogen and oxygen side chain atoms are shown as blue and red squares, respectively.

B) A stereo image of hydrogen bond interactions (dashed lines, < 3.5 Å). The CLIC4 NLS peptide is shown in ball and stick representation. Importin-α is shown in cartoon representation (cyan) with bonded residues shown as sticks.

Other basic residues in the peptide, ^{CLIC4}Lys204 and ^{CLIC4}Arg206, fill negatively charged pockets at P3 and P5 without forming salt bridges (Figure 5.8). At P3, there is a pocket formed between $^{imp\text{-}\alpha}$Trp184 and $^{imp\text{-}\alpha}$Trp231 which favourably accepts the extended and positively charged ^{CLIC4}Lys204 side chain, with $^{imp\text{-}\alpha}$Glu266 (5.8 Å, N^{ζ} - O^{ε} distance) and $^{imp\text{-}\alpha}$Asp270 (4.5 Å, N^{ζ} - O^{δ} distance) positioned at the end of the binding slot. Similarly at P5, a pocket formed between $^{imp\text{-}\alpha}$Trp142 and $^{imp\text{-}\alpha}$Trp184 favourably accepts ^{CLIC4}Arg206, with $^{imp\text{-}\alpha}$Glu180 (4.9 Å, $N^{\zeta 1}$ - $O^{\varepsilon 1}$ distance) positioned at the end of the binding slot.

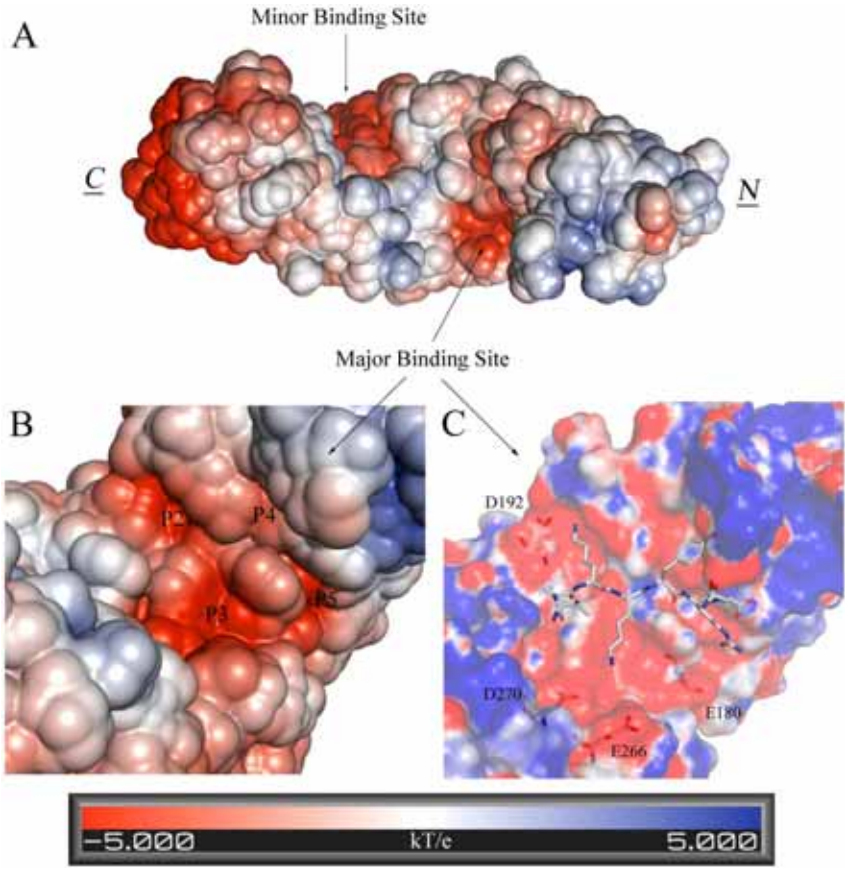

Figure 5.8: The electrostatic surface of the importin-α:CLIC4 NLS complex.

Colouring in all panels has been scaled from red (-5 kT/e) to blue (+5 kT/e). The CLIC4 NLS peptide is omitted from calculations performed using APBS [169]. Figures created using PyMOL [165].

A) The importin-α electrostatic potential is shown across the inner concave solvent accessible surface. The CLIC4 NLS peptide has been omitted from the figure. Major and minor binding sites are labelled, as are the N terminus and C terminus.
B) The importin-α major binding site is shown with electrostatic potential shown on the solvent accessible surface. The core P2 - P5 binding slots are labelled. The CLIC4 NLS peptide is omitted.
C) The electrostatic potential is shown on the transparent molecular or Connolly surface representation of importin-α. Key acidic residues and the CLIC4 NLS peptide are shown in stick representation.

In total, the CLIC4 NLS main chain and side chains make 174 atom to atom van der Waals contacts with importin-α and 13 hydrogen bonds which includes one salt bridge at the P2 major binding site. In addition to this, there are 10 hydrogen bonds formed between the peptide and water molecules, three of which involve the terminating carboxyl group. The van der Waals contact area between the CLIC4 NLS peptide and importin-α has been calculated for each peptide residue by integrating over contact dots using MolProbity [159, 188]. Main chain contributions to the contact area are all approximately the same (2 - 5 $Å^2$). Side chain contributions vary to a greater extent reflecting differences in the binding pockets. The lysine and arginine residues in binding slots P2 and P5 have the largest side chain van der Waals contact area, both corresponding to 15.9 $Å^2$. This is closely followed by the tyrosine at P4, with a contact area of 13.5 $Å^2$ and the lysine at P3 with a contact area of 10.2 $Å^2$. These results are summarised in Table 5.2.

Table 5.2: Characteristics of the importin-α:CLIC4 NLS peptide interaction.

NLS Residue	Interactions[a]		Van der Waals contact area ($Å^2$)[b]			Buried Surface Area ($Å^2$)	RSCC[c]
	Main Chain	Side Chain	Main Chain	Side Chain	Total		
Val201	0/1/0	0/0/0	0.4	0	0.4	26.7	0.89/0.74
^{P1}Ala202	1/8/0	0/6/0	3.5	5.3	8.8	89.4	0.91/0.81
^{P2}Lys203	0/6/1	3/31/1	3	15.9	18.9	145.3	0.96/0.97
^{P3}Lys204	3/5/0	1/20/0	2.2	8.0	10.2	128.4	0.98/0.92
^{P4}Tyr205	0/8/0	1/30/1	3.3	13.5	16.8	133.8	0.96/0.93
^{P5}Arg206	3/9/0	1/45/4	5.6	15.9	21.5	174.8	0.98/0.94
^{P6}Asn207	0/6/3	0/0/0	2.8	0	2.8	46.3	0.96/0.79
Totals	7/42/4	6/132/6	20.8	58.6	79.4	744.7	0.96/0.89

[a] Shown as importin-α hydrogen bonds/van der Waals contacts/ordered solvent hydrogen bonds.
[b] Calculated using MolProbity by integrating contact dots with 16 dots/$Å^2$ [159].
[c] The real space correlation coefficient (RSCC) is listed as main chain/side chain.

The solvent accessible surface area on importin-α buried by the peptide is 513.1 $Å^2$, with the largest contribution of buried surface area due to the tryptophan array that includes Trp142 (51.3 $Å^2$), Trp184 (70.2 $Å^2$) and Trp231 (63.7 $Å^2$). The surface area buried on the peptide is 744.7 $Å^2$, which corresponds to 59.2% of the total

peptide surface area. These results are comparable to the SV40 NLS peptide (^{126}PKKKRKV^{132}) in our importin-α:SV40 NLS structure which buries a surface area of 547.8 Å^2 on importin-α and 761.9 Å^2 on the peptide, which corresponds to 57.0% of the total peptide surface area.

The real space correlation coefficient (RSCC) for each residue has been calculated by comparing importin-α:CLIC4 NLS experimental electron density with density calculated from the model (Table 5.2). We see that the CLIC4 NLS main chain fits the density well, with an average RSCC of 0.96. The fitting of side chain density is more varied with the N-terminal residues (Val201, Ala202) and C-terminus residue (Asn207) having lower RSCC values (Avg. 0.78) than the core binding region comprising residues in binding slots P2 - P5 (KKYR, Avg. 0.94). These RSCC results over P2 - P5 are the same for our importin-α:SV40 NLS complex (KKRK, Avg. 0.94), showing that CLIC4 NLS Tyr205 fits just as well to the electron density as the long positively charged side chain of SV40 NLS Arg130.

5.3.3.3 The CLIC4 NLS tyrosine residue

By solving the structure of the importin-α:CLIC4 NLS complex, we show that the major binding site P4 slot is unambiguously occupied by a tyrosine residue: the first importin-α:NLS structure that has an aromatic residue present in the core binding region. Tyr205 adopts a common rotamer with a score of 82.9% ($\chi_1 \sim 180$, $\chi_2 \sim 80$) [159]. Interestingly, it appears that this unique NLS residue is not only accommodated in the P4 site, but contributes to the importin-α recognition of the CLIC4 NLS. This primarily occurs by the formation of a hydrogen bond between the side chain hydroxyl group of Tyr205 and the importin-α main chain at the C-terminus of the ARM1 H3 helix (see Figure 5.9). The hydrogen bond is possibly shared between the carbonyl oxygen of $^{imp\text{-}\alpha}$Leu104 (bonding distance from oxygen to oxygen, 2.83 Å) and the carbonyl oxygen of $^{imp\text{-}\alpha}$Arg106 (oxygen to oxygen, 2.64 Å). Although the amide nitrogen of $^{imp\text{-}\alpha}$Arg106 (N-O, 3.16 Å) is of hydrogen bonding distance, the geometry for this interaction is not favourable. Analysis of importin-α:CLIC4 NLS contacts using MolProbity [159] suggests the Tyr205-Arg106 (O-O) hydrogen bond is the most favourable with optimal geometry (Figure 5.9).

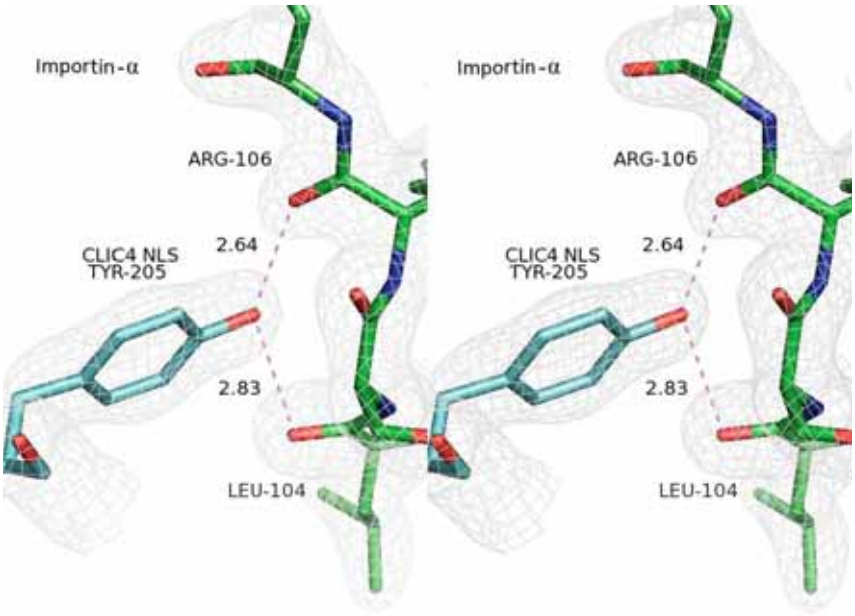

Figure 5.9: Stereo image of the CLIC4 NLS residue, Tyr205, in the importin-α P4 binding slot.

Hydrogen bonds are designated by magenta dashes and distances are shown in ångströms. The CLIC4 NLS is coloured cyan and importin-α is coloured green. The $2F_o$-F_c electron density map is shown in grey, contoured at 1.5σ.

The bulky side chain of Tyr205 also makes extensive hydrophobic interactions with surrounding importin-α residues whereby the phenol ring fits into a hydrophobic pocket formed by the loop connecting ARMs 1 and 2 (Leu104-Pro111). There are 30 atom to atom van der Waals contacts between the Tyr205 side chain and importin-α, which includes the side chains of Pro110, Asn146 and Ser149 as well as main chain atoms in Leu104, Ser105 and Arg106. A kinemage image [189] of Tyr205 van der Waals contact dots is shown in Figure 5.10. There are 13.5 Å^2 of van der Waals contacts made between the ^{CLIC4}Tyr205 side chain and importin-α, comparable to the contact area of basic residues at P2, P3 and P5 (see Table 5.2). These values are also comparable to the contact area of 16.4 Å^2 formed between ^{SV40}Arg130 and importin-α.

