Structural and functional characterisation of the molecular assembly of two pseudokinase scaffolds

Ashleigh Kate Kropp

ORCID: 0000-0002-8027-9050

A dissertation submitted in total fulfillment of the requirements of the degree of Master of Philosophy September 2019

Supervisors:

A/Prof Isabelle Lucet and Dr Onisha Patel

The Walter and Eliza Hall Institute of Medical Research

Department of Medical Biology

Faculty of Medicine, Dentistry and Health Sciences

The University of Melbourne

Abstract

The human pseudokinase SgK269, and its structurally related homologue SgK223, are oncogenic interacting scaffolds that promote the assembly of specific tyrosine kinase signalling pathways [1]. SgK223 and SgK269, as well as the recently discovered PEAK3, belong to the PEAK family of protein pseudokinases. They are large, multidomain proteins that are comprised of an N-terminal region of unknown structure and function, a large unstructured PEST region containing tyrosine phosphorylation sites and a C-terminal domain comprised of a pseudokinase domain flanked by regulatory helices. SgK223 and SgK269 have been shown to localise to focal adhesions and their overexpression leads to increased cell migration and changes in cell morphology, hallmarks of cancerous cells [1-5].

Recent studies from our lab and others have provided structural insight into the C-terminal domain and flanking alpha helices of SgK223 and SgK269. These structures highlighted a conserved mechanism of dimerisation that drives homo- and hetero-association of SgK223 and SgK269 and plays an important role in cell migration [6-8]. Additionally, SgK223 and SgK269 were demonstrated to undergo homo- and hetero-oligomerisation through their pseudokinase domains [6]. In contrast to the C-terminal domain, little is known about the function of the N-terminal domains of SgK223 and SgK269, although there is sequence conservation between them.

In this study, we have begun characterising the N-terminal domains of SgK223 and SgK269 using biophysical and biochemical techniques, initially demonstrating that these domains are monomeric and appear to have no defined secondary structure. To further investigate SgK223 and SgK269 homo- and hetero-association we carried out single site alanine mutagenesis to determine the energetic hotspots at the dimerisation interface of SgK269. Furthermore, we carried out mutagenesis within the N-lobe of SgK223 and SgK269, to investigate the role of this interface in homo- and hetero-oligomerisation.

Additionally, we characterised the PEAK family interactions with the critical interacting signalling adaptor protein, CrkII, using biophysical assays and X-ray crystallography. We found that each member of the PEAK family has a proline-rich motif within their PEST linker that interacts with CrkII N-SH3 domain with ~1-3 μM affinity. The crystal structure of the CrkII N-SH3 domain bound to the SgK269 proline-rich motif demonstrated the critical consensus residues for the PEAK family interaction with CrkII.

To further investigate the role of SgK223 and SgK269 homo- and hetero-association in cells, these studies were complemented with localisation microscopy techniques. Utilising mutants of SgK223 and SgK269 that can no longer dimerise or oligomerise, we investigated the importance of SgK223 and SgK269 associations for their localisation and thus, role in signalling. Insights into the scaffolding functions of SgK223 and SgK269 will inform how they contribute to the assembly of signalling pathways and hence their role in cancer.

Declaration

This is to certify that:

(i) this thesis comprises only my original work towards the MPhil except where indicated in the Preface and the text,

(ii) due acknowledgement has been made in the text to all other material used, and

(iii) the thesis is fewer than 50,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Ashleigh Kate Kropp

Preface

I acknowledge that majority of the work in this Thesis I carried out, however some of the work in this Thesis was carried out with the contributors listed below:

Chapter 3 was primarily my own work, with the exception of the analytical ultracentrifugation experiments, in which I carried out the protein preparation and data analysis and Dr. Michael Griffin from Bio21 operated the ultracentrifuges and carried out the AUC experiments (Figures 3.2.2b, 3.4.3c, 3.4.4c). Additionally, Weiwen Dai provided assistance with some of the protein purification.

Chapter 4 had the assistance of Dr. Michael Roy and Dr. Onisha Patel from WEHI. Specifically, Dr. Michael Roy and Dr. Onisha Patel collected data on the crystals at the Australian Synchrotron and Dr. Michael Roy processed the data, after which I solved, refined and analysed the structures (Figures 4.2.3, 4.2.4, 4.2.5, 4.2.6b).

Chapter 5 was assisted by Dr. Kelly Rogers and Dr. Michael Mlodzianoski (from WEHI) who operated the microscopes the data were acquired on (Figures 5.2.2, 5.2.3, 5.2.4). Dr. Onisha Patel assisted in some of the cloning.

Acknowledgements

There are numerous people who contributed to my MPhil Thesis and also my life over the past two years, without whom I could not have achieved what I have. To begin, I must acknowledge my supervisors, Isabelle Lucet and Onisha Patel. I appreciate all the help, insight and knowledge they have passed on to me over the past (nearly) four years I have been with them. From Honours, to working for them as a Research Assistant, and now my MPhil, they have been great supervisors and brilliant scientists, always with my best interests at heart. I have enjoyed watching our lab expand and achieve great things over the years and I have no doubt this will continue.

I would like to also thank my committee, Kelly, Roger, James and Mike, who have been supportive throughout my MPhil and provided great ideas for my project. All of my committee were collaborators and helped in some aspect of my project, for which I am thankful. I would like to specifically thank Mike Griffin for his help with AUC and taking the time to teach me AUC data analysis. Additionally, Kelly Rogers and Mike Mlodzianoski were so incredibly helpful and full of so many great ideas and advice for Chapter 5 of my Thesis. Undoubtedly without them I would not have had the opportunity to learn microscopy and use it for my project. Additionally, Janet Newman and the C3 facility as well as the Australian synchrotron have been invaluable for my project, facilitating the crystallography work in Chapter 4.

I need to thank Keely Bumsted-Obrien and Sue Hardy who supported me and continue to support every student at WEHI greatly, especially through the difficult times. I have to thank WEHI for the top up scholarship and the Melbourne Research Scheme scholarship for funding my studies.

I also am so thankful to the members of the Lucet lab, who have been nothing but supportive, friendly and knowledgeable. Lung-Yu has been a wonderful student lab-mate, who is always smiling and ready to help. Journal club with

Lung-Yu was a great initiative that I really enjoyed over the past year. Weiwen has been my longest lab-mate, she taught me most of what I know in the lab and is a master of protein research techniques. Thank you for always teaching me, helping me and being such a lovely person to spend days in the lab with. I am so grateful to have had you as our RA. I received so much support and insight from Michael, especially for Chapter 4. Michael is so knowledgeable and skilled, but also a very patient and thorough teacher, and I believe he will make an incredible supervisor in the future.

I was fortunate to be in the Chemical Biology division, particularly on 5E. My colleagues here became my close friends, and although the division has changed vastly since when I began, I maintained those friendships. I want to thank the OG division members, and some of my closest friends and mentors: Kate, Marija and Lucy. You are all the best. I have no doubt that I would not be here without your constant support and love. In addition, Michelle who was incredibly supportive as our Division Coordinator and then as a friend after this. I also want to thank Leeanne, her tireless work to make CBD a great place to be is appreciated and I am so thankful for your friendship. To my current CBD lab-mates, Mengxiao, Leigh, Ying, James, Stef, Michael and Alan, thank you for your ongoing friendship, support, help and advice over the past years. You all make CBD a great place to be and I'm sad to be leaving such a lovely work environment.

WEHI has been an incredible place to work and study and I will be forever grateful for the opportunity to work here. From biscuits every day, to the wonderful student cohort, WEHI has enriched my life significantly. There are so many students who have made my MPhil enjoyable. As well as being a supportive, they are always willing to listen to my complaints or triumphs. Specifically, I'd like to thank Beth, who has been writing her thesis at the same time as I was, which meant that we shared many frustrations and joys. It has been great to have someone at the same stage as me, going through the same process, so we can share in how we feel and be a constant support for each other. I would like to thank the WESA committee members, who make WEHI such a fun and supportive place for

students. A highlight of my time here has been being on that committee and I am so thankful for meeting everyone on the committee and working closely with them.

Finally, I need to thank my family and friends. Shaz and Mikey, my mum and dad, are always so loving and encouraging of everything I do, and this is no exception. Although they may not quite understand what it is I research or how I go about it, they always endeavor to understand and help in any way they can. Without their support, I would not be in the position I am. Thank you also to Nick and Kim, who ask about my research even when they are confused and I make no sense, your support is appreciated. One of the MVP's of my MPhil is my housemate Kayla, who really endured it all, the highs to the lows, she was a constant support and always on Team Ash. Rosie is another MVP, being a fellow WEHI student and friend who I could talk to about anything and would listen without judgement. I'm sad that I will no longer be able to go on snack and coffee runs, or vent to you over tea at WEHI, but I know that our friendship began well before we ever knew WEHI existed and will continue beyond WEHI. The last MVP is James who really came in at the worst time but endured it all, providing support and helping me believe in myself.