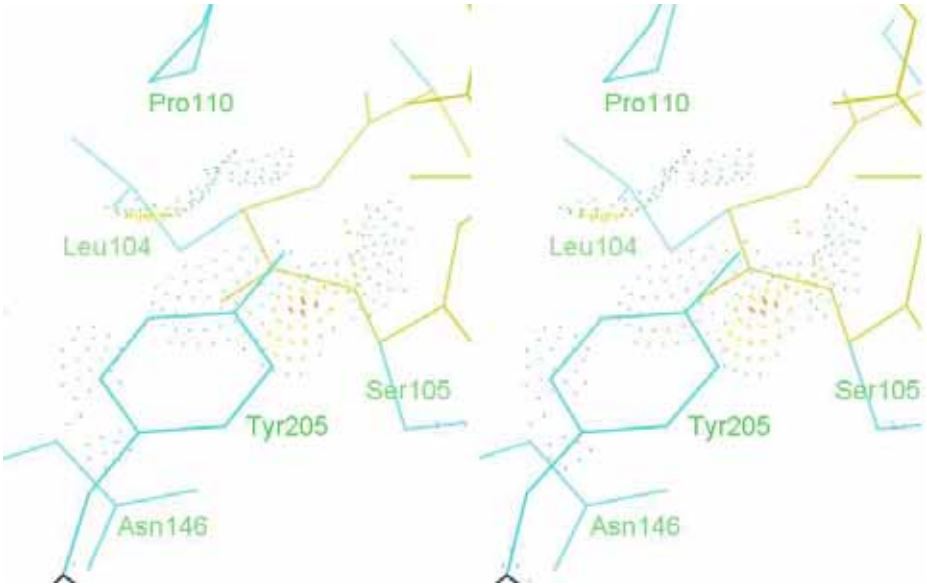

Figure 5.10: The van der Waals contact surface area between Tyr205 and importin-α.

Dark blue dots represent van der Waals radii at 0.5 Å separation, green dots at 0.25 Å and yellow and red dots are contact overlaps. No hydrogen bonds are shown. Hydrogen atoms used in the calculation of contact dots are omitted for clarity. Image created using kinemage [189] and prepared by MolProbity [159].

5.3.3.4 Normalised B-factor analysis

In order to analyse changes in the flexibility of binding site residues in importin-α due to the presence of a bound CLIC4 NLS peptide, the relative B-factor score, B_z^{-apo}, was calculated. The B_z^{-apo} score was determined by comparing normalised B-factors from the importin-α:CLIC4 NLS structure and those from *apo* importin-α, as described in Section 4.2.7.2. This method ensures that any significant differences in residue flexibility will be due to local changes in mobility rather than large global effects in the crystal. The importin-α C-terminus (residues 430-496) contains high B_z^{-apo} scores which are considered outliers (Z > 4) and thus excluded from analysis. These outliers cover a similar residue range in both the importin-α:SV40 NLS structure and importin-α:S21 structure. The B_z^{-apo} scores have a zero mean and standard deviation of 1, and are plotted in Figure 5.11 as an average over atoms in each residue ($B_z^{-apo\ res}$). The graph shows there are three main negative B_z^{-apo} peaks that correspond to residues in the ARM1 and ARM2 domains as well as Trp184 in ARM3.

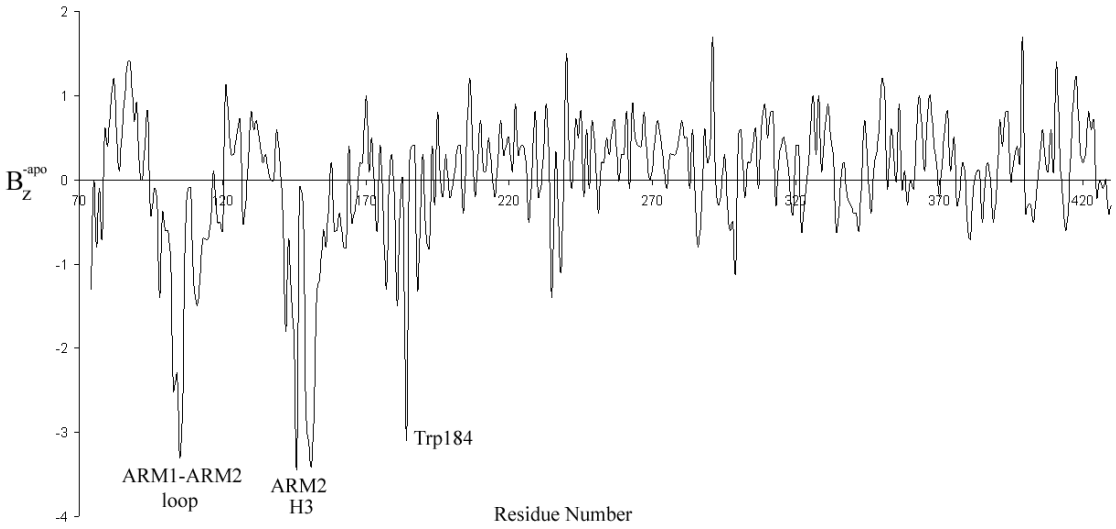

Figure 5.11: A plot of normalised B-factor scores for the importin-α:CLIC4 NLS complex.

The B_z^{-apo} value plotted here has been averaged over all atoms in each residue. Three peaks have been labelled corresponding to the ARM1-ARM2 loop, the ARM2:H3 helix and Trp184 in ARM3:H3. Scores for residues 72-430 are shown.

The B_z^{-apo} score was colour mapped onto a molecular surface representation of the importin-α molecule, as shown in Figure 5.12. The scores were averaged over main chain ($B_z^{-apo\ mc}$) atoms and side chain ($B_z^{-apo\ sc}$) atoms to illustrate key regions in importin-α. Outliers are coloured grey while other residues are coloured in a blue-white-magenta spectrum from -3 to +3 normalisation units or standard deviation units. The scores represent changes to residue flexibility that should only arise due to the binding of the CLIC4 NLS. The figure shows that the major binding site corresponds to a cluster of negative B_z^{-apo} scores, whereas the minor binding site has no significant scores.

The reduced flexibility of residue side chains in the major binding site is similar to that observed in our importin-α:SV40 NLS structure, particularly over the tryptophan and asparagine arrays. This includes residues that directly interact with the CLIC4 NLS peptide backbone such as Asn146 (-4.8^{sc}), Asn188 (-2.4^{sc}), Asn235 (-2.0^{sc}), Trp142 (-2.2^{sc}) and Trp184 (-3.9^{sc}). Also, Ser149 (-3.3^{sc}) is notably constrained by the CLIC4 NLS in a single conformation compared with dual rotamer conformations in the *apo* structure.

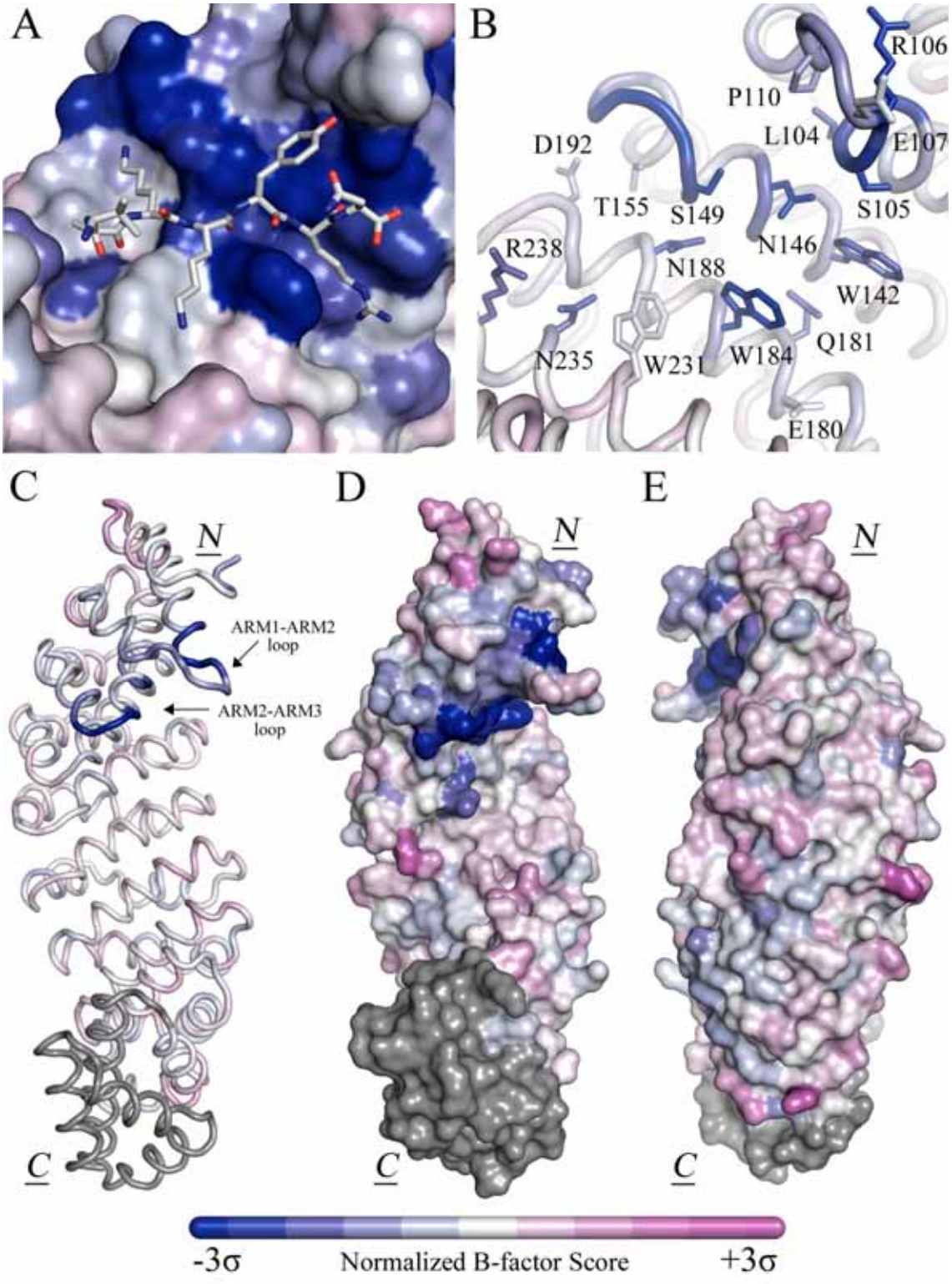

Figure 5.12: The importin-α:CLIC4 NLS peptide complex coloured by a normalised B-factor score, B_z^{-apo}.

The blue-white-magenta colour spectrum is shown in the figure, from -3σ to +3σ. Outliers near the C-terminus are shown in grey.

A) The SV40 large T-antigen NLS bound to importin-α in the major binding site. **B)** Important residues are shown in the major binding site. **C)** Cartoon tube representation of the importin-α backbone. **D)** The importin-α inner concave surface. **E)** The importin-α outer convex surface.

The presence of the Lys203 side chain at P2 has marginal effects on hydrogen bonding partners Thr155 (-0.6^{res}) or Asp192 (-0.8^{res}), but a significant effect on the third hydrogen bonding partner, the carbonyl group of Gly150 (-3.1^{res}). This can be explained by the presence of a water molecule in the *apo* structure which is predicted to make hydrogen bonds with Thr155 and Asp192 but not Gly150. We also note that the lysine N$^{\zeta}$:Gly150 bond has the added effect of reducing flexibility in the connecting loop between ARMs 1 and 2, (Ser149-Ser152) consistent with the effect observed in the importin-α:SV40 NLS complex.

At the P4 binding site we see that ^{CLIC4}Tyr205 has a very similar effect on residues in the ARM1:ARM2 connecting loop as was seen for ^{SV40}Arg130. The hydrogen bond formed between ^{CLIC4}Tyr205 and the carbonyl oxygen of $^{imp\text{-}\alpha}$Arg106 (-2.8^{res}) significantly decreases the B_z^{-apo} score for the residue. This is also seen for the other ^{CLIC4}Tyr205 shared hydrogen bonding partner, Leu104 (-2.3^{res}). Van der Waals contacts involving ^{CLIC4}Tyr205 have a stabilising effect on Pro110 (-1.2^{res}) and Ser105 (-3.3^{res}). Interestingly residue $^{imp\text{-}\alpha}$Glu107 which forms a hydrogen bond to ^{SV40}Arg130 with a B_z^{-apo} score of -3.4^{res}, has a less significant negative score in the importin-α:CLIC4 NLS structure (-0.6^{res}) where no hydrogen bonding occurs. The CLIC4 NLS, consistent with the SV40 NLS, tightens residue mobility across the ARM2-ARM3 connecting loop from Leu103 to Ile112.

The ^{CLIC4}Arg206 residue at P5 forms a hydrogen bond with $^{imp\text{-}\alpha}$Gln181 which has a corresponding decrease in its B_z^{-apo} score of -1.9^{sc}. We note that the B-factor analysis does not reflect significant changes in residue flexibility due to long range electrostatic interactions between CLIC4 NLS side chains at P3 (Lys204) or P5 (Arg206) and acidic residues in importin-α (Glu180, Glu266, Asp270).

5.4 Discussion

5.4.1 The importin-α:CLIC4 NLS peptide complex

There have been a number of key recent findings in regards to the physiological role of CLIC4 where it has been implicated in disease processes including angiogenesis [129, 131, 132] and the regulation of tumour growth [110, 204]. It is also well established that CLIC4 forms redox sensitive ion channels, although they are poorly selective [86, 124]. A further line of study has been the investigation of the mechanism whereby CLIC4 translocates to the nucleus in response to multiple types of cellular stress inducers. At sufficient levels, nuclear CLIC4 is found to be associated with apoptosis and the translocation event coincides with keratinocyte differentiation *in vitro* (reviewed in Suh *et al*. 2007, [128]).

Looking further into the cellular mechanisms behind translocation, it was revealed that nuclear CLIC4 interacts with two major nuclear import receptors: importin-α and Nuclear Transport Factor-2 (NTF-2) [123]. These import receptors operate within two different nuclear transport pathways, apparently independent of each other. In order to elucidate the nuclear import pathway in which CLIC4 utilises, mutations were made to the putative CLIC4 NLS site that resulted in the abolition of nuclear transfer [123]. Up until now, this was the strongest evidence suggesting a direct importin-α:CLIC4 interaction.

A blue native-PAGE binding assay was unsuccessful in obtaining clear evidence for a full length CLIC4:importin-α (70-529) complex. This experiment was performed at a physiologically relevant pH but without any structural modifications to CLIC4 that may be necessary to facilitate the binding mechanism. We also sought to establish if a cluster of basic residues (^{60}KRK^{62}) in CLIC4, corresponding to a similar motif in CLIC1 which was previously identified as a potential NLS [90], may also bind importin-α. By solving the crystal structure of *apo* importin-α (70-529) soaked with a CLIC4 peptide encompassing these residues (CLIC4$^{58-65}$), we demonstrate that there is no binding interaction consistent with other importin-α:NLS complexes. Noting that although the soaking method of crystallization is not preferred due to the associated

degradation of diffraction quality, it was the methods also used previously in the first crystal structure of the yeast importin-α:SV40 NLS complex [30].

In this chapter, we have presented the 2.0 Å X-ray crystal structure of the CLIC4 NLS peptide (^{198}VKVVAKKYRN^{207}) bound to importin-α (70-529). The monopartite NLS peptide binds in the major binding groove of importin-α in an extended conformation consistent with previously solved structures [39]. In the case of CLIC4, this extended NLS conformation differs greatly from its helical conformation as seen in the soluble CLIC4 structure (refer to Section 5.4.4, Figure 5.14) [86]. This important feature means that the putative CLIC4 NLS needs to undergo a structural transition if it is to be a biologically active NLS. It also suggests there is tighter control over CLIC4 translocation to the nucleus in comparison to other nuclear bound proteins that do not require their NLS to undertake a structural re-arrangement.

Electron density for the CLIC4 NLS peptide is unambiguous over the core binding site residues (^{203}KKYR^{206}, P2 - P5), while there is less well defined density at the N-terminus (^{201}VA) and C-terminus (Asn207). The first 3 residues (^{198}VKV) are not included in the peptide model since they are disordered in the crystal. With no evidence that Lys199 interacts with importin-α from the structure, the CLIC4 NLS motif should be refined to include fewer residues. The literature currently indicates the CLIC4 NLS motif corresponds to ^{199}KVVAKKYR^{206} [123], but we hypothesise the motif should be abbreviated to ^{202}AKKYRN^{207}, in line with our structure that indicates these are the residues recognised by importin-α. Although amino acids N-terminally flanking the NLS have previously been shown to be important for binding importin-α [67], it is difficult to reconcile this with our importin-α:CLIC4 NLS peptide complex.

The $F_o^{clic4} - F_o^{apo}$ data-data difference Fourier was calculated to ensure that electron density in the importin-α major binding site is not confused with residual density from the *apo* importin-α structure. The results of this difference Fourier show that all peptide side chains in P1 - P6 correspond to $F_o^{clic4} - F_o^{apo}$ density, supporting the assignation of CLIC4 NLS residues in these binding slots. This analysis has been performed on the importin-α:SV40 NLS structure with similar results and is therefore a recommended method of structural examination in other importin-α:NLS complexes.

In the importin-α:CLIC4 NLS peptide complex, the majority of key backbone and side chain interactions are conserved when compared to other importin-α bound peptides. In fact P1, P3 and P5 main chain hydrogen bonding to the invariant importin-α tryptophan and asparagine arrays are identical to previously solved structures such as the importin-α:SV40 NLS peptide complex (Section 4.4.2). Lysine side chain interactions compared between the CLIC4 NLS and SV40 NLS (PKKKRKV), are also identical in binding slots P2 and P3. This includes the key electrostatic interaction at P2 involving ^{CLIC4}Lys203 and $^{imp\text{-}\alpha}$Asp192. In classical NLS sequences, the lysine at P2 has been shown to provide the most substantial energetic contribution to peptide binding [62] and is critical for nuclear import which can be eliminated through mutagenesis of this single amino acid [60]. Although work presented by Suh *et al*. [123] obliterated the putative CLIC4 NLS motif (^{199}KVVAKKYR^{206} to ^{199}TVVAITYG^{206}), rendering CLIC4 unable to translocate to the nucleus, a single point mutation at ^{CLIC4}Lys203 should be sufficient to prevent nuclear translocation.

A prominent difference between the CLIC4 NLS and classical NLS sequences is the presence of a bulky aromatic residue within the core NLS motif. The Tyr205 CLIC4 NLS residue fits in the importin-α P4 binding slot and its presence is confirmed by strong, unambiguous density in the $F_o^{clic4} - F_o^{apo}$ difference map. Nearby weak electron density also suggests there is a cation - π interaction with the phenyl ring of Tyr205. The likely cation is the monovalent Na^+ ion which is abundant in the crystallization buffer, although its occupancy in the site is low (< 50%). Metal cation - π bonds are not commonly investigated in protein structures although they are known to form strong and stable interactions with all aromatic amino acids [205]. Due to the weak nature of the Tyr205 cation - π bond, it seems unlikely it will have a significant effect on the CLIC4 NLS peptide binding to importin-α.

The tyrosine is found to interact with importin-α through both hydrophobic and hydrogen bonding interactions. Although it is an unusual NLS residue containing a bulky phenol side group, it neatly fits at its location with minimal van der Waals clashes. Most of the hydrophobic contacts with importin-α residues lining the P4

pocket are equivalent to those made by the aliphatic portion of the SV40 NLS residue, ^{SV40}Arg130. The side chain hydroxyl group reaches the importin-α main chain near the C-terminus of the ARM1:H3 helix. The hydrogen bonds shared between the tyrosine hydroxyl group and carbonyl groups from $^{imp\text{-}\alpha}$Leu104 and $^{imp\text{-}\alpha}$Arg106 are reproduced in the importin-α:SV40 NLS P4 binding slot involving the N^{ζ} atom of ^{SV40}Arg130. Because of the proximity of the ^{CLIC4}Tyr205 hydroxyl group to importin-α, CLIC4 NLS peptide binding should be prevented by tyrosine phosphorylation. Therefore the phosphorylation state of CLIC4 could foreseeably act as a switch between an 'active' and 'inactive' CLIC4 NLS. The effects of phosphorylation on the nuclear import of the SV40 large T-antigen have previously been investigated [206, 207]. Although it was shown that the phosphorylation of serine residues N-terminal to the SV40 NLS significantly enhances nuclear import, structural studies revealed that these serine residues do not directly interact with the importin-α binding site [67].

The importin-α P4 binding slot has been shown to tolerate a wide range of residues including: arginine from the SV40 NLS (Section 4.4.2), lysine in the nucleoplasmin NLS (PDB:1EE5, [31]) and androgen receptor NLS (PDB:3BTR, [53]), a valine in the c-*myc* NLS (PDB:1EE4, [31]), a leucine in the retinoblastoma NLS or a serine in the N1N2 NLS (1PJM and 1PJN, [50]) and an isoleucine in the influenza A PB2 subunit NLS (2JDQ, [68]). The contribution to NLS binding energy of the residue at P4 is similar whether it is occupied by a negatively charged arginine side chain or a hydrophobic valine side chain [62]. This is despite the extra hydrogen bonding formed by arginine at P4 and a possible helix dipole interaction with H3 helices of ARMs 1 and 2. Although the hydrogen bonds formed by Tyr205 may suggest that a tyrosine at P4 is more favourable than a phenylalanine or other bulky residue, it is likely to have a similar energetic contribution to peptide binding as valine or arginine.

The CLIC4 NLS contains an arginine residue (Arg206) in binding slot P5 which adopts a common and extended rotamer conformation. The P5 slot is defined by importin-α residues Trp142 and Trp184, which both adopt a rare rotamer conformation that facilitates the formation of this hydrophobic pocket. The site is normally occupied by a NLS lysine residue in importin-α:NLS structures (with a dual conformation in our importin-α:SV40 NLS structure), where the aliphatic portion of the side chain fits into

the hydrophobic alcove and allows the charged head group access to the solvent on the other side. The extended side chain of ^{CLIC4}Arg206 allows for the same favourable interactions as a lysine residue, including a hydrogen bond to $^{imp\text{-}\alpha}$Gln181 and the formation of a network of solvent interactions. The positively charged guanidinium group of ^{CLIC4}Arg206 is also compensated by the nearby acidic residue $^{imp\text{-}\alpha}$Glu180, and we note the nearest interaction distance between them ($N^{\zeta 1}$ - $O^{\varepsilon 1}$, 4.9 Å) is significantly closer than the equivalent interaction involving the P5 lysine in our importin-α:SV40 NLS structure (N^{ζ} - $O^{\varepsilon 1}$, 6.1 Å). While an arginine at P5 is somewhat rare, two previously solved importin-α:NLS structures also consist of a P5 arginine, modelled with the same conformation as in our structure: the importin-α:retinoblastoma NLS structure [50] and the importin-α:influenza A PB2 subunit structure [68].

A normalised B-factor analysis has demonstrated the effect on the flexibility of importin-α binding site residues when the CLIC4 NLS peptide binds. The calculation of the B_z^{-apo} score is a novel approach to investigating features of an importin-α:NLS complex and also serves as an independent method of confirming the presence of the CLIC4 NLS peptide in our importin-α:CLIC4 NLS complex. We see that the constraints placed on residues which participate in hydrogen bonds to the peptide are reflected in significantly negative B_z^{-apo} scores. This is most pronounced in the conserved tryptophan and asparagine arrays where the flexibility of side chains would be expected to decrease with the presence of a bound peptide.

There is a more general decrease in flexibility through main chain and side chain atoms in loop regions connecting ARMs 1 and 2, as well as ARMs 2 and 3. The ARM1:ARM2 connecting loop is stabilised by the side chain of ^{CLIC4}Tyr205 at binding slot P4, similar to the effect observed for ^{SV40}Arg130. The ARM2:ARM3 connecting loop is stabilised by ^{CLIC4}Lys203 at P2, which forms a hydrogen bond with $^{imp\text{-}\alpha}$Gly150 that is not present in the *apo* structure.

5.4.2 Conservation of the NLS sequence in CLICs

The CLIC4 NLS sequence ^{199}KVVAKKYR^{206} is highly conserved in vertebrate species and across CLICs 1-6, except CLIC3 (Figure 5.13). The NLS can be broken up into separate structural parts according to our importin-α:CLIC4 NLS structure, with an N-terminal sequence that does not participate in binding and a C-terminal sequence that directly binds. The C-terminal KKYR motif that occupies the core binding region P2 - P5, appears to control importin-α recognition of the NLS and is conserved in all human CLICs except CLIC3. The significance of the motif is apparent when looking at the cluster of contacts made between the CLIC4 NLS and importin-α in which the KKYR region accounts for 92% of hydrogen bonds and 92% of van der Waals contacts.

Figure 5.13: Multiple sequence alignment of the CLIC4 NLS motif in *human* CLIC proteins.

A sequence alignment of human CLICs 1-6 is shown. Conserved residues are coloured red, non-conserved residues are black and perfect conservation is highlighted with red fill in. The sequence of CLIC3 does not participate in conservation colouring. Binding slots P1 to P6 are shown in an alignment that corresponds to our solved importin-α:CLIC4 NLS peptide complex.