Table of contents

Abbreviations

Activation loop	A-loop
Adenosine triphosphate	ATP
Ammonium persulfate	APS
Analytical ultracentrifugation	AUC
Big dye terminator	BDT
CaM-activated serine-threonine kinase	CASK
Catalytic spine	C-spine
CCAAT-enhancer-binding proteins	C/EBP
Circular dichroism	CD
c-Jun kinase	JNK
Column volume	CV
Constitutive photomorphogenesis protein 1	COP-1
C-terminal Src kinase	Csk
Degrees Celsius	°C
Dimethylsulfoxide	DMSO
Dithiothreitol	DTT
Dulbecco's modified eagle's medium	DMEM
Elution volume	V_e
Epidermal growth factor	EGF
Epidermal growth factor receptor	EGFR
Escherichia coli	E.coli
Ethylenediaminetetraacetic acid	EDTA
1-ethyl-3(3-dimethylaminopropyl) carbodiimide hydrochloride	EDC
Extracellular matrix	ECM
Extracellular-signal-regulated kinase 5	ERK5
Family with sequence similarity 20	FAM20
Focal adhesion	FA
Focal adhesion kinase	FAK
Fetal calf serum	FCS
Food and drug administration	FDA
Gastric cancer	GC
Geometric column volume	V_c
GTPase activating proteins	GAP
Guanosine diphosphate	GDP
Guanosine triphosphate	GTP
Human epidermal growth factor receptor	HER
Human pancreatic duct epithelial cells	HPDE
Integrin-linked kinase	ILK
Isopropyl β-D-1-thiogalactopyranoside	IPTG
Janus homology domain	JH
KiloDalton	kDa
Kinase suppressor of Ras	KSR
Litres	L
Liver kinase B1	LKB1
Long chain biotin tag	LcBiotin

Luria broth	LB
Mammalian target of rapamycin	mTOR
Mammary epithelial cells	MCF-10A
Mean residual MW	MRW
microMolar	μM
microLitres	μL
Midpoint of melting	Tm
milliAmp	mA
milliLitres	mL
milliMolar	mM
Mitogen activated protein kinase	MAPK
Mitogen activated protein kinase kinase	MEK
Molar	M
Molecular weight	MW
Mouse protein 25	MO25
nanoLitres	nL
Nerve growth factor	NGF
New kinase family 3	NKF3
Nickel-nitrilotriacetic acid	Ni-NTA
N-Hydroxysuccinimide	NHS
Non-small cell lung cancer	NSCLC
Pancreatic ductal adenocarcinoma	PDAC
Particularly interesting new Cys-His protein 1	PINCH1
Partition coefficient	K_{av}
Phosphate buffered saline	PBS
Phosphotyrosine binding	PTB
Pleckstrin homology	PH
Polyacrylamide gel electrophoresis	PAGE
Polyhydramnios megalencephaly symptomatic epilepsy	PMSE
Post translational modification	PTM
Proline glutamate serine threonine linker	PEST
Proline rich motif	PRM
Protein data bank	PDB
Protein kinase A	PKA
Protein kinase R	PKR
Pseudokinase	PsK
Pseudopodium enriched atypical kinase 3	PEAK3
Rapidly accelerated fibrosarcoma	RAF
Regulatory spine	R-spine
Response units	RU
Revolutions per minute	RPM
Rho family GTPase 2	RND2
Root mean square deviation	RMSD
Roswell Park Memorial Institute medium	RPMI
Sedimentation coefficient	*c(s)*
Small angle X-ray scattering	SAXS
Src-homology 2	SH2

Src-homology 3	SH3
Src family kinase	SFK
Signal transducer and activator of transcription 3	STAT3
Size exclusion chromatography	SEC
Sodium chloride	NaCl
Sodium dodecyl sulfate	SDS
STE20-relalted kinase adaptor protein alpha	STRADα
Sugen kinase 223	SgK223
Sugen kinase 269	SgK269
Super broth	SB
Surface Plasmon Resonance	SPR
Tetramethylethylene-diamine	TEMED
Thermal stability assay	TSA
Tobacco Etch Virus	TEV
Total internal reflection fluorescence	TIRF
tris(2-carboxyethyl)phosphine	TCEP
Vaccinia related kinase 3	VRK3
Void volume	V_o
Wild type	WT

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 8:

List of Tables

1.0 Introduction

1.1 Cellular signalling through kinases

Cellular signalling underpins all cellular functions and aberrations to this leads to diseases. Signalling is the process of a cell detecting a stimulus in the extracellular environment and transmitting this internally, to stimulate a host of cellular responses. It is critical for cells to correctly respond to their environment through signalling pathways, and in doing so, maintain homeostasis [9]. Signalling pathways are reliant on chemical modifications of many proteins within signalling cascades, in order to transmit the message to different parts of the cell [9]. There are a multitude of modifications that occur in cells post-translationally, one important signalling modification is the addition of a phosphate group, a reaction catalysed by protein kinases.

There are over 500 kinases identified in humans [10]. Kinases catalyse the transfer of the terminal phosphate group of adenosine triphosphate (ATP) onto the hydroxyl group of an acceptor substrate. All kinases share a distinctive conserved kinase domain, comprised of two lobes: a large C-lobe consisting of α-helices, and a smaller N-lobe consisting of primarily β-strands and one helix, the critical αC helix (Fig. 1.1a) [10]. At the core of the kinase domain, between these two lobes, is the ATP binding site. All kinases must have conformational flexibility around this critical ATP binding site in order to facilitate binding of ATP and switching between an active and inactive state. This switching allows signalling cascades to be dynamic; turning on and off rapidly and as required, thereby supporting correct signalling. There are a number of key catalytic motifs that are conserved amongst kinases and are critical to their function. These include: the His-Arg-Asp (HRD) motif of the catalytic loop, that contributes the catalytic aspartate that activates the incoming substrate for phospho-transfer; the Asp-Phe-Gly (DFG) motif that coordinates the β and γ phosphates of ATP via a divalent ion; Val-Ala-Ile-Lys (VAIK) motif, whose lysine interacts with the α and β phosphates of ATP; a critical glutamate in the αC helix that coordinates the VAIK

lysine; and the Ala-Pro-Glu (APE) motif of the activation loop (A-loop) (Fig. 1.1a) [11]. Activation of a kinase usually relies on phosphorylation of the A-loop on a Ser/Thr/Tyr residue to induce a conformational change required for correct positioning of the catalytic motifs for catalysis [11].

Additionally, kinase activation relies on the correct alignment of the residues that make up the kinase hydrophobic core, which comprises two spines, the catalytic spine (C-spine) and regulatory spine (R-spine) [12]. The hydrophobic residues that form the C-spine stem from the αF-helix, highlighted on PKA in Figure 1.1a. Residues from the αD helix and the β7 strand contribute to this spine and the adenine ring of ATP stacks between the hydrophobic residues to complete the C-spine (Fig. 1.1a) [12]. The R-spine is formed by residues in the important catalytic motifs, the HRD motif through the histidine (YRD in PKA, and the tyrosine is involved in the spine), the DFG motif through the phenylalanine, and the αC helix (Fig.1.1a) [12, 13]. Mutations of the critical catalytic residues or misalignment of the R- and C-spines can lead to an inactive kinase.

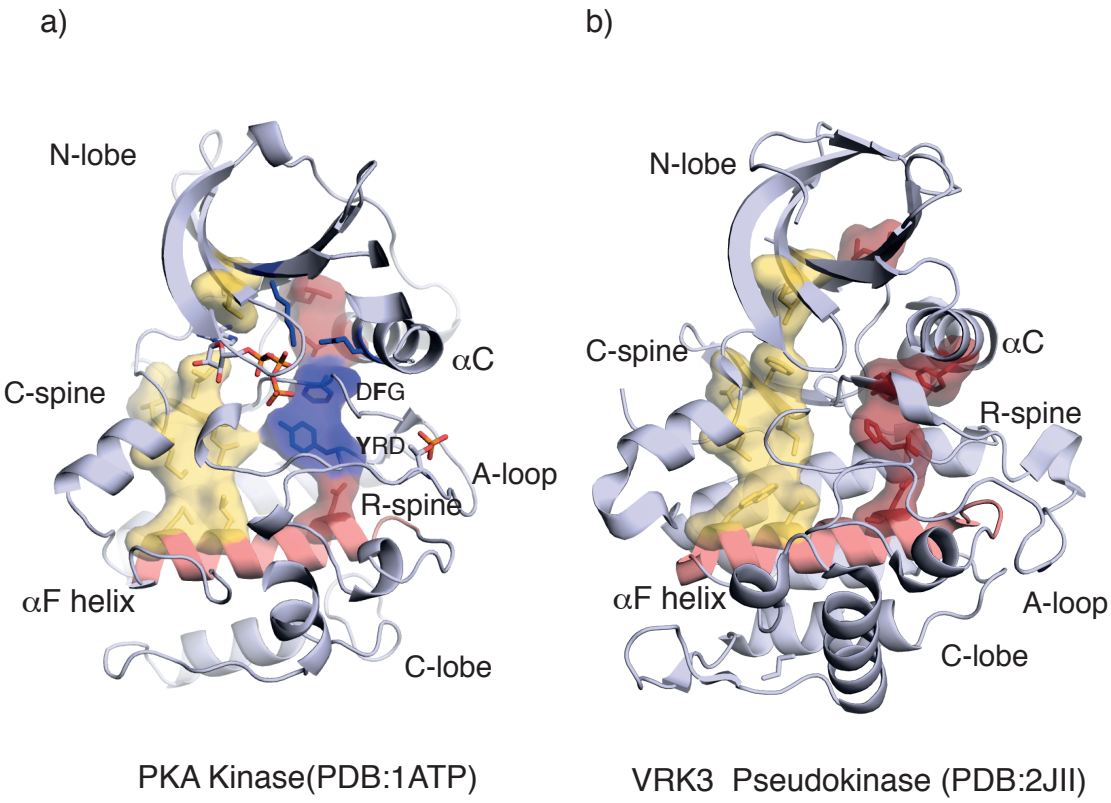


Figure 1.1. Typical kinase PKA (left) and pseudokinase VRK3 (right).
a) PKA represents the structure of a typical kinase. The regulatory spine (R-spine) is shown in red and catalytic spine (C-spine) in yellow. The αF-helix is shown in salmon. ATP is bound and the spines are aligned for catalysis. Residues important for catalysis are shown in blue.
b) VRK3 represents the structure of a pseudokinase. The R-spine is highlighted in red, not aligned in this case, and C-spine in yellow. The αF-helix is shown in salmon. VRK3 cannot bind ATP so cannot complete the C-spine with the ATP adenine ring as seen in PKA. The R-spine is not aligned and hence catalysis cannot occur.

1.2 Non-catalytic functions of kinases

Aside from their catalytic function, kinases have other critical functions, referred to as the non-catalytic functions, that drive signalling. Here, we present two examples of the non-catalytic functions of kinases, and how these can contribute to signalling and malignancies, demonstrating the importance of investigating the functions of kinases outside of their enzymatic functions. By no means are the non-catalytic functions highlighted here limited to these kinases, as it is becoming increasingly apparent through kinase knock-out versus kinase inhibition studies, that many kinases contribute to signalling beyond catalysis [14].

Rapidly accelerated fibrosarcoma kinases (RAF) play a critical role in mitogen-activated protein kinase (MAPK) signalling, responsible for cell proliferation and survival. RAF is an active kinase that phosphorylates MEK1/2, which in turn phosphorylates and activates ERK1/2 [15, 16]. However, it has been demonstrated that the kinase dead form of B-RAF can actually activate this pathway, through dimerisation with a wild type (WT) C-RAF kinase molecule. The kinase dead B-RAF induces a conformational change in the WT C-RAF, switching it into its active conformation and thus, stimulating downstream signalling [15, 16]. Through the design of RAF kinase inhibitors to prevent MAPK signalling, it was discovered that when one RAF member of the C-RAF-C-RAF homodimer or B-RAF-C-RAF heterodimer is inhibited, the subsequent signalling is potentiated [17]. This was due to the inhibited RAF molecule activating the other kinase molecule in the dimer, promoting its kinase activity and leading to ERK activation and signalling [17].

Similarly, protein kinase R (PKR) forms a homodimer whereby one monomer activates the other monomer in the dimer [18]. PKR has a critical role responding to cellular and viral stress and is responsible for the phosphorylation of the translation initiation factor eIF2α. Using yeast cell growth as a readout of functional PKR signalling, it was demonstrated that the kinase dead variant of PKR still promoted yeast growth, although it could not catalyse phosphotransfer.

Through dimerisation of the kinase dead PKR with an active PKR molecule, the WT PKR was held in its active state due to the conformational restraints of the dimer, thus potentiating its activity and stimulating PKR signalling and yeast growth [18].

These examples illustrate the non-catalytic functions of the kinase domain and the importance of the kinase fold for signalling, rather than just as catalysts. Thus, it is unsurprising that there is a category of kinases that have no catalytic activity, yet maintain the canonical kinase fold and have critical roles in cell signalling.

1.3 Pseudokinases, the catalytically dead kinases

Approximately 10% of all kinases have no catalytic activity; these are termed pseudokinases [19]. Pseudokinases have a mutation in one or more of the critical catalytic residues, disrupting the R- and C- spines and rendering them catalytically inactive (shown in the VRK3 structure in Fig. 1.1b) [10]. Currently, it is unknown whether pseudokinases diverged from functional kinases during evolution, or whether they underwent co-evolution alongside kinases. However, although pseudokinases lack catalytic activity, they are still vital members of signalling cascades, and more importantly contribute to a multitude of malignancies, offering novel drug targets. Broadly, pseudokinases are classified into two groups: the allosteric regulators and the molecular scaffolds. Allosteric regulators utilise their pseudokinase fold and ability to undergo conformational changes to modulate the activity of active enzymes [20]. Comparatively, molecular scaffolds are more rigid and often unable to bind ATP, demonstrating a reduced conformational flexibility. They instead use their pseudokinase domain and associated protein interaction modules to act as a signalling hub, promoting interaction and association of signalling proteins [21]. Table 1.1 outlines the pseudokinases that fall into these two categories.

[22-39]

Table 1.1 Allosteric regulator and Scaffold Pseudokinases in the Protein Data Bank

ALLOSTERIC REGULATORS				
Pseudokinase	**Function**	**ATP binder?**	**Disease relevance**	**Reference/PDBID**
HER3	Forms asymmetric dimers with HER2/4 and EGFR to activate their kinase activity	Yes	Her3 mutations that stabilise its interaction with EGFR and promote EGFR activity are associated with cancer	Littlefield et al. 2014; 4RIW
STRADα	Activates LKB1 kinase through formation of a trimeric complex with MO25	Yes	Truncation of STRADα leads to PMSE through its inability to activate LKB1	Zeqiraj et al. 2009, Zeqiraj et al. 2009, 3GNI, 2WTK
JH2 (of JAK family)	Regulates JAK kinase (JH1 domain) activity	Yes JAK1/2, No TYK2	JH2 mutations can dysregulate JH1 domains causing haematological malignancies through aberrant JAK signalling	Bandaranayake et al. 2012, Min et al. 2015, Toms et al. 2013, Lupardus et al. 2014; 5C01 (TYK2), 4FVP (JAK2), 4L00 (JAK1), 4OLI (TYK2)
FAM20A	Activates FAM20C kinase, which can then phosphorylate proteins secreted from the cell	Yes	FAM20A mutation leads to a dental condition through the inability to activate FAM20C	Cui et al. 2017, Zhang et al. 2018; 5WRR, 5YH2
MLKL	Key effector of necroptosis, initiates cell death	Yes	Dysregulation of MLKL signalling can lead to inflammatory diseases	Murphy et al. 2013, Murphy et al. 2014; 4BTF, 4MWI
SCAFFOLDS				
Pseudokinase	**Function**	**ATP binder?**	**Disease relevance**	**Reference/PDBID**
KSR1/2	Localizes components of the RAS/ERK/MAPK pathway to promote signal transduction	Yes	Aberrant signalling in the RAS/ERK/MAPK pathway involved in many cancers	Brennan et al. 2011, Dhawan et al. 2016; 2Y4I, 5KKR
ILK	Forms heterotrimeric complex with PINCH and Parvin to regulate focal adhesion signalling	Yes	Overexpression of ILK shown in many cancers; increases cell migration and invasion	Fukuda et al. 2009; 3KMU
Tribbles1	Regulates levels of C/EBP transcription factors through recruitment to the COP-1 ubiquitin ligase for proteasomal degradation	No	Overexpressed in acute myeloid leukaemia (AML)	Murphy et al. 2015; 5CEK, 5CEM
STK40	Binds COP-1 E3 ubiquitin ligase, to either regulate ligase activity or as an adaptor to determine substrate specificity	No	Overexpressed in a number of cancers. In triple negative breast cancer (TNBC), depletion of STK40 lead to decreased cell proliferation and colony formation, as well as apoptosis	Durzynska et al. 2017, Maubant et al. 2018; 5L2Q
SgK223	Influences cell morphology through interactions with RND2 and Csk	No	Overexpressed in many cancers, increases cell migration and elongation	Patel et al. 2017, Lecointre et al. 2018; 5VE6, 6EWX
SgK269	Phosphorylated by Src family kinases and localises to focal adhesions to promote cell migration signalling	No	Overexpressed in many cancers, increases cell migration and elongation	Ha et al. 2017;.; 6BHC
CASK	Plays a role in neural development and neuronal gene transcription regulation	Yes and No has been reported	Mutations lead to conditions such as cleft palate and mental retardation in humans and knock-out in mice is lethal	Mukherjee et al. 2008, Mukherjee et al. 2010; 3C0G, 3C0H, 3MFR, 3MFS, 3MFT
VRK3	Interacts with VHR phosphatase to promote deactivation of ERK	No	Important Erk regulator, dysregulation of Erk signalling leads to malignancies	Scheef et al. 2009; 2JII

1.3.1 Pseudokinases as allosteric regulators

A number of pseudokinases function as allosteric regulators of active kinases or enzymes, inducing conformational changes to alter their activity and manipulate signalling. Table 1.1 lists these pseudokinases, their roles in cells and the diseases they contribute to. The first section of the Table details allosteric regulators, a subset of which are described in detail in this section.

Human epidermal growth factor receptor 3 (HER3), a pseudokinase and an allosteric regulator, is a member of the HER family of tyrosine kinase receptors. HER3 is one of the best characterised allosteric regulator pseudokinases and contributes to cell proliferation and survival signalling. The HER family members (epidermal growth factor receptor (EGFR), HER2/3/4) must asymmetrically dimerise through their intracellular kinase domains in order to propagate signalling (Fig. 1.2) [35, 40]. Figure 1.2 illustrates an asymmetric dimer of HER3 and EGFR kinase domains. The C-lobe of the HER3 pseudokinase domain forms an interaction interface with the N-lobe of the EGFR kinase domain and induces a conformational change in the position of the αC helix of EGFR thus allowing catalytic residues to align for EGFR kinase activity [35]. As HER3 is the only pseudokinase member of the family, hetero-dimerisation of HER3 with other HER family members always results in the activation of the other HER family kinases [41]. Thus, upon HER3 overexpression or deregulation, EGFR signalling could be promoted through consistent activation of its kinase activity. Additionally, the HER3/EGFR structure demonstrated that oncogenic mutations of HER3 stabilise the asymmetric dimer it forms with EGFR, and in doing so, promotes EGFR activity (Fig. 1.2) [35]. Pseudokinases have indispensable roles in modulating signalling, thus there has been a large effort to design drugs to target HER3 and modulate its ability to interact with other HER family members. This would provide a novel way to manipulate, in this case, EGFR signalling and treat EGFR associated cancers [42, 43].

STE20-related kinase adapter protein alpha (STRADα), an allosteric activator pseudokinase, trimerises with adaptor mouse protein 25 (MO25) and liver kinase B1 (LKB1) to activate LKB1 kinase activity (Fig. 1.3) [33, 34]. The interaction of the C-lobe αG helix of STRADα with the LKB1 kinase domain N-lobe switches LKB1 into an active conformation through alignment of the LKB1 catalytic motifs (Fig. 1.3) [33]. The epileptic syndrome polyhydramnios megalencephaly symptomatic epilepsy (PMSE) is a condition whereby the C-lobe of the pseudokinase domain of STRADα is largely missing. Thus, STRADα is unable to interact with and activate LKB1, a negative regulator of the mammalian target of rapamycin (mTOR) pathway. This leads to increased mTOR signalling and increased cell growth, survival and proliferation signalling [34].

Another allosteric regulator pseudokinase is the Janus homology domain 2 (JH2) pseudokinase domain, within the JAK family, that plays a critical role in the regulation of the JH1 JAK kinase domain activity. A mutation of the JH2 domain, V617F (residue number from JAK2), is known to cause hematological malignancies. This mutation hyper-stabilises the interaction of the JH2 pseudokinase domain with the JH1 kinase domain, leading to constitutive kinase activity and thus, constitutive JAK signalling [28]. Similarly, family with sequence similarity 20 (FAM20A) pseudokinase allosterically increases the activity of its functional kinase family member, FAM20C. Mutation of FAM20A prevents its interaction with FAM20C, decreasing the catalytic activity of FAM20C and causing defective tooth enamel formation, a condition called Amelogenesis Imperfecta [24].