The N-terminal KVVA motif is mostly conserved in CLICs, although there are hydrophobic amino acid substitutions that should not significantly alter the NLS binding interaction. The first residue in the motif, Lys199, does not make any observed interactions in our importin-α:CLIC4 NLS structure and is conspicuously absent in CLIC1, replaced by a glutamine. Furthermore, CLIC1 contains a cysteine residue (Cys191) within the NLS which will occupy the major binding site P1 position if a CLIC1 NLS were to bind importin-α with a similar alignment to the CLIC4 NLS. The presence of the cysteine residue at this position suggests a tantalising possibility that

redox activity may regulate the interaction of the putative CLIC1 NLS with importin-α. In binding slot P6, the CLIC4 NLS asparagine residue that makes no interactions with importin-α is only present in CLIC5. CLIC2 and CLIC6 both contain an aspartic acid at P6, which is an uncommon NLS residue compared with the positively charged nature of an NLS sequence. The relatively short KKYR sequence that dictates the most important CLIC4 NLS interactions with importin-α could indicate that residue variability in the local flanking regions is acceptable so long as the core region satisfies key binding requirements.

Similarities in the putative NLS sequence therefore suggests that all CLICs, apart from CLIC3, may interact with importin-α by utilising a binding mode similar to that observed in our importin-α:CLIC4 NLS peptide complex, providing the NLS region is exposed and extended. This could suggest there is a shared role in the response to cell stress by members of the CLIC family which can participate in the importin-α nuclear import pathway. To date, reported experiments have focused on CLIC4 and little data exists that links other members of the CLIC family to the cells response to stress [123]. However, we do note that CLIC1 has been shown to localise in the nucleus with an unexplained import pathway [90]. Thus, the possibility of a functioning CLIC1 NLS operating in much the same manner as the CLIC4 NLS needs to be explored, particularly in conditions of a controlled redox environment.

5.4.3 Import pathways

The importin-α:CLIC4 NLS complex presented in this chapter supports a model in which the full length CLIC4 protein binds to importin-α through an interaction in the major binding site. This could suggest that CLIC4 can enter the nucleus via an importin-α mediated nuclear import pathway. The recognition of the CLIC4 NLS by the importin-α import receptor is strictly dependent on the ability of the NLS to undergo a change in conformation that would expose a linear NLS ready for docking to importin-α.

Experiments that show CLIC4 translocates to the nucleus via various import machinery components were not clear on the exact mechanism by which translocation

occurs [123]. Although mutagenesis of the putative CLIC4 NLS implicated an interaction with importin-α, immunoprecipitation experiments showed that CLIC4 was associating with importin-α as well as NTF-2 and Ran [123]. These other proteins are part of the nuclear import machinery and work together to maintain a concentration gradient of Ran across the nuclear membrane [9]. The NTF-2 pathway is thought to be contained in an independent transport cycle, shuttling Ran between the cytoplasm and the nucleus but not interacting with cargo complexes [74].

One possible explanation for the association between NTF-2 and CLIC4 is that there could be an interaction during intermediate processes in the nuclear import cycle. This may occur during CLIC4 cargo dissociation from the importin-α:β complex or in cargo export where a number of proteins facilitate the release of imported cargo [39]. Such an intermediate complex may also explain the association between CLIC4 and Ran, which is likely facilitated through an indirect interaction by a common binding partner. An example of this is the importin-α:Cse1p:RanGTP export complex [17], which could provide an indirect association of importin-α cargo with Ran as it is being released in the nucleus. However, while Ran is often involved in transient import and export processes, NTF-2 has not been implicated with such a role.

More recently, the direct involvement of NTF-2 and Ran in cargo transport has been shown. This occurs in the case of the filamentous actin capping protein CapG [24]. Here it is seen that CapG is transported to the nucleus via direct association with NTF-2 and Ran, and furthermore that it can bind directly to the NTF-2:Ran complex. It was further noted that CapG is not transported to the nucleus via an importin-α:β mechanism due to the inability of the SV40 NLS [208] and the GTPγS classical nuclear import inhibitor [24] to affect nuclear accumulation of CapG.

This finding may have implications in the study of CLIC4 nuclear transport since it provides support for a multiple nuclear import pathway hypothesis. If CLIC4 utilises the NTF-2 transporter in conjunction with the importin-α:β transport pathway, this would be a unique property among nuclear proteins deserving of further investigation. Ascertaining the correct CLIC4 import system could reveal useful new information

about the variability of cargo recognition in nuclear import pathways and how the transfer of cargo proteins is regulated.

5.4.4 Triggers for NLS exposure

The CLIC4 NLS is unusual not only because the sequence does not resemble a classical NLS but also because the NLS is contained within a structured region of the folded, full length CLIC4 protein. Typically NLS sequences are located in solvent exposed regions of unstructured domains such as at flexible termini. This promotes receptor recognition of cargo in which the NLS binds to importin-α in an elongated fashion that accommodates interaction with the tryptophan and asparagine arrays that line the major binding site.

The classical NLS sequence of the SV40 large T-antigen (^{126}PKKKRKV^{132}) is not located at the protein terminus but is inserted between the N-terminal J-domain (PDB:1GH6, includes residues 7-117) and the central original-binding domain or o.b.d. (PDB:2FUF, includes residues 131-260). The crystal structures of these SV40 domains have disordered termini suggesting the linking region containing the NLS is flexible and solvent exposed. The NLS motif is relatively close to the o.b.d. which suggests a long loop is not needed to ensure accessibility to the importin-α binding site. Since the binding motif is close to the folded o.b.d., it is quite probable that SV40 experiences a steric hindrance due to the proximity of importin-α in a full length importin-α:SV40 complex.

The CLIC4 NLS motif, as seen in the soluble CLIC4 crystal structure, is unusual for an NLS as it is neither unfolded nor situated toward domain termini (PDB:2AHE). It is in fact internally located in a structurally stable helical conformation. Since the CLIC4 NLS binds to importin-α in an extended conformation this means there must be an event that triggers structural changes in CLIC4 to expose the NLS region in preparation for importin-α recognition. Without such a conformational change, it is difficult to envisage CLIC4 being imported via the importin-α:β pathway.

The NLS sequence in CLIC4 is situated at the C-terminus of helix 6 and includes residues that lead into a connecting loop between helix 6 and helix 7. The key KKYR binding motif in the NLS is positioned so that Arg206 terminates the helix and the two lysine residues are solvent exposed (Figure 5.14). In addition to backbone hydrogen bonding in helix 6, the region contains a number of side chain hydrogen bonding interactions that includes a salt bridge between Lys203 and Asp209. We also find Tyr205 and Arg206 side chains form hydrogen bonds with residues in helix 9 and the C-terminal extended tail. If an unfolding event exposes the NLS region of CLIC4 it could foreseeably initiate through an unravelling of the C-terminal tail.

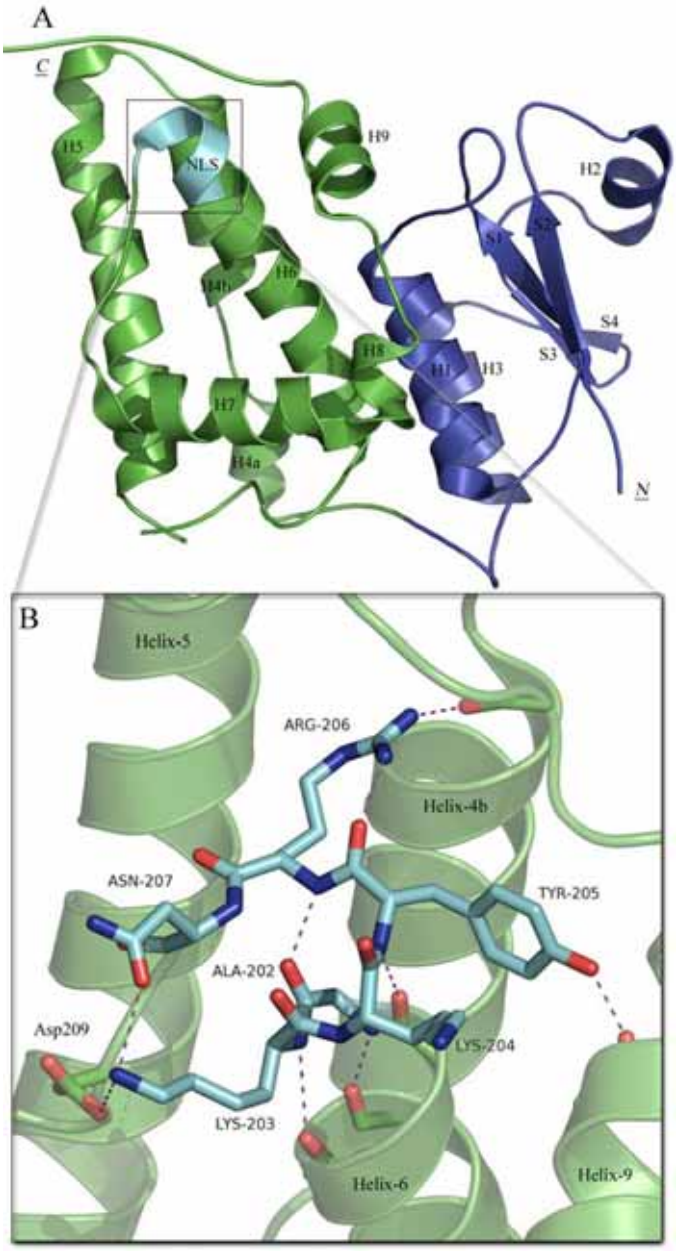

Figure 5.14: The putative NLS motif in the soluble CLIC4 crystal structure.

A) The full length human CLIC4 structure is shown in cartoon representation (PDB:2AHE). The N-terminal domain is coloured blue, the C-terminal domain is green and NLS residues are cyan.

B) The folded NLS is shown with carbons in cyan, oxygens in red and nitrogens in blue. Hydrogen bonds are represented by magenta dashes.

It is clear that a significant structural modification needs to occur that will expose the CLIC4 NLS in an extended conformation prior to binding importin-α. What remains unclear is the transitional mode that will accomplish this. One possible post-translational modification is the S-nitrosylation of a cysteine residue that is in close proximity to the CLIC4 NLS site, Cys234, which resides between helix 8 and helix 9.

It has been shown that Cys234 is S-nitrosylated when exposed to a nitric oxide agent, S-nitrosocysteine [209], a modification which could result in the disruption of structural stability in the C-terminal domain including helix 6. The nitrosylation of Cys234 occurs despite the thiol side chain being buried in the inter-domain interface. The function of S-nitrosylation as a trigger for conformational change has been previously observed in blackfin tuna myoglobin [210]. The *in vitro* S-nitrosylation of a cysteine residue causes a structural transition whereby a helix and loop region shifts to accommodate the S-nitrosocysteine and avoid steric clashes.

Using the 1.65 Å resolution structure of human thioredoxin (PDB:2HXK) which contains a buried S-nitrosocysteine (Cys62), the coordinates of the SNO group were transferred to CLIC4 Cys234 in PDB:2AHE (Figure 5.15). Both cysteines adopt the most common rotamer assisting the residue transfer. The presence of an SNO group at Cys234 fits relatively well in the central hydrophobic region of CLIC4. Without predicting other changes in the structure, MolProbity [159] calculates a strong steric clash between the S-nitrosylated Cys234 and His196 suggesting there will be a structural rearrangement to alleviate the clash. With no clearly favourable alternative rotamer, it is likely in such an event that the helix containing His196 (helix 6) will shift at least 1 Å to remove the more serious steric overlaps. Since helix 6 forms an inter-domain interface with helix 1, S-nitrosylation could have a more global effect on the structural stability of the CLIC module. In support of this process as a means to allow CLIC4 to interact with importin-α, there is a correlation between nitric oxide donors and the rate of nuclear translocation of CLIC4 [128]. Furthermore, trophic agents known to induce nitric oxide synthase 2 encourage CLIC4 nuclear translocation and the S-nitrosylation of Cys234.

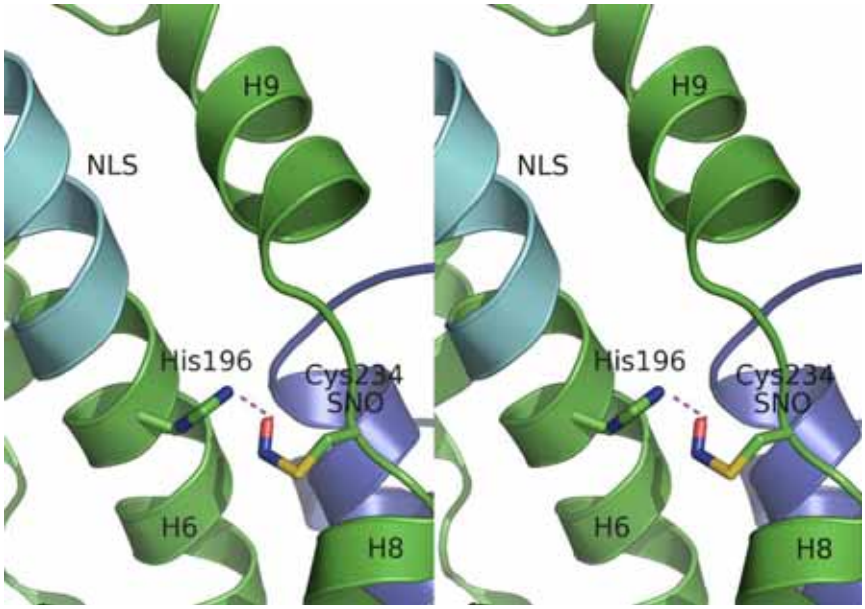

Figure 5.15: A stereo image of CLIC4 with a hypothetical S-nitrosylated Cys234.

CLIC4 is shown in cartoon representation with the C-terminal domain in green and the N-terminal domain in blue. The NLS motif (KVVAKKYR) is shown in cyan. A steric overlap Cys234 (SNO) and His196 is represented by a dashed line.

Another possible post-translational modification involves another cysteine residue in CLIC4: the redox sensitive Cys35. This is a highly conserved residue in CLICs, which may be important for a structural transition leading to a membrane bound form of CLIC within the cell [86]. Experiments have shown that Cys35 may be susceptible to oxidation from H_2O_2 [86], nitric oxide and oxidised glutathione [211]. It is yet to be seen how post-translational modification of Cys35 affects the structural conformation of CLIC4. A relevant example of signalling via reactive oxygen species is shown by the importin-α dependent nuclear import of thioredoxin-1, which is regulated by H_2O_2 and nitric oxide [212].

The requirement for a major rearrangement of the CLIC module to expose the NLS may well occur through a cumulation of post-translational modifications in addition to changes to Cys35 or Cys234. In CLIC1, the cysteine corresponding to Cys35 in CLIC4 (Cys24) forms a disulfide bond with Cys59, trapping a structural transition that allows the formation of a non-covalent dimer [102]. Although this has not been shown to occur in other CLIC proteins, it suggests a degree of plasticity in the CLIC module that could accommodate such necessary triggered structural changes.

The prevalence of ‘metamorphic’ proteins, such as CLIC, capable of dramatic conformational rearrangements, is becoming increasingly recognised [213].

The possibility of trigger events to expose NLS sequences offers an interesting criterion for cargo selection by nuclear import receptors. Typically NLS sequences are identifiable due to their high content of basic residues which allow them to bind tightly to an acidic binding site in importin-α. However it is clear that the pre-requisite for NLS binding, based on sequence alone, is very limited since a core region of just four residues are necessary for importin-α peptide recognition. Furthermore, there is a certain degree of tolerance for non-basic residues within the importin-α binding region, particularly at P4 where we find the CLIC4 NLS tyrosine residue. In order to select specific cargo proteins destined for the nucleus, it is likely importin-α screens additional criteria such as the flexibility of the NLS and its solvent exposure within a folded protein. This form of regulation is governed more so by the cargo protein itself rather than importin-α and as such is a specific control mechanism that has not been thoroughly considered among the many levels of import regulation [8].

For a protein such as CLIC4 where nuclear translocation occurs in response to cellular events, it is reasonable that a higher level of regulation is needed compared to proteins that regularly shuttle across the nuclear membrane. The folded state of the CLIC4 NLS may provide the necessary barrier to binding importin-α, so that an environmental trigger is required before the NLS unfolds into an extended and exposed sequence that the import receptor recognises.

5.5 Conclusion

In this chapter we have solved the importin-α:CLIC4 NLS crystal structure at 2.0 Å, demonstrating that a CLIC4 NLS bearing peptide is capable of binding the nuclear import receptor, importin-α. The presence of the peptide was confirmed by the $F_o^{clic4nls} - F_o^{apo}$ data-data difference Fourier which contains strong density for side chain features unique to the CLIC4 NLS. The interaction between the peptide and importin-α occurs in the major binding site and is comparable to the interaction involving classical NLS sequences, although the presence of an aromatic residue, Tyr205, occupying a key binding slot at P4 is observed for the first time in our importin-α:CLIC4 NLS complex. Surprisingly, the bulky side chain of the tyrosine residue fits into the P4 slot without significant steric clashes and forms hydrogen bonds to the importin-α main chain through the hydroxyl group. We demonstrated there is a reduction in importin-α residue flexibility near the peptide binding site by analysing the normalised B-factor score, B_z^{-apo}. Although such an approach has not been used before in examining importin-α:NLS complexes, it proves to be a useful tool to analyse the effects of NLS binding.

The implications of the importin-α:CLIC4 NLS peptide complex span two separate fields: nucleocytoplasmic transport and CLIC4 function. The high resolution importin-α:CLIC4 NLS structure, together with nuclear import assays and mutagenesis results [123], provide a compelling case that CLIC4 enters the nucleus via the classical nuclear import pathway. Further work may investigate the ability of other CLIC family members to interact with importin-α.

Regarding nuclear import, the CLIC4 NLS demonstrates that only short basic sequences are necessary to bind importin-α, which has a degree of tolerance for non-basic residues both outside and inside the core binding slots P2 - P5. While it has been considered that the importin-α binding slot P4 can accommodate a number of non-basic residues, current sequence databases such as PredictNLS [214] and NLStradamus [215], ignore the KKYR motif as a potential NLS. This suggests there could be a number of NLS sequences that have gone undetected and that the criteria

for predicting an NLS sequence needs to be expanded to account for this residue variability.

Studying the trigger mechanism that unfolds the CLIC4 NLS is a worthwhile avenue for further experimentation. The strongest evidence for induced structural change is in the S-nitrosylation of Cys234 and therefore obtaining a structure of CLIC4 with the S-nitrosylation modification could be useful for determining if this is enough to expose the NLS. If the structural changes are small, it may be that the disruption to local stability around the NLS is sufficient to trigger further unfolding, and so experiments to measure the structural stability of CLIC4 with an S-nitrosylated Cys234 could also be useful. The possibility of a conformational switch that triggers NLS recognition by import receptors is a fascinating subject, and one that must be explored to fully elucidate the mechanisms that govern nuclear import pathways.

Chapter 6

Final conclusions and future directions

This thesis has presented structural investigations into members of the CLIC family of proteins and the nuclear import receptor, importin-α. It focuses on two main avenues of work: the first relating to structural insights into the invertebrate CLIC from *Drosophila melanogaster* (*Dm*CLIC) and the other the structural characterisation of the interaction between *Homo sapiens* CLIC4 and mouse importin-α. Performed in parallel with these studies was the investigation of the importin-α recognition of an NLS-like sequence in the *E. coli* 30S ribosomal subunit S21.

In Chapter 2 the 1.80 Å structure of *Dm*CLIC was presented and comparisons were made to other known structures of vertebrate and invertebrate CLIC homologues. The structure of *Dm*CLIC corresponds to a revised and improved model that built upon previously published work [137]. Like other CLICs, *Dm*CLIC adopts the characteristic GST fold that consists of an N-terminal thioredoxin-like domain and all α-helical C-terminal domain. Some structural differences are noted in *Dm*CLIC compared to vertebrate CLICs, particularly over flexible loop regions and minor helical extensions. One of the more significant differences is the invertebrate specific C-terminal extension which adopts an elongated conformation and interacts with the surface of the CLIC module over the C-terminal domain. An invertebrate specific metal binding site in *Dm*CLIC is another distinctive structural feature, properties also shared by the *C. elegans* CLIC homologue, EXC-4. An enhanced analysis of the site, using single wavelength anomalous dispersion data and the bond valence method, has identified the bound metal as a K^+ cation. The assignment of this metal clarifies earlier concerns over the proposed isoelectronic Ca^{2+} cation in the published *Dm*CLIC structure, where a high concentration of the metal chelating agent, EDTA, in the crystallization buffer does not appear to affect Ca^{2+} binding. These findings also apply to EXC-4, where the conserved metal binding site is also likely to be homogeneously occupied by a bound K^+ cation, rather than Ca^{2+}, as assigned in the published structure. The proximity of the metal binding site to the conserved CxxC glutaredoxin-like motif may suggest an allosteric role for K^+ in putative GSH binding.

Future experiments will be needed to investigate the possibility that other CLIC proteins can also bind K^+, particularly given the conserved conformational geometry of the metal's coordination sphere. In addition, the possibility that K^+ binding acts as a structural switch should be examined by determining structures of invertebrate CLICs in the unbound state via deprivation of potassium during crystallization. Finally, the results need to be considered in the context of electrophysiology where the absence of potassium may have a direct impact on the ability of invertebrate CLICs to form functioning ion channels.

In Chapter 3, preliminary studies to determine structures of importin-α:NLS peptide complexes are presented. Although the experimental methods that were used followed widely accepted protocols published in the literature, the interpretation of electron density for a peptide bound in the importin-α binding site was ambiguous. By mutating an NLS-like sequence in the artificial N-terminal domain of the importin-α (70-529) construct, AKFERQ to AEFEGQ, an intrasteric interaction was shown not to be responsible for the unaccountable electron density. Instead, it was discovered that importin-α (70-529) had a rogue peptide or protein occupying the binding site. These findings were critical in highlighting the necessity of an improved purification protocol required to obtain a structure of importin-α (70-529) with an empty binding site (*apo* importin-α).

In Chapter 4 the crystal structure of *apo* importin-α (70-529) is presented at a resolution of 1.77 Å, which corresponds to the highest resolution structure of any importin-α currently published. The improved purification protocol that was used to obtain *apo* protein involved ion exchange chromatography and size exclusion chromatography, additional steps that should be adopted for future experiments using this importin-α (70-529) construct. Following the determination of the *apo* structure, the importin-α:SV40 NLS peptide complex was solved at 1.85 Å, the highest resolution importin-α complex currently determined. A thorough analysis of the SV40 NLS peptide binding interaction was performed using a range of methods including the $F_o^{sv40} - F_o^{apo}$ data-data difference Fourier and a novel normalised B-factor score, B_z^{-apo}. These methods rely on information obtained from the *apo* importin-α (70-529)

structure and thus *apo* serves as a useful and necessary control in establishing the authenticity of the importin-α:NLS structures.

Also explored in Chapter 4 is an unexpected discovery involving an interaction between endogenous *E. coli* proteins and importin-α (70-529). An NLS-like motif in the 30S ribosomal subunit S21, ^{P1}RKRAKA^{P6}, was shown to bind to the importin-α (70-529) major binding site and is likely responsible for the rogue peptide electron density observed in preliminary structures. It may be that the 30S ribosomal subunit S11 also binds importin-α (70-529), possibly as an S11:S21 heterodimer which is disordered in the crystal structure. The only ordered regions of the S11:S21 complex appear to correspond to where S21 interacts with the major binding site and at Cys133, where S11 may form a disulfide bond with importin-α (70-529).

This opens up a fortuitous opportunity for further study, as there are few importin-α complexes solved with full length substrates. The use of small angle X-ray scattering (SAXS) to probe for an S11:S21 dimer bound to importin-α is a suitable technique in this instance. However, it must be noted that an importin-α:S11:S21 complex is artificial and not strictly biologically relevant. Hence the interaction may simply prove to be coincidental. However, further work should consider if the NLS-like motif in S21 functions as a signal in *E. coli*, where the importin-α recognition of the motif may resemble a binding interaction that is relevant to bacterial systems.

It is recommended that future work involving importin-α, particularly truncation constructs expressed in *E. coli* expression systems, experimentally characterise the *apo* form and compare this to the substrate bound forms. This would have, for example, highlighted inaccuracies in the published structure of the importin-α:PLSCR1 NLS complex [75], which is most likely an importin-α:S21 complex. In future crystallography studies, the determination of an unambiguous importin-α:NLS complex may include similar methods to the ones employed in this thesis, such as the data-data difference Fourier and normalised B-factor analysis. Due to the time consuming nature of crystallography work, a quick assay system to determine if the importin-α binding site is occupied will be useful. Due to the prevalence of tryptophan

residues lining the binding surface, tryptophan quenching detected by fluorescence may prove to be such an experimental technique that can determine the presence of a bound peptide.