These examples highlight the importance of allosteric activator pseudokinases in signalling pathways and how they can contribute to diseases, further highlighting the potential of pseudokinases as drug targets.

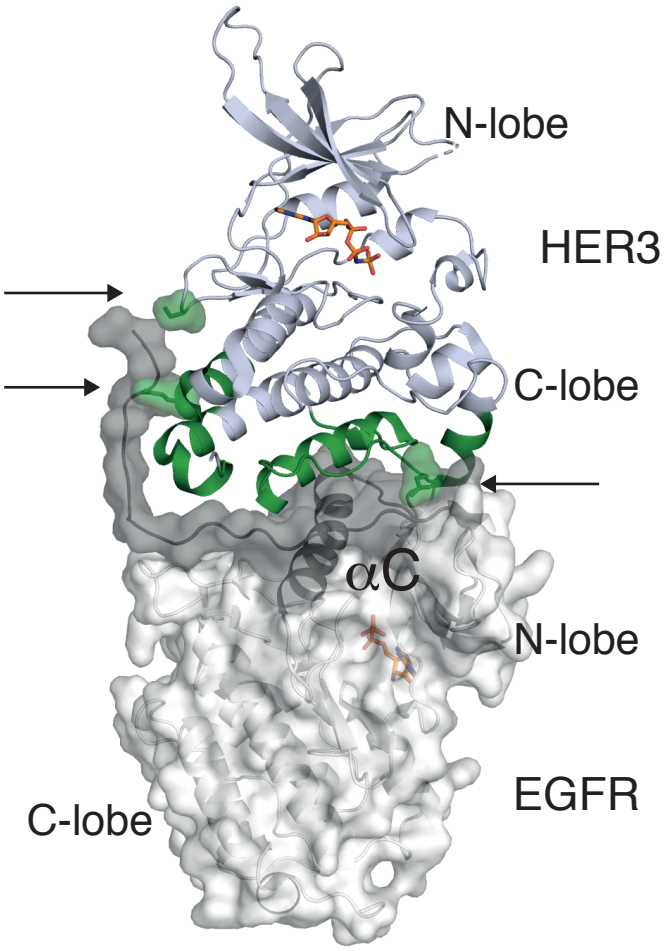
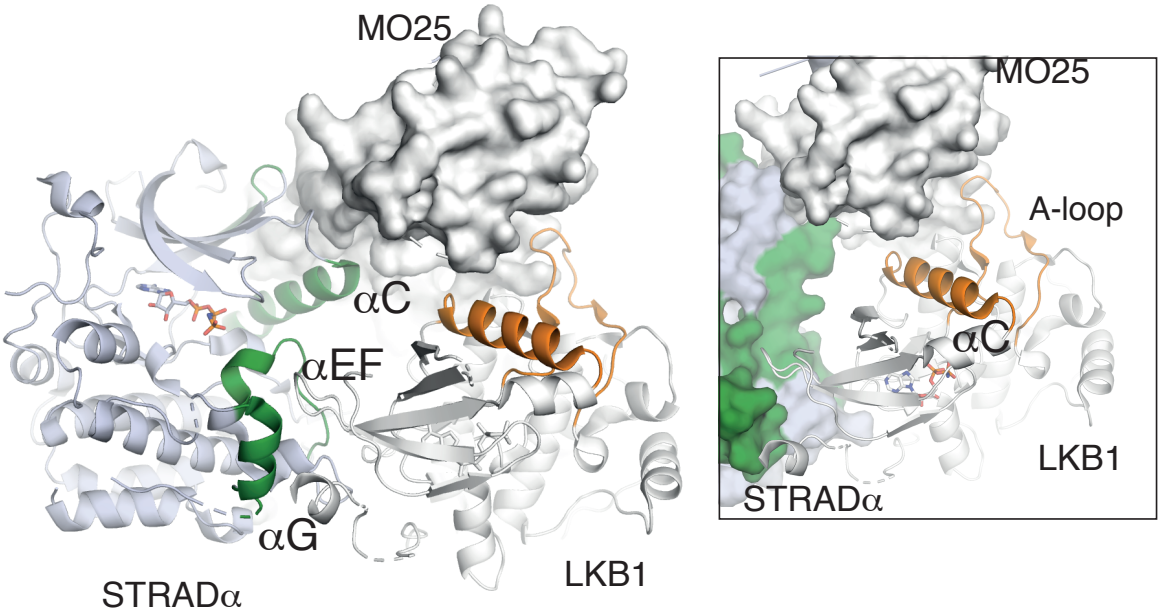


Figure 1.2. Allosteric activator pseudokinase HER3.
Structure of HER3 pseudokinase domain (pale blue) dimerised with EGFR kinase (white). Interaction interfaces of HER3 are highlighted in green. Arrows indicate regions of cancer-associated mutations in HER3 that enhance EGFR activity.



STRADα and MO25 and LKB1 (PDB:2WTK)

Figure 1.3. Allosteric activator pseudokinase STRADα.
STRADα (pale blue) in complex with MO25 and LKB1 (white), activating the LKB1 kinase domain. Interaction interfaces of STRADα with MO25 and LKB1 are highlighted in green. Inset of LKB1 kinase domain in the activate conformation with the A-loop out and the αC helix (orange) correctly positioned for catalysis.

1.3.2 Pseudokinases as scaffolds

A number of pseudokinases have reduced conformational flexibility, often due to an occluded ATP binding site, and therefore, participate in the assembly of multi-protein complexes as scaffolding hubs to influence signalling [21, 44]. The role of scaffolding proteins in cells spans from localising pathway components to coordinating an assembly line of proteins for efficient signalling output [45]. Scaffolds can be large with numerous domains, often containing protein-protein interaction modules in addition to the pseudokinase domain. Examples of protein interaction domains are the Src-homology 2 (SH2) or the phosphotyrosine binding (PTB) domains, enabling interaction with phosphorylated tyrosine residues, to facilitate the association of scaffolding proteins with their interaction partners [19]. One example of a scaffold pseudokinase with auxiliary protein interaction domains is CaM-activated serine-threonine kinase (CASK). CASK has been demonstrated to be critical for neural development and mutations to CASK can lead to mental retardation and cleft palate (Table 1.1) [46-48]. CASK contains an N-terminal CaM pseudokinase domain, followed by a PDZ domain, a central SH3 domain and a C-terminal inactive guanylate kinase domain [46, 47]. The PDZ domain interacts with the cytoplasmic region of cell surface proteins, localising CASK to neural synapses, the SH3 domain interacts with calcium ion channels and the CaM pseudokinase domain interacts with calcium/calmodulin-dependent protein kinase II (CaMKII) to modulate its activity. These domains are necessary for the scaffolding function of CASK, facilitating its protein-protein interactions and thus, role in signalling pathways.

Integrin-linked kinase (ILK) is a scaffolding pseudokinase that forms a complex with the adapter protein particularly interesting new Cys-His protein 1 (PINCH-1), through its N-terminal Ankyrin repeat domains, and α-parvin, through the αG and αEF helices of ILK pseudokinase domain (Fig. 1.4a) [27]. They trimerise and migrate to focal adhesions, the contact point between the extracellular matrix (ECM) and the cell, recruited by focal adhesion associating proteins such as paxillin [49]. Here, the pseudokinase domain of ILK can interact with the focal

adhesion receptors, integrins, and the adaptor paxillin. In addition, α-parvin and PINCH-1 interact with actin. The scaffolding role of ILK acts to link integrins to the cytoskeleton, facilitating cell migration in response to environmental cues [27]. Upon mutation of residues in the αG helix of ILK, the interaction between ILK and α-Parvin is destabilised and ILK is unable to localise to focal adhesions. This impairs focal adhesion signalling, highlighting the importance of the αG as a structural feature that mediates their interaction and the scaffolding function of ILK (Fig. 1.4a) [27]. Additionally, ILK is implicated in numerous cancers. Specifically, ILK is overexpressed in melanoma leading to increased disease progression and increased lymph node invasion [25]. Through its role as a scaffold that links actin to the ECM, overexpression of ILK increases cell migration and enhances cell survival pathways, contributing to malignancies [25].

Kinase suppressor of Ras (KSR) is an important scaffold in the MAPK signalling pathway. KSR is composed of five domains, an N-terminal 40-amino acid region unique to KSR, followed by a proline-rich region, a cysteine-rich zinc finger domain, a serine/threonine-rich region and a C-terminal pseudokinase domain. KSR facilitates its scaffolding function through its numerous domains, for example the zinc finger domain is critical for its localisation and serine/threonine-rich region acts as a binding site for MAPK [50]. Additionally, the pseudokinase domain of KSR interacts with Rapidly Accelerated Fibrosarcoma (RAF) kinase and mitogen activated protein kinase kinase (MEK), both key kinases in the MAPK signalling pathway [51]. Through 'face-to-face' interactions, whereby the αG and A-loop from KSR interact with the αG and A-loop from either RAF or MEK, KSR stabilises the active conformation of RAF or MEK, enhancing their activity (Fig. 1.4b). Thus, each domain of KSR plays a critical role in the scaffolding function of KSR, enabling its role in enhancing MAPK signalling.

This thesis focuses on two scaffolding pseudokinases, SgK223 and SgK269 that belong to the pseudopodium-enriched atypical kinase (PEAK) family and will be discussed in the following sections.