The importin-α:CLIC4 NLS crystal structure is presented in Chapter 5 at a resolution of 2.0 Å. It is thus demonstrated that a CLIC4 NLS bearing peptide (^{198}VKVVAKKYRN^{207}) is capable of binding the nuclear import receptor, importin-α, supporting the hypothesis that CLIC4 translocates to the nucleus via a classical NLS interaction [123]. Since the CLIC4 NLS has α-helical secondary structure in the native CLIC4 structural fold, binding to importin-α will require a structural rearrangement to expose the NLS sequence in an elongated conformation. The CLIC4 NLS peptide binding mode imitates the classical mono-partite SV40 NLS interaction, where only the short basic motif, KKYR, conserved in all human CLICs apart from CLIC3, interacts with importin-α (70-529) in the core binding region over binding slots P2 - P5. The N-terminal NLS lysine (Lys199) is disordered in the crystal and therefore future experiments should consider a revised CLIC4 NLS sequence. Such a revised sequence should include residues in the P1 - P6 binding slots (AKKYRN). The structure confirms that Lys203 fits in the critical binding slot at P2 and therefore a single point mutation at this residue should be sufficient to obliterate CLIC4 NLS interaction with importin-α and hence prevent nuclear translocation. This compares with the quadruple point mutation to the CLIC4 NLS (^{199}TVVAITYG^{206}) which was shown to prevent nuclear translocation [123]. Mutation of the other core basic residues should disrupt the CLIC4 NLS interaction with importin-α. To a lesser degree, mutation of the tyrosine (Tyr205) may also impact on the NLS interaction, since the bulky side chain makes favourable interactions in the P4 binding slot.

The importin-α:CLIC4 NLS peptide complex shows for the first time that an aromatic NLS residue can occupy a core binding slot in the importin-α binding site. The presence of tyrosine in an NLS sequence should therefore not exclude it as a potential NLS. However, although it is reported that the importin-α P4 binding slot can tolerate various other non-basic residues, current generation sequence databases discard the KKYR motif as an NLS-like sequence. Therefore the criteria for predicting

an NLS needs to be further expanded to account for a tyrosine that can fit in the P4 slot.

Further work examining the binding of importin-α and CLIC4 may also benefit from investigating the interaction with other CLIC family members, considering that the core NLS sequence is highly conserved. However the most significant barrier to understanding CLIC4 translocation to the nucleus is the unknown trigger mechanism that is needed to unfold the CLIC4 NLS in preparation for docking to importin-α. Triggering and thus determining such a modified conformation of CLIC4 necessary for interaction with importin-α is integral to understanding the process. Perhaps the strongest evidence to date for a possible induced structural change is the S-nitrosylation of Cys234 and therefore structurally characterising the S-nitrosylated form of CLIC4, will be useful for assessing if this post-translational modification is adequate to allow for exposure of the NLS. The effects of S-nitrosylation on other homologous CLIC members could also be concurrently examined. Unravelling the mechanisms of a conformational switch that triggers NLS recognition by import receptors will be an immense leap forward in our understanding of CLICs, nuclear import pathways and indeed the metamorphic behaviour of proteins in general.

References

1. Fahrenkrog, B., J. Koser, and U. Aebi, *The nuclear pore complex: a jack of all trades?* Trends Biochem Sci, 2004. **29**(4): p. 175-82.
2. Rout, M.P., et al., *The yeast nuclear pore complex: composition, architecture, and transport mechanism.* J Cell Biol, 2000. **148**(4): p. 635-51.
3. Rout, M.P. and S.R. Wente, *Pores for thought: nuclear pore complex proteins.* Trends Cell Biol, 1994. **4**(10): p. 357-65.
4. Stewart, M., *Molecular mechanism of the nuclear protein import cycle.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 195-208.
5. Rout, M.P. and G. Blobel, *Isolation of the yeast nuclear pore complex.* J Cell Biol, 1993. **123**(4): p. 771-83.
6. Cronshaw, J.M., et al., *Proteomic analysis of the mammalian nuclear pore complex.* J Cell Biol, 2002. **158**(5): p. 915-27.
7. Reichelt, R., et al., *Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components.* J Cell Biol, 1990. **110**(4): p. 883-94.
8. Terry, L.J., E.B. Shows, and S.R. Wente, *Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport.* Science, 2007. **318**(5855): p. 1412-6.
9. Smith, A., A. Brownawell, and I.G. Macara, *Nuclear import of Ran is mediated by the transport factor NTF2.* Curr Biol, 1998. **8**(25): p. 1403-6.
10. Conti, E. and E. Izaurralde, *Nucleocytoplasmic transport enters the atomic age.* Curr Opin Cell Biol, 2001. **13**(3): p. 310-9.
11. Cingolani, G., et al., *Structure of importin-beta bound to the IBB domain of importin-alpha.* Nature, 1999. **399**(6733): p. 221-9.
12. Bayliss, R., et al., *GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta.* J Biol Chem, 2002. **277**(52): p. 50597-606.
13. Lee, S.J., et al., *Structural basis for nuclear import complex dissociation by RanGTP.* Nature, 2005. **435**(7042): p. 693-6.
14. Bischoff, F.R. and D. Gorlich, *RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors.* FEBS Lett, 1997. **419**(2-3): p. 249-54.
15. Matsuura, Y. and M. Stewart, *Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling.* Embo J, 2005. **24**(21): p. 3681-9.
16. Solsbacher, J., et al., *Nup2p, a yeast nucleoporin, functions in bidirectional transport of importin alpha.* Mol Cell Biol, 2000. **20**(22): p. 8468-79.
17. Matsuura, Y. and M. Stewart, *Structural basis for the assembly of a nuclear export complex.* Nature, 2004. **432**(7019): p. 872-7.
18. Solsbacher, J., et al., *Cse1p is involved in export of yeast importin alpha from the nucleus.* Mol Cell Biol, 1998. **18**(11): p. 6805-15.
19. Conti, E., C.W. Muller, and M. Stewart, *Karyopherin flexibility in nucleocytoplasmic transport.* Curr Opin Struct Biol, 2006. **16**(2): p. 237-44.
20. Stewart, M., H.M. Kent, and A.J. McCoy, *Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran.* J Mol Biol, 1998. **277**(3): p. 635-46.
21. Vetter, I.R., et al., *Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport.* Nature, 1999. **398**(6722): p. 39-46.
22. Forwood, J.K., et al., *Kap95p binding induces the switch loops of RanGDP to adopt the GTP-bound conformation: implications for nuclear import complex assembly dynamics.* J Mol Biol, 2008. **383**(4): p. 772-82.
23. Bayliss, R., et al., *Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats.* Embo J, 2002. **21**(12): p. 2843-53.
24. Van Impe, K., et al., *A new role for nuclear transport factor 2 and Ran: nuclear import of CapG.* Traffic, 2008. **9**(5): p. 695-707.

25. Strom, A.C. and K. Weis, *Importin-beta-like nuclear transport receptors.* Genome Biol, 2001. **2**(6): p. REVIEWS3008.
26. Bayliss, R., T. Littlewood, and M. Stewart, *Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking.* Cell, 2000. **102**(1): p. 99-108.
27. Bednenko, J., G. Cingolani, and L. Gerace, *Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport.* J Cell Biol, 2003. **162**(3): p. 391-401.
28. Lee, S.J., et al., *The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor.* Science, 2003. **302**(5650): p. 1571-5.
29. Cingolani, G., et al., *Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta.* Mol Cell, 2002. **10**(6): p. 1345-53.
30. Conti, E., et al., *Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha.* Cell, 1998. **94**(2): p. 193-204.
31. Conti, E. and J. Kuriyan, *Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha.* Structure, 2000. **8**(3): p. 329-38.
32. Andrade, M.A., et al., *Comparison of ARM and HEAT protein repeats.* J Mol Biol, 2001. **309**(1): p. 1-18.
33. Mitrousis, G., et al., *Molecular basis for the recognition of snurportin 1 by importin beta.* J Biol Chem, 2008. **283**(12): p. 7877-84.
34. Liu, S.M. and M. Stewart, *Structural basis for the high-affinity binding of nucleoporin Nup1p to the Saccharomyces cerevisiae importin-beta homologue, Kap95p.* J Mol Biol, 2005. **349**(3): p. 515-25.
35. Vetter, I.R., et al., *Structural view of the Ran-Importin beta interaction at 2.3 A resolution.* Cell, 1999. **97**(5): p. 635-46.
36. Fukuhara, N., et al., *Conformational variability of nucleo-cytoplasmic transport factors.* J Biol Chem, 2004. **279**(3): p. 2176-81.
37. Kobe, B., *Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha.* Nat Struct Biol, 1999. **6**(4): p. 388-97.
38. Goldfarb, D.S., et al., *Importin alpha: a multipurpose nuclear-transport receptor.* Trends Cell Biol, 2004. **14**(9): p. 505-14.
39. Lange, A., et al., *Classical nuclear localization signals: definition, function, and interaction with importin alpha.* J Biol Chem, 2007. **282**(8): p. 5101-5.
40. Weis, K., U. Ryder, and A.I. Lamond, *The conserved amino-terminal domain of hSRP1 alpha is essential for nuclear protein import.* Embo J, 1996. **15**(8): p. 1818-25.
41. Riggleman, B., E. Wieschaus, and P. Schedl, *Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene.* Genes Dev, 1989. **3**(1): p. 96-113.
42. Peifer, M., S. Berg, and A.B. Reynolds, *A repeating amino acid motif shared by proteins with diverse cellular roles.* Cell, 1994. **76**(5): p. 789-91.
43. Komiya, Y. and R. Habas, *Wnt signal transduction pathways.* Organogenesis, 2008. **4**(2): p. 68-75.
44. Daniels, D.L., K. Eklof Spink, and W.I. Weis, *beta-catenin: molecular plasticity and drug design.* Trends Biochem Sci, 2001. **26**(11): p. 672-8.
45. Fagotto, F., U. Gluck, and B.M. Gumbiner, *Nuclear localization signal-independent and importin/karyopherin-independent nuclear import of beta-catenin.* Curr Biol, 1998. **8**(4): p. 181-90.
46. Eleftheriou, A., M. Yoshida, and B.R. Henderson, *Nuclear export of human beta-catenin can occur independent of CRM1 and the adenomatous polyposis coli tumor suppressor.* J Biol Chem, 2001. **276**(28): p. 25883-8.
47. Yokoya, F., et al., *beta-catenin can be transported into the nucleus in a Ran-unassisted manner.* Mol Biol Cell, 1999. **10**(4): p. 1119-31.

48. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
49. Gouet, P., et al., *ESPript: analysis of multiple sequence alignments in PostScript.* Bioinformatics, 1999. **15**(4): p. 305-8.
50. Fontes, M.R., et al., *Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha.* J Biol Chem, 2003. **278**(30): p. 27981-7.
51. Matsuura, Y., et al., *Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import.* Embo J, 2003. **22**(20): p. 5358-69.
52. Mayrose, I., et al., *Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior.* Mol Biol Evol, 2004. **21**(9): p. 1781-91.
53. Cutress, M.L., et al., *Structural basis for the nuclear import of the human androgen receptor.* J Cell Sci, 2008. **121**(Pt 7): p. 957-68.
54. Moroianu, J., G. Blobel, and A. Radu, *The binding site of karyopherin alpha for karyopherin beta overlaps with a nuclear localization sequence.* Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6572-6.
55. Gorlich, D., et al., *A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus.* Embo J, 1996. **15**(8): p. 1810-7.
56. Gorlich, D., et al., *Isolation of a protein that is essential for the first step of nuclear protein import.* Cell, 1994. **79**(5): p. 767-78.
57. Rexach, M. and G. Blobel, *Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins.* Cell, 1995. **83**(5): p. 683-92.
58. Fanara, P., et al., *Quantitative analysis of nuclear localization signal (NLS)-importin alpha interaction through fluorescence depolarization. Evidence for auto-inhibitory regulation of NLS binding.* J Biol Chem, 2000. **275**(28): p. 21218-23.
59. Catimel, B., et al., *Biophysical characterization of interactions involving importin-alpha during nuclear import.* J Biol Chem, 2001. **276**(36): p. 34189-98.
60. Kalderon, D., et al., *A short amino acid sequence able to specify nuclear location.* Cell, 1984. **39**(3 Pt 2): p. 499-509.
61. Robbins, J., et al., *Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence.* Cell, 1991. **64**(3): p. 615-23.
62. Hodel, M.R., A.H. Corbett, and A.E. Hodel, *Dissection of a nuclear localization signal.* J Biol Chem, 2001. **276**(2): p. 1317-25.
63. Hodel, A.E., et al., *Nuclear localization signal receptor affinity correlates with in vivo localization in Saccharomyces cerevisiae.* J Biol Chem, 2006. **281**(33): p. 23545-56.
64. Colledge, W.H., et al., *Extensive mutagenesis of the nuclear location signal of simian virus 40 large-T antigen.* Mol Cell Biol, 1986. **6**(11): p. 4136-9.
65. Chelsky, D., R. Ralph, and G. Jonak, *Sequence requirements for synthetic peptide-mediated translocation to the nucleus.* Mol Cell Biol, 1989. **9**(6): p. 2487-92.
66. Kosugi, S., et al., *Six classes of nuclear localization signals specific to different binding grooves of importin alpha*

Nuclear export signal consensus sequences defined using a localization-based yeast selection system

Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling. J Biol Chem, 2009. **284**(1): p. 478-85.
67. Fontes, M.R., et al., *Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha.* Biochem J, 2003. **375**(Pt 2): p. 339-49.

68. Tarendeau, F., et al., *Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit.* Nat Struct Mol Biol, 2007. **14**(3): p. 229-33.
69. Dias, S.M., et al., *The molecular basis for the regulation of the cap-binding complex by the importins.* Nat Struct Mol Biol, 2009. **16**(9): p. 930-7.
70. Kalderon, D., et al., *Sequence requirements for nuclear location of simian virus 40 large-T antigen.* Nature, 1984. **311**(5981): p. 33-8.
71. Gorlich, D., et al., *Identification of different roles for RanGDP and RanGTP in nuclear protein import.* Embo J, 1996. **15**(20): p. 5584-94.
72. Pyhtila, B. and M. Rexach, *A gradient of affinity for the karyopherin Kap95p along the yeast nuclear pore complex.* J Biol Chem, 2003. **278**(43): p. 42699-709.
73. Efthymiadis, A., et al., *Kinetic characterization of the human retinoblastoma protein bipartite nuclear localization sequence (NLS) in vivo and in vitro. A comparison with the SV40 large T-antigen NLS.* J Biol Chem, 1997. **272**(35): p. 22134-9.
74. Hu, W. and D.A. Jans, *Efficiency of importin alpha/beta-mediated nuclear localization sequence recognition and nuclear import. Differential role of NTF2.* J Biol Chem, 1999. **274**(22): p. 15820-7.
75. Chen, M.H., et al., *Phospholipid scramblase 1 contains a nonclassical nuclear localization signal with unique binding site in importin alpha.* J Biol Chem, 2005. **280**(11): p. 10599-606.
76. Lam, M.H., et al., *Importin beta recognizes parathyroid hormone-related protein with high affinity and mediates its nuclear import in the absence of importin alpha.* J Biol Chem, 1999. **274**(11): p. 7391-8.
77. Chook, Y.M. and G. Blobel, *Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp.* Nature, 1999. **399**(6733): p. 230-7.
78. Chook, Y.M., et al., *Uncoupling Kapbeta2 substrate dissociation and ran binding.* Biochemistry, 2002. **41**(22): p. 6955-66.
79. Lee, B.J., et al., *Rules for nuclear localization sequence recognition by karyopherin beta 2.* Cell, 2006. **126**(3): p. 543-58.
80. Cansizoglu, A.E., et al., *Structure-based design of a pathway-specific nuclear import inhibitor.* Nat Struct Mol Biol, 2007. **14**(5): p. 452-4.
81. Imasaki, T., et al., *Structural basis for substrate recognition and dissociation by human transportin 1.* Mol Cell, 2007. **28**(1): p. 57-67.
82. Jentsch, T.J., et al., *Molecular structure and physiological function of chloride channels.* Physiol Rev, 2002. **82**(2): p. 503-68.
83. Landry, D., et al., *Molecular cloning and characterization of p64, a chloride channel protein from kidney microsomes.* J Biol Chem, 1993. **268**(20): p. 14948-55.
84. Redhead, C.R., et al., *A ubiquitous 64-kDa protein is a component of a chloride channel of plasma and intracellular membranes.* Proc Natl Acad Sci U S A, 1992. **89**(9): p. 3716-20.
85. Shanks, R.A., et al., *AKAP350 at the Golgi apparatus. II. Association of AKAP350 with a novel chloride intracellular channel (CLIC) family member.* J Biol Chem, 2002. **277**(43): p. 40973-80.
86. Littler, D.R., et al., *Crystal structure of the soluble form of the redox-regulated chloride ion channel protein CLIC4.* FEBS J, 2005. **272**(19): p. 4996-5007.
87. Harrop, S.J., et al., *Crystal structure of a soluble form of the intracellular chloride ion channel CLIC1 (NCC27) at 1.4-A resolution.* J Biol Chem, 2001. **276**(48): p. 44993-5000.
88. Waldegger, S. and T.J. Jentsch, *From tonus to tonicity: physiology of CLC chloride channels.* J Am Soc Nephrol, 2000. **11**(7): p. 1331-9.
89. Dulhunty, A., et al., *The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator.* J Biol Chem, 2001. **276**(5): p. 3319-23.

90. Valenzuela, S.M., et al., *Molecular cloning and expression of a chloride ion channel of cell nuclei.* J Biol Chem, 1997. **272**(19): p. 12575-82.
91. Tulk, B.M. and J.C. Edwards, *NCC27, a homolog of intracellular Cl- channel p64, is expressed in brush border of renal proximal tubule.* Am J Physiol, 1998. **274**(6 Pt 2): p. F1140-9.
92. Berryman, M. and A. Bretscher, *Identification of a novel member of the chloride intracellular channel gene family (CLIC5) that associates with the actin cytoskeleton of placental microvilli.* Mol Biol Cell, 2000. **11**(5): p. 1509-21.
93. Petrova, D.T., et al., *Expression of chloride intracellular channel protein 1 (CLIC1) and tumor protein D52 (TPD52) as potential biomarkers for colorectal cancer.* Clin Biochem, 2008. **41**(14-15): p. 1224-36.
94. Ulmasov, B., et al., *Tissue and subcellular distribution of CLIC1.* BMC Cell Biol, 2007. **8**: p. 8.
95. Myers, K., et al., *Identification of chloride intracellular channel proteins in spermatozoa.* FEBS Lett, 2004. **566**(1-3): p. 136-40.
96. Tonini, R., et al., *Functional characterization of the NCC27 nuclear protein in stable transfected CHO-K1 cells.* Faseb J, 2000. **14**(9): p. 1171-8.
97. Singh, H. and R.H. Ashley, *Redox regulation of CLIC1 by cysteine residues associated with the putative channel pore.* Biophys J, 2006. **90**(5): p. 1628-38.
98. Valenzuela, S.M., et al., *The nuclear chloride ion channel NCC27 is involved in regulation of the cell cycle.* J Physiol, 2000. **529 Pt 3**: p. 541-52.
99. Warton, K., et al., *Recombinant CLIC1 (NCC27) assembles in lipid bilayers via a pH-dependent two-state process to form chloride ion channels with identical characteristics to those observed in Chinese hamster ovary cells expressing CLIC1.* J Biol Chem, 2002. **277**(29): p. 26003-11.
100. Tulk, B.M., S. Kapadia, and J.C. Edwards, *CLIC1 inserts from the aqueous phase into phospholipid membranes, where it functions as an anion channel.* Am J Physiol Cell Physiol, 2002. **282**(5): p. C1103-12.
101. Ashley, R.H., *Challenging accepted ion channel biology: p64 and the CLIC family of putative intracellular anion channel proteins (Review).* Mol Membr Biol, 2003. **20**(1): p. 1-11.
102. Littler, D.R., et al., *The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition.* J Biol Chem, 2004. **279**(10): p. 9298-305.
103. Edwards, J.C., *The CLIC1 chloride channel is regulated by the cystic fibrosis transmembrane conductance regulator when expressed in Xenopus oocytes.* J Membr Biol, 2006. **213**(1): p. 39-46.
104. Singh, H., M.A. Cousin, and R.H. Ashley, *Functional reconstitution of mammalian 'chloride intracellular channels' CLIC1, CLIC4 and CLIC5 reveals differential regulation by cytoskeletal actin.* Febs J, 2007. **274**(24): p. 6306-16.
105. Novarino, G., et al., *Involvement of the intracellular ion channel CLIC1 in microglia-mediated beta-amyloid-induced neurotoxicity.* J Neurosci, 2004. **24**(23): p. 5322-30.
106. Milton, R.H., et al., *CLIC1 function is required for beta-amyloid-induced generation of reactive oxygen species by microglia.* J Neurosci, 2008. **28**(45): p. 11488-99.
107. Paradisi, S., et al., *Blockade of chloride intracellular ion channel 1 stimulates Abeta phagocytosis.* J Neurosci Res, 2008. **86**(11): p. 2488-98.
108. Board, P.G., et al., *Identification, characterization, and crystal structure of the Omega class glutathione transferases.* J Biol Chem, 2000. **275**(32): p. 24798-806.
109. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.* Nucleic Acids Res, 2007. **35**(Database issue): p. D61-5.
110. Suh, K.S., et al., *Reciprocal modifications of CLIC4 in tumor epithelium and stroma mark malignant progression of multiple human cancers.* Clin Cancer Res, 2007. **13**(1): p. 121-31.

111. Cocco, R., et al., *The folding and stability of human alpha class glutathione transferase A1-1 depend on distinct roles of a conserved N-capping box and hydrophobic staple motif.* J Biol Chem, 2001. **276**(34): p. 32177-83.
112. Cromer, B.A., et al., *Structure of the Janus protein human CLIC2.* J Mol Biol, 2007. **374**(3): p. 719-31.
113. Sheehan, D., et al., *Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily.* Biochem J, 2001. **360**(Pt 1): p. 1-16.
114. Fanucchi, S., R.J. Adamson, and H.W. Dirr, *Formation of an unfolding intermediate state of soluble chloride intracellular channel protein CLIC1 at acidic pH.* Biochemistry, 2008. **47**(44): p. 11674-81.
115. Howell, S., R.R. Duncan, and R.H. Ashley, *Identification and characterisation of a homologue of p64 in rat tissues.* FEBS Lett, 1996. **390**(2): p. 207-10.
116. Duncan, R.R., et al., *Rat brain p64H1, expression of a new member of the p64 chloride channel protein family in endoplasmic reticulum.* J Biol Chem, 1997. **272**(38): p. 23880-6.
117. Edwards, J.C., *A novel p64-related Cl- channel: subcellular distribution and nephron segment-specific expression.* Am J Physiol, 1999. **276**(3 Pt 2): p. F398-408.
118. Fernandez-Salas, E., et al., *p53 and tumor necrosis factor alpha regulate the expression of a mitochondrial chloride channel protein.* J Biol Chem, 1999. **274**(51): p. 36488-97.
119. Chuang, J.Z., et al., *A 29 kDa intracellular chloride channel p64H1 is associated with large dense-core vesicles in rat hippocampal neurons.* J Neurosci, 1999. **19**(8): p. 2919-28.
120. Proutski, I., N. Karoulias, and R.H. Ashley, *Overexpressed chloride intracellular channel protein CLIC4 (p64H1) is an essential component of novel plasma membrane anion channels.* Biochem Biophys Res Commun, 2002. **297**(2): p. 317-22.
121. Berryman, M.A. and J.R. Goldenring, *CLIC4 is enriched at cell-cell junctions and colocalizes with AKAP350 at the centrosome and midbody of cultured mammalian cells.* Cell Motil Cytoskeleton, 2003. **56**(3): p. 159-72.
122. Suh, K.S., et al., *CLIC4 mediates and is required for Ca2+-induced keratinocyte differentiation.* J Cell Sci, 2007. **120**(Pt 15): p. 2631-40.
123. Suh, K.S., et al., *The organellular chloride channel protein CLIC4/mtCLIC translocates to the nucleus in response to cellular stress and accelerates apoptosis.* J Biol Chem, 2004. **279**(6): p. 4632-41.
124. Singh, H. and R.H. Ashley, *CLIC4 (p64H1) and its putative transmembrane domain form poorly selective, redox-regulated ion channels.* Mol Membr Biol, 2007. **24**(1): p. 41-52.
125. Fernandez-Salas, E., et al., *mtCLIC/CLIC4, an organellular chloride channel protein, is increased by DNA damage and participates in the apoptotic response to p53.* Mol Cell Biol, 2002. **22**(11): p. 3610-20.
126. Shiio, Y., et al., *Quantitative proteomic analysis of myc-induced apoptosis: a direct role for Myc induction of the mitochondrial chloride ion channel, mtCLIC/CLIC4.* J Biol Chem, 2006. **281**(5): p. 2750-6.
127. Ronnov-Jessen, L., et al., *Differential expression of a chloride intracellular channel gene, CLIC4, in transforming growth factor-beta1-mediated conversion of fibroblasts to myofibroblasts.* Am J Pathol, 2002. **161**(2): p. 471-80.
128. Suh, K.S., et al., *CLIC4, skin homeostasis and cutaneous cancer: surprising connections.* Mol Carcinog, 2007. **46**(8): p. 599-604.
129. Bohman, S., et al., *Proteomic analysis of vascular endothelial growth factor-induced endothelial cell differentiation reveals a role for chloride intracellular channel 4 (CLIC4) in tubular morphogenesis.* J Biol Chem, 2005. **280**(51): p. 42397-404.
130. Berry, K.L., et al., *A C. elegans CLIC-like protein required for intracellular tube formation and maintenance.* Science, 2003. **302**(5653): p. 2134-7.