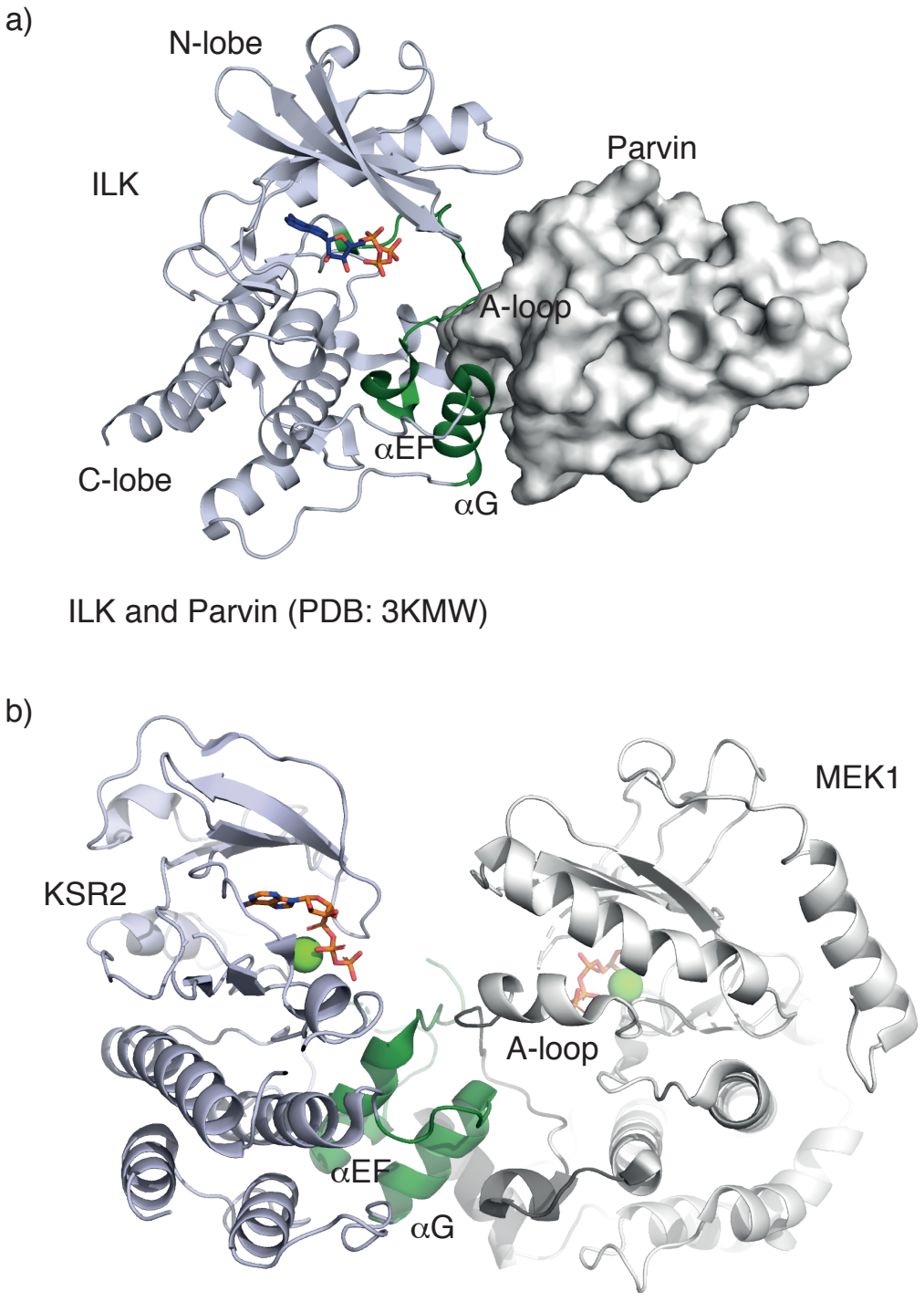


Figure 1.4. Structures of scaffolding pseudokinases ILK and KSR.
a) Structure of ILK pseudokinase domain (pale blue) in complex with one of its interacting partners α-Parvin (white surface representation). Interaction interfaces of ILK are highlighted in green.
b) Structure of KSR2 pseudokinase domain (pale blue) in complex with the active kinase MEK1 (white). Interaction interfaces of KSR2 are highlighted in green.

1.4 The PEAK family of pseudokinases

The PEAK family of scaffolding pseudokinases, SgK223, SgK269, and PEAK3, have been shown to be important regulators of signalling that controls cell migration and cell morphology signalling [1, 6, 52]. SgK223 and SgK269 share a similar domain layout characterised by an N-terminal region of unknown structure and function, a PEST linker, and a pseudokinase domain flanked by an N-terminal helix (αN1) and four C-terminal helices (αJ, αK, αL, αM) (Fig. 1.5) [1]. The PEST linker has multiple tyrosine residues that are phosphorylated and subsequently serve as specific interaction points for adaptors and signalling proteins with SH2 and PTB domains [1, 2, 5, 6, 53-60]. Additionally, SgK223 and SgK269 have been demonstrated to undergo dimerisation and oligomerisation. This is facilitated by the C-terminal end of SgK223 and SgK269, specifically the pseudokinase domain and helices that flank it. The association of these two proteins has been demonstrated to be critical for their role in promoting cell migration and causing elongated cell morphology [1]. It has been suggested the dimerisation of these proteins localises multiple different interactors and adaptors of SgK223 and SgK269 in close proximity, allowing for their interaction and subsequent signalling [6].

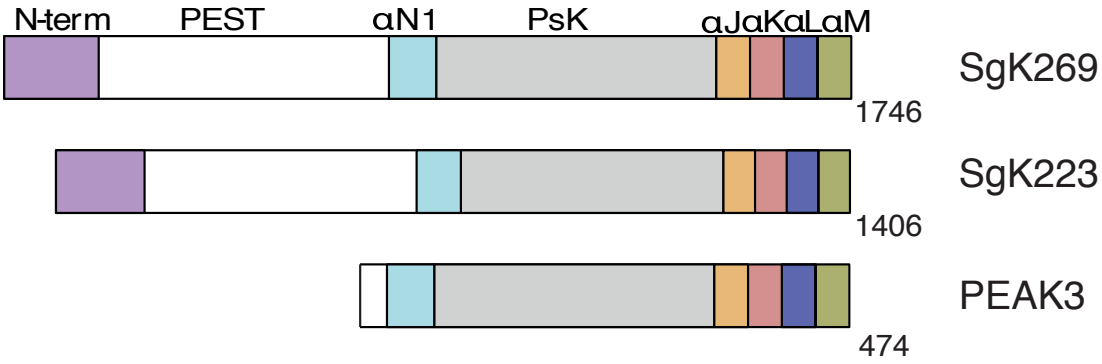
Recently, the newest member of the family, PEAK3, was identified. PEAK3 is significantly shorter than SgK223 and SgK269 containing a shorter N-terminal region (Fig. 1.5) [52]. Much like SgK223 and SgK269, PEAK3 has a pseudokinase domain flanked N- and C-terminally by helices with sequence similarity to those that form the dimerisation interface of SgK223 and SgK269 (Fig. 1.5b) [52]. In contrast to SgK223 and SgK269, whose overexpression in cells leads to cell elongation, PEAK3 overexpression leads to a decrease in the number of actin filaments in cells [52]. Additionally, recent research demonstrated that PEAK3 can interact with the adaptor protein CrkII, and in doing so prevents the ability of CrkII to promote membrane ruffling and elongation [52]. Thus, the role of PEAK3 in cells is seemingly different to that of SgK223 and SgK269.

1.4.1 The recently solved structures of SgK223 and SgK269

Recently the structures of the pseudokinase domain and the flanking helices of SgK223 (PDB: 5VE6 (human), 6EWX (rat)) and SgK269 (PDB: 6BHC) were solved as dimers. The mode of dimerisation in all these structures was conserved, mediated by the five helices that flank the pseudokinase domain, with a large buried surface area (BSA) (Fig. 1.6a) [6-8]. Within a monomer, the three structures showed a conserved canonical kinase fold, consisting of a N-lobe of primarily β-sheets and a larger C-lobe of α-helices (Fig.1.6a). The αC helix is unresolved in the rat and human SgK223 structures, however in the SgK269 structure a part of the αC helix, which is stabilised by a salt bridge between Arg1378 of the αC helix and Asp1390 of the β4 strand is resolved (Fig. 1.6a). SgK223 and SgK269 have substitutions to the critical glutamate that forms a salt bridge with the lysine of the VAIK motif to coordinate the αC helix in the correct orientation for catalysis. Thus, these pseudokinases are unable to form this stabilising interaction and cannot correctly position the αC helix [6-8]. The conformation of the catalytic loop HRD motif is similar to that of a *bona fide* kinase in the SgK223 structures, however in SgK269 the arginine is substituted with a cysteine. The DFG motif of a canonical kinase, at the start of the A-loop, is SNF and NFS in SgK223 and SgK269, respectively [6-8]. The A-loop is unresolved in the SgK223 structures, however in the SgK269 structure a small portion of A-loop is resolved (Fig. 1.6a). The A-loop is stabilised by a salt bridge between Arg1576 of the A-loop and Glu1601 of the αF helix of the SgK269 pseudokinase domain (Fig. 1.6a) [6, 7]. Interestingly, these residues are conserved in SgK223, indicating the SgK223 A-loop can also adopt this conformation and that the A-loop of these pseudokinases are conformationally flexible and affected by crystal packing. It was noted in the SgK269 structure the residues within the A-loop have high *B*-factors, suggesting this region is flexible [7]. The inherent flexibility of the A-loop could thus be a regulatory mechanism for the association of SgK223 and SgK269 and other interaction partners, similar to the A-loop regulatory mechanism in *bona fide* kinases. In addition, the structures confirm that SgK223 and SgK269 are pseudokinases, as both have misaligned R- and C-spines,

compared to the canonical PKA (Fig. 1.1a), and an occluded ATP binding site, discussed in detail in the next section (Fig. 1.6b) [6-8].

a)



b)

```
SgK223 ----------------------------------------------------------GSTQLQ 6
SgK269 ------------------------------------------------------------MQ 2
PEAK3  KKILTRTQSLPTRRTLHPSSIQVQPPRRPFLGSHSVDKSQAAVGPACLPAELTFGPADAP 120

                          αN1
SgK223 LHGLLSNISSKEGTYAKLGGLYTQSLARLVAKCEDLFMGGQKKELHFNENNWSLFKLTCN 66
SgK269 RQALYRGLENREEVVGKIRSLHTDALKKLAVKCEDLFMAGQKDQLRFGVDSWSDFRLTSD 62
PEAK3  LGLSLRDLHSPEAVHTALAARQLQGLRTIYARLRARLMGGHPGPCHPG----HSFRLLDS 176
          .: . * .   : .   :.*  : .: .  :*.*:    : .      *:*  .
          β1    β2        β3     αC                              
SgK223 KPCCDSGDAIYYCATCSEDPGSTYAVKICKAPEPK-TVSYC-----SPSVPVHFNIQQDC 120
SgK269 KPCCEAGDAVYYTASYAKDPLNNYAVKICKSKAKE-SQQYYHSLAVRQSLAVHFNIQQDC 121
PEAK3  SPCAESGDALYYRVVRAHEDAW--HILVAKVPKPGADVPHPWGLELQASLSPHFNLQGLC 234
       .**.::***:** .  :.:      : :.*       :       *:  ***:*  *
          β4                          β5     β5D   αD
SgK223 GHFVASVPSSMLSSPDAPKDPVPALPTHPPAQEQDCVVVITREVPHQTASDFVRDSAASH 180
SgK269 GHFLAEVPNRLLPWEDP------------SKKQRSHVVVITREVPCLTVADFVRDSLAQH 169
PEAK3  GLVPEGT----LP-----------------GAPWRGAVALAAEVPERTVAQWLAEAC--- 270
       * .   .    *                   .*.:: ***  *.:::: ::
          αE                    β7
SgK223 QAEPEAYERRVCFLLLQLCNGLEHLKEHGILHRDLCLENLLLVHCTLQAGPGPAPAPAPA 240
SgK269 GKSPDLYERQVCLLLLQLCSGLEHLKPYHVTHCDLRLENLLLVHYQPGGTAQ-------- 221
PEAK3  TQPPEEFVWAVALLLLQLSAALKFLEAWGAALVELRPENLLLVAPRGCATTG-------- 322
          *: :   *.:*****. .*:.*:     :*  ******      .
                                Catalytic loop  β8   A-loop
SgK223 PAAAAPPCSSAAPPAGGTLSPAAGPASPEGPREKQLPRLIISNFLKAKQKPGGTPNLQQK 300
SgK269 ---------------------GFGPAEP-TPTSSYPTRLIVSNFSQAKQKSHL-VDPEIL 258
PEAK3  ------------------------------------PPRLLLTDFGRVCLQPPGPPGSPG- 346
                                          **:::* :.  :      .
                   αF                      αG
SgK223 KSQARLAPEIVSASQYRKFDEFQTGILIYELLHQPNPFEVRAQLRERDYRQEDLPPLPAL 360
SgK269 RDQSRLAPEIITATQYKKCDEFQTGILIYEMLHLPNPFDENPELKEREYTRADLPRIPFR 318
PEAK3  ----PHA-------------PQLGSLLRALLSLAA----------------------PST 367
          *              * * *:  :*                          *
          αH          αI          αJ       αK
SgK223 SLYSPGLQQLAHLLLEADPIKRIRIGEAKRVLQCLLWGPRRELVQQPGTSEE--ALCGTL 418
SgK269 SPYSRGLQQLASCLLNPNPSERILISDAKGILQCLLWGPREDLFQTFTACPSLVQRNTLL 378
PEAK3  TPLAAGLELLAAQLTRLR----PSASRTRGALQALLWGPGPELRGRGA---PLGPWLRAL 420
       :  : **: **  * .         . ::  **.*****  :*              *
          αL       αM
SgK223 HNWIDMKRALMMMKFAEKAVDRRRGVELEDWLCCQYLASAEPGALLQSLKLLQLL 473
SgK269 QNWLDIKRTLLMIKFAEKSLDREGGISLEDWLCAQYLAFATTDSLSCIVKILL-- 431
PEAK3  GPWLRVRRGLLVLRLAERAAGG-EAPSLEDWLCCEYLAEATESSMGQALALLWD- 473
         *: ::* *::::::*:: .   . .******.:*** *  .::   : :*
```

Figure 1.5. Domain layout and sequence alignment of SgK223, SgK269 and PEAK3.

a) Domain layout of SgK269, SgK223 and PEAK3 with the important regions mapped and labelled.

b) Sequence alignment of SgK223, SgK269 and PEAK3 based on sequences used for the crystal structures of SgK223 and SgK269 proteins and the corresponding region in PEAK3. Structural motifs are shown above the alignment and labelled. Helices appear as cylinders and β-strands appear as arrows.

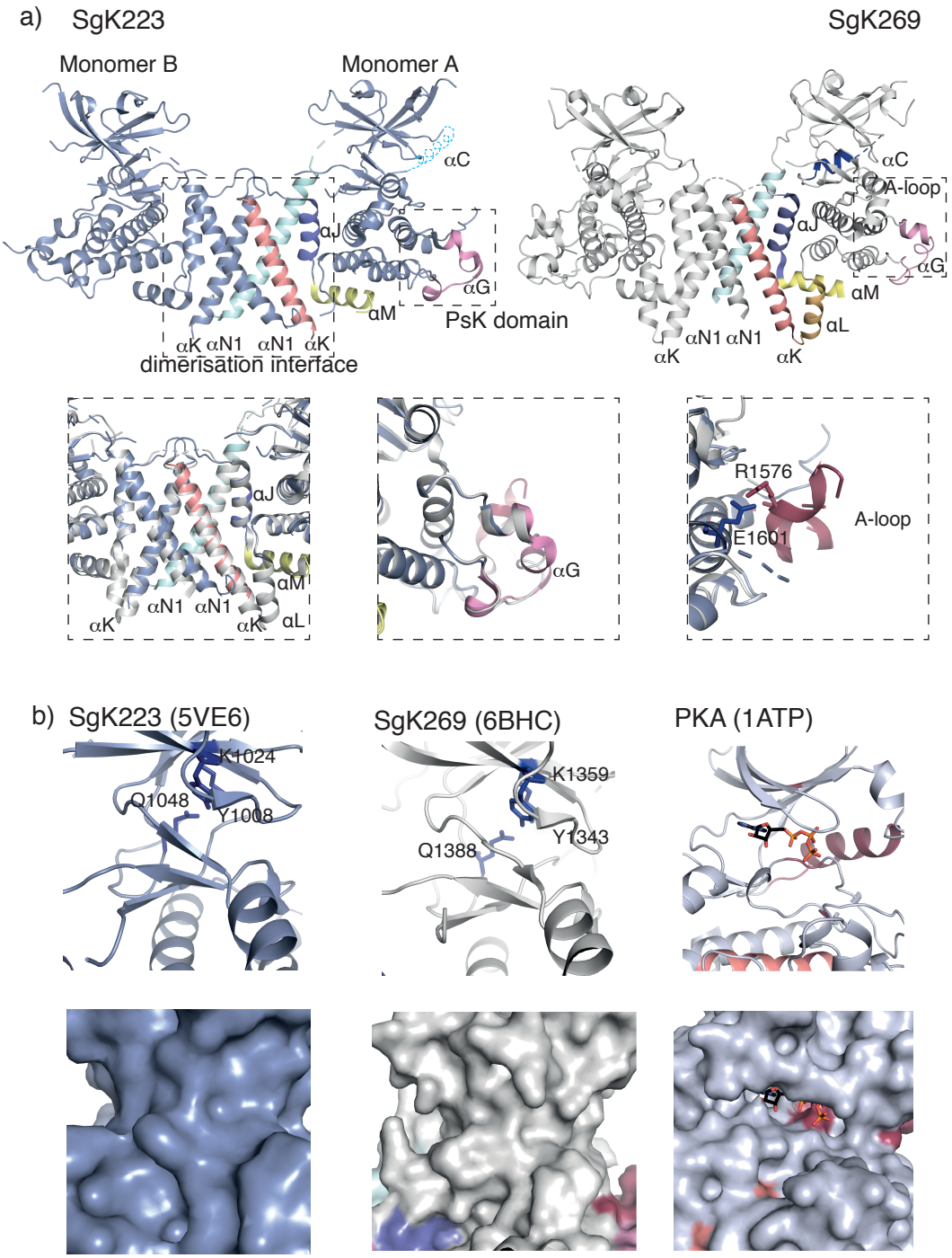


Figure 1.6. Structures of SgK223 human and SgK269.
a) SgK223 (PDB: 5VE6) left and SgK269 (PDB: 6BHC) right. The inset shows the overlay of SgK223 and SgK269 pseudokinase domains at the A-loop region. The A-loop is stabilised by a salt bridge formed between the residues Arg1576 and Glu1601 in SgK269 and is unresolved in SgK223. SgK223 in blue; SgK269 in white; the dimerisation interface helices αN1 in blue, αJ in pink, αK in purple and αL in yellow; αC helix in blue; A-loop in dark red.
b) ATP binding sites of SgK223, SgK269 and PKA (PDB: 1ATP) illustrating that SgK223 and SgK269 have occluded ATP binding sites and cannot bind ATP, compared to PKA, a *bona fide* kinase that binds ATP.

1.4.2 The ATP binding site of SgK223 and SgK269 is occluded

The structures of SgK223 and SgK269 demonstrated they both have an occluded ATP binding pocket, which is filled with the sidechains of bulky residues due to the reorientation of the N-lobe in respect to the C-lobe to accommodate the dimerisation domain (Fig. 1.6b) [6-8]. In addition, residues that are occupying the ATP binding pocket are conserved in PEAK3, suggesting it most likely cannot bind ATP [52]. The conservation of residues in the ATP binding pockets could suggest that the PEAK family evolved the inability to bind ATP through the same selection pressures, possibly through the dimerisation interface that forces conformational restraints on the pseudokinase domain.