131. Tung, J.J., et al., *Chloride intracellular channel 4 is involved in endothelial proliferation and morphogenesis in vitro.* Angiogenesis, 2009. **27**: p. 27.
132. Ulmasov, B., et al., *Chloride intracellular channel protein-4 functions in angiogenesis by supporting acidification of vacuoles along the intracellular tubulogenic pathway.* Am J Pathol, 2009. **174**(3): p. 1084-96.
133. Li, Y., et al., *Trimeric structure of the wild soluble chloride intracellular ion channel CLIC4 observed in crystals.* Biochem Biophys Res Commun, 2006. **343**(4): p. 1272-8.
134. Berry, K.L. and O. Hobert, *Mapping functional domains of chloride intracellular channel (CLIC) proteins in vivo.* J Mol Biol, 2006. **359**(5): p. 1316-33.
135. Guindon, S. and O. Gascuel, *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.* Syst Biol, 2003. **52**(5): p. 696-704.
136. Buechner, M., et al., *Cystic canal mutants in Caenorhabditis elegans are defective in the apical membrane domain of the renal (excretory) cell.* Dev Biol, 1999. **214**(1): p. 227-41.
137. Littler, D.R., et al., *Comparison of vertebrate and invertebrate CLIC proteins: the crystal structures of Caenorhabditis elegans EXC-4 and Drosophila melanogaster DmCLIC.* Proteins, 2008. **71**(1): p. 364-78.
138. Schmidt, P.H., et al., *AKAP350, a multiply spliced protein kinase A-anchoring protein associated with centrosomes.* J Biol Chem, 1999. **274**(5): p. 3055-66.
139. Suginta, W., et al., *Chloride intracellular channel protein CLIC4 (p64H1) binds directly to brain dynamin I in a complex containing actin, tubulin and 14-3-3 isoforms.* Biochem J, 2001. **359**(Pt 1): p. 55-64.
140. Lehner, B., et al., *Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region.* Genomics, 2004. **83**(1): p. 153-67.
141. Lehner, B. and C.M. Sanderson, *A protein interaction framework for human mRNA degradation.* Genome Res, 2004. **14**(7): p. 1315-23.
142. Fan, L., W. Yu, and X. Zhu, *Interaction of Sedlin with chloride intracellular channel proteins.* FEBS Lett, 2003. **540**(1-3): p. 77-80.
143. Maeda, K., et al., *CLIC4 interacts with histamine H3 receptor and enhances the receptor cell surface expression.* Biochem Biophys Res Commun, 2008. **369**(2): p. 603-8.
144. Hill, S.J., et al., *International Union of Pharmacology. XIII. Classification of histamine receptors.* Pharmacol Rev, 1997. **49**(3): p. 253-78.
145. Qian, Z., et al., *Molecular cloning and characterization of a mitogen-activated protein kinase-associated intracellular chloride channel.* J Biol Chem, 1999. **274**(3): p. 1621-7.
146. Abe, M.K., et al., *Extracellular signal-regulated kinase 7 (ERK7), a novel ERK with a C-terminal domain that regulates its activity, its cellular localization, and cell growth.* Mol Cell Biol, 1999. **19**(2): p. 1301-12.
147. Griffon, N., et al., *CLIC6, a member of the intracellular chloride channel family, interacts with dopamine D(2)-like receptors.* Brain Res Mol Brain Res, 2003. **117**(1): p. 47-57.
148. Berryman, M., et al., *CLIC-5A functions as a chloride channel in vitro and associates with the cortical actin cytoskeleton in vitro and in vivo.* J Biol Chem, 2004. **279**(33): p. 34794-801.
149. Edwards, J.C., B. Tulk, and P.H. Schlesinger, *Functional expression of p64, an intracellular chloride channel protein.* J Membr Biol, 1998. **163**(2): p. 119-27.
150. Tamir, H., et al., *Secretogogue-induced gating of chloride channels in the secretory vesicles of parafollicular cells.* Endocrinology, 1994. **135**(5): p. 2045-57.
151. Edwards, J.C. and S. Kapadia, *Regulation of the bovine kidney microsomal chloride channel p64 by p59fyn, a Src family tyrosine kinase.* J Biol Chem, 2000. **275**(41): p. 31826-32.

152. Edwards, J.C., et al., *c-Src control of chloride channel support for osteoclast HCl transport and bone resorption.* J Biol Chem, 2006. **281**(38): p. 28011-22.
153. Stein, N., *CHAINSAW: a program for mutating pdb files used as templates in molecular replacement.* Journal of Applied Crystallography, 2008. **41**(3): p. 641-643.
154. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(4): p. 658-674.
155. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
156. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
157. Langer, G., et al., *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.* Nat Protoc, 2008. **3**(7): p. 1171-9.
158. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.
159. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.
160. Meritt, E.A., *X-ray Anomalous Scattering.* Biomolecular Structure Center, http://skuld.bmsc.washington.edu/scatter/.
161. Cromer, D.T. and D. Liberman, *Relativistic calculation of anomalous scattering factors for X-rays.* J. Chem. Phys., 1970. **53**: p. 1891-1898.
162. Muller, P., S. Kopke, and G.M. Sheldrick, *Is the bond-valence method able to identify metal atoms in protein structures?* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 1): p. 32-7.
163. Read, R.J., *Improved Fourier coefficients for maps using phases from partial structures with errors.* Acta Crystallographica Section A, 1986. **42**(3): p. 140-149.
164. Rhodes, G., *Crystallography Made Crystal Clear*. 3rd ed. 2006.
165. DeLano, W.L., *The PyMol Molecular Graphics System.* http://www.pymol.org, 2002.
166. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-23.
167. *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
168. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.
169. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
170. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
171. Martell, A.E. and R.M. Smith, *Critical Stability Constants*. 1974, Plenum Press: New York. p. 204-211.
172. Mi, W., et al., *The crystal structure of human chloride intracellular channel protein 2: a disulfide bond with functional implications.* Proteins, 2008. **71**(1): p. 509-13.
173. Yang, Y.F. and W.W. Wells, *Identification and characterization of the functional amino acids at the active center of pig liver thioltransferase by site-directed mutagenesis.* J Biol Chem, 1991. **266**(19): p. 12759-65.
174. Nelson, J.W. and T.E. Creighton, *Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation in vivo.* Biochemistry, 1994. **33**(19): p. 5974-83.
175. Kortemme, T. and T.E. Creighton, *Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family.* J Mol Biol, 1995. **253**(5): p. 799-812.

176. Di Cera, E., *A structural perspective on enzymes activated by monovalent cations.* J Biol Chem, 2006. **281**(3): p. 1305-8.
177. Teh, T., T. Tiganis, and B. Kobe, *Crystallization of importin alpha, the nuclear-import receptor.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 2): p. 561-3.
178. Fontes, M.R., T. Teh, and B. Kobe, *Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha.* J Mol Biol, 2000. **297**(5): p. 1183-94.
179. Kussel, P. and M. Frasch, *Yeast Srp1, a nuclear protein related to Drosophila and mouse pendulin, is required for normal migration, division, and integrity of nuclei during mitosis.* Mol Gen Genet, 1995. **248**(3): p. 351-63.
180. Leslie, A.G., *The integration of macromolecular diffraction data.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 48-57.
181. Evans, P., *Scaling and assessment of data quality.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.
182. A. J. McCoy, R.W.G.-K.P.D.A.M.D.W.L.C.S. and R.J. Read, *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(4): p. 658-674.
183. Wouters, J., *Cation-pi (Na+-Trp) interactions in the crystal structure of tetragonal lysozyme.* Protein Sci, 1998. **7**(11): p. 2472-5.
184. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
185. Smith, D.K., et al., *Improved amino acid flexibility parameters.* Protein Sci, 2003. **12**(5): p. 1060-72.
186. Branden, C. and A. Jones, *Between objectivity and subjectivity.* Nature, 1990. **343**: p. 687-689.
187. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. **28**(1): p. 235-242.
188. Word, J.M., et al., *Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms.* J Mol Biol, 1999. **285**(4): p. 1711-33.
189. Richardson, D.C. and J.S. Richardson, *The kinemage: a tool for scientific communication.* Protein Sci, 1992. **1**(1): p. 3-9.
190. Gallivan, J.P. and D.A. Dougherty, *Cation-pi interactions in structural biology.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9459-64.
191. Moreno, J.M., et al., *Characterization of the domains of E. coli initiation factor IF2 responsible for recognition of the ribosome.* FEBS Lett, 1999. **455**(1-2): p. 130-4.
192. Moreno, J.M., et al., *Binding of Escherichia coli initiation factor IF2 to 30S ribosomal subunits: a functional role for the N-terminus of the factor.* Biochem Biophys Res Commun, 1998. **252**(2): p. 465-71.
193. Bollen, A., et al., *Cross-linking of initiation factor IF-2 to Escherichia coli 30 S ribosomal proteins with dimethylsuberimidate.* J Biol Chem, 1975. **250**(11): p. 4310-4.
194. Wickner, S., et al., *A molecular chaperone, ClpA, functions like DnaK and DnaJ.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 12218-22.
195. Hoskins, J.R., et al., *The role of the ClpA chaperone in proteolysis by ClpAP.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12135-40.
196. Schuwirth, B.S., et al., *Structures of the bacterial ribosome at 3.5 A resolution.* Science, 2005. **310**(5749): p. 827-34.
197. Carugo, O. and P. Argos, *Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors.* Proteins, 1998. **31**(2): p. 201-13.
198. Laursen, B.S., et al., *Initiation of protein synthesis in bacteria.* Microbiol Mol Biol Rev, 2005. **69**(1): p. 101-23.
199. Wu, X.Q. and U.L. RajBhandary, *Effect of the amino acid attached to Escherichia coli initiator tRNA on its affinity for the initiation factor IF2 and on the IF2 dependence of its binding to the ribosome.* J Biol Chem, 1997. **272**(3): p. 1891-5.

200. Caldas, T., S. Laalami, and G. Richarme, *Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2.* J Biol Chem, 2000. **275**(2): p. 855-60.
201. Neuwald, A.F., et al., *AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes.* Genome Res, 1999. **9**(1): p. 27-43.
202. Guo, F., et al., *Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease.* J Biol Chem, 2002. **277**(48): p. 46743-52.
203. Bubunenko, M., et al., *30S ribosomal subunits can be assembled in vivo without primary binding ribosomal protein S15.* Rna, 2006. **12**(7): p. 1229-39.
204. Suh, K.S., et al., *CLIC4, an intracellular chloride channel protein, is a novel molecular target for cancer therapy.* J Investig Dermatol Symp Proc, 2005. **10**(2): p. 105-9.
205. De Wall, S.L., et al., *Synthetic receptors as models for alkali metal cation-pi binding sites in proteins.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6271-6.
206. Hubner, S., C.Y. Xiao, and D.A. Jans, *The protein kinase CK2 site (Ser111/112) enhances recognition of the simian virus 40 large T-antigen nuclear localization sequence by importin.* J Biol Chem, 1997. **272**(27): p. 17191-5.
207. Xiao, C.Y., S. Hubner, and D.A. Jans, *SV40 large tumor antigen nuclear import is regulated by the double-stranded DNA-dependent protein kinase site (serine 120) flanking the nuclear localization sequence.* J Biol Chem, 1997. **272**(35): p. 22191-8.
208. De Corte, V., et al., *Increased importin-beta-dependent nuclear import of the actin modulating protein CapG promotes cell invasion.* J Cell Sci, 2004. **117**(Pt 22): p. 5283-92.
209. Greco, T.M., et al., *Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells.* Proc Natl Acad Sci U S A, 2006. **103**(19): p. 7420-5.
210. Schreiter, E.R., et al., *S-nitrosylation-induced conformational change in blackfin tuna myoglobin.* J Biol Chem, 2007. **282**(27): p. 19773-80.
211. Weerapana, E., G.M. Simon, and B.F. Cravatt, *Disparate proteome reactivity profiles of carbon electrophiles.* Nat Chem Biol, 2008. **4**(7): p. 405-7.
212. Schroeder, P., et al., *Nuclear redox-signaling is essential for apoptosis inhibition in endothelial cells--important role for nuclear thioredoxin-1.* Arterioscler Thromb Vasc Biol, 2007. **27**(11): p. 2325-31.
213. Murzin, A.G., *Biochemistry. Metamorphic proteins.* Science, 2008. **320**(5884): p. 1725-6.
214. Cokol, M., R. Nair, and B. Rost, *Finding nuclear localization signals.* EMBO Rep, 2000. **1**(5): p. 411-5.
215. Nguyen Ba, A.N., et al., *NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction.* BMC Bioinformatics, 2009. **10**(202): p. 202.