1.5 SgK223 and SgK269 as scaffolds in signalling and cancer

1.5.1 SgK223

SgK223 is a signalling hub that interacts with multiple different pathways dependent on the specific interaction partner it associates with. SgK223 was first discovered as an interactor of Rho family GTPase 2 (RND2), a member of the Rho family of small GTPases in neuronal cells. This study showed that RND2 interacts with the pseudokinase domain of SgK223 (Fig. 1.7a) [60]. It was demonstrated that ectopic expression of SgK223 in HeLa cells increased cell contraction and upon co-expression of SgK223 and RND2 cell contraction was substantially increased [60]. In addition, changes to cell morphology induced by SgK223 and RND2 overexpression is reliant on Rho-kinase/RhoA signalling, as when this signalling is inhibited, SgK223 overexpression cannot induce cell contraction. PC12 cells, a neuronal model cell line, undergo neurite outgrowth upon treatment with nerve growth factor (NGF), extending long processes from their cell body. Co-expression of RND2 and SgK223 lead to a decrease in neurite outgrowth in response to NGF and an altered cell morphology to a rounded phenotype [60]. Using truncation variants of SgK223, it was found that only the construct lacking the C-terminal region (residues 1-829) was capable of suppressing neurite outgrowth and inducing cell contractility in PC12 cells and

stimulating RhoA activity [60]. This phenotype was not seen upon expression of the N-terminal truncation variant of SgK223, lacking the first ~800 residues. Thus, the first ~800 residues of SgK223 are responsible for the changes in cell morphology induced by SgK223 expression in this cell model [60]. Localisation studies revealed that SgK223 was diffusely distributed across the cytoplasm when expressed alone in HeLa cells, and upon co-expression with RND2, SgK223 localised to regions in the cell periphery [60]. These data demonstrated that the interaction of SgK223 and RND2 is critical for modulating neuronal cell morphology and SgK223 acts as a downstream effector of Rho-kinase/RhoA signalling, through direct interactions with RND2 and RhoA, enhancing cell contractility.

SgK223 was also shown to interact with the SH2 domain of C-terminal Src Kinase (Csk), a negative regulator of Src kinase activity [5, 58]. Tyr411 of SgK223 PEST linker is phosphorylated by Src family kinases (SFKs), allowing it to interact specifically with the SH2 domain of Csk. The SgK223-Csk interaction sequesters Csk from Src to potentiate Src activity (Fig. 1.7a) [5, 58]. SgK223 is one of the few mammalian proteins that has an EPIYA (Glu-Pro-Ile-Tyr-Ala) motif, which includes Tyr411. Many bacterial species have effector proteins containing the EPIYA motif to subvert host cell signalling [58]. It is thought that the EPIYA motif of bacterial effectors are promiscuous, capable of interacting with multiple different SH2 domains, whereas the SgK223 EPIYA motif interacts specifically with that of Csk [58]. SgK223 co-localisation studies show that it localises to focal adhesions with Csk and vinculin, a focal adhesion associating protein, however SgK223 is primarily distributed to the cytoplasm [5]. Upon ectopic co-expression of SgK223 and Csk, MKN7 gastric cells had an altered cell morphology compared to the WT cells. The cells overexpressing SgK223 and Csk demonstrated increased cell elongation. This was not evident upon co-expression of Csk and SgK223 Y411F mutant, which was unable to be phosphorylated. The cellular morphology change was not as substantial upon single ectopic expression of SgK223 or Csk, compared to co-expression of the two proteins [5]. The interaction of SgK223 and Csk is able to induce morphological changes in cells,

suggesting a role of these two proteins in cell migration and morphology signalling. Consistent with this, overexpression of SgK223 in immortalised mammary epithelial cells (MCF-10A) leads to cell elongation and an increased rate of cell migration [1, 6]. This is in contrast to the cell contractility induced through interaction of SgK223 with RND2, highlighting that the morphological outcomes for the cell differ dependent on the SgK223 interaction partner, the signalling pathway SgK223 associates with, and the tissue type SgK223 is expressed in.

SgK223 has been implicated in a number of cancers. In SW620 colon cancer cells that overexpress Src, expression of SgK223 increases their invasive potential and upon knockdown of SgK223, this increase was negated [61]. In addition, knockdown of SgK223 in non-small cell lung cancer (NSCLC) cells reduces cell migration, invasion and proliferation [4, 57]. SgK223 is overexpressed in pancreatic ductal adenocarcinoma (PDAC), NSCLC and esophageal adenocarcinoma compared to normal tissue, and is associated with poor prognosis in NSCLC [4, 62, 63]. Additionally, the invasive phenotype seen in human pancreatic duct epithelial cells (HPDE) upon SgK223 overexpression, was decreased when STAT3 was inhibited. As STAT3 activation is required for cell migration and proliferation, SgK223 could exert its influence on cell migration through regulation of STAT3 signalling [59].

SgK223 has been reported to interact with a number of adaptor proteins at focal adhesions. Adaptor proteins are a critical element of signalling pathways that through their multiple interaction domains, they can facilitate the associations of many different signalling components. One such adaptor that SgK223 has been reported to interact with is vinculin, a 117 kDa protein composed of globular N-terminal head domain linked to the tail domain by a proline-rich region. Through immunoprecipitation and co-localisation studies, it was proposed that vinculin is a driver of SgK223 localisation to focal adhesions [58]. Additionally, SgK223 was implicated in the Bcr-Abl signalling network, proposed to interact with Csk SH2 domain (as previously mentioned), C10 regulator of kinase I (CrkI) SH2 domain

and Alpha-Actinin-1 [64]. There are a number of motifs in the PEST linker of SgK223 that are yet to be characterised. Namely, a number or tyrosine residues that upon phosphorylation may interact with SH2 domain containing proteins and proline-rich regions, that could interact with SH3 domains. Chapter 4 of this Thesis will investigate the interactions between the adaptor protein CrkII and the PEAK family of proteins, characterising them structurally and biophysically.

1.5.2 SgK269

The closely-related homologue of SgK223, SgK269, is also a scaffolding protein important for the nucleation of many signalling pathways. To date, three important tyrosine residues within the PEST linker region of SgK269 have been characterised (Fig. 1.7b). Tyr635 is phosphorylated by Lyn SFK and serves as a docking site for the SH2 domain of Grb2, an adaptor protein for the Ras/Erk signalling pathways. Interaction of Grb2 and SgK269 enhanced Ras/Erk signalling, leading to increased cell migration and invasion [54]. Mutation of Tyr635 to Phe635 abrogated the increased cell invasion seen when SgK269 is overexpressed in MCF-10A cells, highlighting the importance of phosphorylation of Tyr635 for SgK269 induced cell motility [54].

In another study, phosphorylation of Tyr665 by SFKs was shown to localise SgK269 to focal adhesions (Fig. 1.7b) [2]. Disruption of Tyr665 phosphorylation decreased cell migration, focal adhesion assembly and reduced SgK269 co-localisation with paxillin, suggesting this post-translational modification is critical for the role of SgK269 at focal adhesions [2]. Inhibition of Src through a small molecule inhibitor or small interfering RNA mediated knockdown diminished SgK269 induced cell motility and co-localisation of SgK269 and paxillin at focal adhesions [2]. Thus, the function of SgK269 in focal adhesion assembly and cell migration is dependent on Src phosphorylation of Tyr665 [2].

SgK269 has been implicated in epidermal growth factor (EGF) signalling, with Zheng *et al*. demonstrating that at later stages of signalling, Tyr1188 of SgK269

is phosphorylated and this becomes a site for Shc1 PTB interaction (Fig. 1.7b) [56]. At early EGF signalling, Grb2 interacts with Shc1 PTB domain and this switch in EGF signalling to SgK269 as the scaffold in this pathway is required to link Shc1 to other signaling proteins, such as phosphatases and GTPase activating proteins (GAP) [56]. The SgK269-Shc1 interaction promotes cell migration and invasion pathways and demonstrates the scaffolding capacity of SgK269 to modulate signalling pathways [56].

Similar to SgK223, overexpression of SgK269 in MCF-10A cells increases cell elongation and migration, hallmarks of a malignant phenotype [1, 6]. Breast cancer cells with high expression of SgK269 had elevated expression of markers of epithelial-to-mesenchymal transition (EMT) and displayed a mesenchymal morphology [53]. In pancreatic ductal adenocarcinoma (PDAC), mutations of the *KRas* oncogene lead to an overexpression of SgK269. Higher levels of SgK269 lead to increased Src kinase activity, reportedly through the interaction of the SH2 domain of Src with phosphorylated Tyr665 of SgK269. This interaction stabilised the Src kinase active conformation, thus promoting its signalling and aberrant cell behaviour [55].

Although the interaction partners of SgK269 have been investigated in greater detail than SgK223, there are still a number of interaction motifs and residues that are yet to be directly characterised. Specifically, structural, biochemical and biophysical insights into these interactions, to verify that they are direct interactions, is lacking. We begin to address this gap in the literature in Chapter 4, however more research is required to understand which proteins SgK269 and SgK223 are directly interacting with, and which are indirect interactions, mediated by adaptors such as Grb2 and CrkII.

1.6 The homo- and hetero- association of the PEAK family

Recently, it was shown that SgK223 and SgK269 homo-and hetero-associate. They both form a symmetric dimer, whereby the αN1 and αK helices arrange into a four-helix bundle and the αJ, αL and αM helices form a junction between the

pseudokinase and dimerisation domain (Fig. 1.6a) [6-8]. There is 45% sequence identity between SgK223 and SgK269 pseudokinase and dimerisation domains (Fig. 1.5b) [1]. This is reflected in the striking similarities between the structures when they are aligned, with high conservation of key structural motifs (Fig. 1.6a). Additionally, the third member of the PEAK family, PEAK3, has similarity in the key dimerisation domain helices, predicted to also undergo dimerisation (Fig. 1.5b). Mutational analysis has been carried out on SgK223 and SgK269 to reveal hot spots that drive their homo-dimerisation (see Chapter 3 for further details) [6-8]. The mutants that disrupted SgK223 dimerisation were found to prevent the increase in cell migration caused by overexpression of SgK223 in the MCF-10A human cell line [6]. Additionally, these mutants failed to promote cell elongation that usually occurs upon SgK223 overexpression [6].

Interestingly, through recombinant protein studies, it was also demonstrated that not only are SgK223 and SgK269 capable of homo-dimerisation, they can also hetero-dimerise [6]. Hetero-dimerisation of SgK223 and SgK269 demonstrates the multiple modes of interaction they can undergo, however, the importance of homo-and hetero-dimerisation of these scaffolding pseudokinases is yet to be established.

It was previously found that SgK223 and SgK269 pseudokinase domains can homo-oligomerise to form higher molecular weight complexes, as evident through size exclusion chromatography (SEC) experiments [1]. However, until recently the interface driving these interactions was unknown. Analysis of the SgK223 crystal structure revealed an interface consisting of residues from the αG, αEF and end of the A-loop of the C-lobe that pack against the N-lobe of the symmetry related molecule (see Chapter 3 for further detail) [6]. Through mutational analysis, similar to that carried out on the dimerisation interface, residues of the αG, αEF and end of the A-loop that drive oligomerisation of SgK223 were identified. Further *in vitro* experiments, using recombinant SgK223 and SgK269, confirmed that they can hetero-oligomerise [1]. Constructs of either SgK223 and/or SgK269 pseudokinase domain mutated at residues critical for

oligomerisation were incapable of forming higher molecular weight hetero-complexes, demonstrating the importance of this region for their homo- and hetero-association [6].

Further highlighting the importance of their hetero-association, SgK223 was shown to immunoprecipitate with Grb2, an adaptor protein known to associate with SgK269 through Tyr635 [1]. However, in cells lacking SgK269, SgK223 could no longer immunoprecipitate with Grb2, indicating the association of SgK269 with SgK223 mediates the interaction between Grb2 and SgK223 [1]. Furthermore, overexpression of SgK269 in MCF-10A cells lead to increased cell migration, but in the absence of SgK223, SgK269 was unable to increase cell motility to the levels seen when both SgK223 and SgK269 were present [1]. Overexpression of both SgK269 and SgK223 lead to signal transducer and activator of transcription 3 (STAT3) activation, however in cells lacking SgK223, STAT3 activation could not be induced solely by SgK269 expression. Thus, the hetero-association of SgK223 and SgK269 clearly plays a role in their signalling and could be a mechanism for amplification and diversification of signalling outputs.

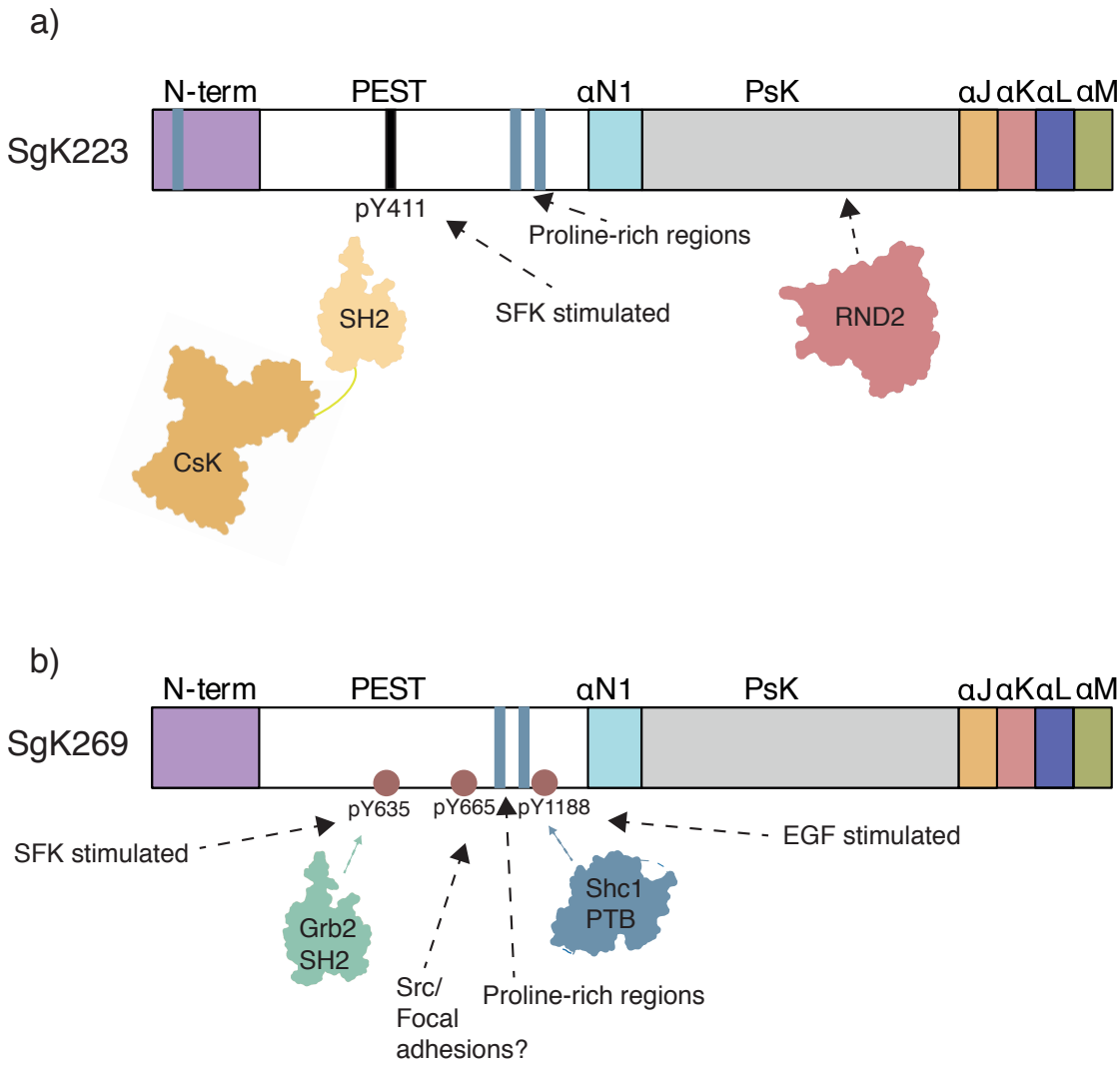


Figure 1.7. Domain layout of SgK223 and SgK269 and with their interacting partners mapped to their interaction sites.
a) SgK223 domain layout, N-terminal domain, PEST linker with Tyr411, pseudokinase domain flanked by regulatory helices. Residue Tyr411 is phosphorylated by Src family kinases and the SH2 domain of Csk interacts with the phosphorylated tyrosine. RND2 was demonstrated to interact with SgK223 pseudokinase domain. There are also a number of proline-rich regions in the PEST region and N-terminal domain (blue lines on the domain).
b) SgK269 domain layout, resembling the SgK223 layout. Three phosphorylation sites are known in the PEST linker, Tyr635 is phosphorylated by Src family kinases and is a docking site for Grb2 SH2 domain, Tyr665 is known to play a role in focal adhesion signalling and Tyr1188 is phosphorylated after EGF stimulation and a docking site for Shc1 PTB domain. There are also a number of proline-rich regions in the PEST region (blue lines on the domain).

1.7 Outstanding research questions

Although a few studies have now been reported investigating the homo- and hetero-dimerisation and oligomerisation of SgK223 and SgK269, there are a number of outstanding questions that still remain to be answered. The oligomeric interface of SgK223 and SgK269 was shown to be driven by the αG and A-loop within the pseudokinase domains. However, the complementary region, based on the SgK223 structure, involves the N-lobe of the pseudokinase domain of the interacting monomer and has not yet been investigated. Additionally, it is unknown whether this same region is involved in oligomerisation of SgK269 or even hetero-oligomerisation between SgK223 and SgK269.

The structures of SgK223 and SgK269 pseudokinase and dimerisation domains have provided insights into the mechanisms of SgK223 and SgK269 self-association. However, the N-terminal domain in these proteins has not been characterised.

In addition, although there have been several PEAK family interacting partners described in the literature, very little structural and biophysical data validating these interactions is currently available. One of the focal points of this Thesis is to begin to verify some of these interactions, specifically to investigate an adaptor protein CrkII and its interactions with the PEAK family. This will be done using binding assays and X-ray crystallography. This will form the basis for understanding the scaffolding function of the PEAK family and how they interact with various adaptor proteins in order to drive signalling pathways.

Previous research has established that dimerisation of SgK223 and SgK269 is important for their role in increasing cell migration and altering cell morphology and that both localise to focal adhesions. However, how disruption of their ability to form homo- or hetero- dimers or oligomers can impact on their localisation is currently unknown. To understand this further, this Thesis will design and develop

new tools and techniques that can be used to investigate these proteins in live or fixed cells using various microscopy based methods.

1.8 Thesis aims

Thus, in this Thesis we aim to:

a) Characterise the N-terminal domains of SgK223 and SgK269 using biochemical and biophysical techniques and characterise the homo- and hetero-association of SgK223 and SgK269 using mutational and biochemical analysis.

b) Structurally and biophysically characterise the interaction of CrkII with the PEAK family proteins.

c) Investigate the role of SgK223 and SgK269 homo- and hetero- association in cells, specifically for their sub-cellular localisation using microscopy techniques.

2.0 Experimental procedures

2.1 Cloning

2.1.1 E.coli expression constructs

The human genes of SgK223, SgK269, CrkII C-SH3 domain and CrkII N-SH3 domain were synthesised by Genscript. The genes were individually cloned using restriction digestion, using BamHI and HindIII restriction sites, into a modified pCOLD IV vector (Takara) containing an N-terminal His6 TEV cleavage tag (provided by Associate Professor Isabelle Lucet) (Fig. 2.1). pCOLD IV expression vectors with the SgK223 and SgK269 N-terminal domains were provided by Associate Professor Isabelle Lucet. All *E.coli* expression vectors contained the ampicillin resistance gene for selection of *E.coli* expressing the correct proteins. See Figure 2.1 and Table 8.1 in the appendices for details of gene sequences, restriction sites and vectors maps.

2.1.2 Imaging constructs

The human genes of SgK269 and SgK223 were synthesised by Genscript with an internal BamHI site, prior the αN1 helix, and a HindIII restriction site at the C-terminus of SgK223 and SgK269, to facilitate sub-cloning of mutants of the pseudokinase domain and flanking helices (Fig. 2.1). To generate the internal BamHI site there was a change in sequence of the SgK223 gene, without altering the codon (TCT to TCC), and a change in the sequence of the SgK269 gene, without altering the codon (GCA to GGA). These genes were sub-cloned through restriction digestion, using an N-terminal AgeI restriction site and C-terminal XhoI restriction site, into a pMIH mammalian expression vector kindly gifted by Dr. Kate McArthur. SgK223 and SgK269 were fused N-terminally to an mNeonGreen (Allele Biotechnology) and a Halo tag (Promega), respectively. Additionally, mApple and mEmerald Paxillin (a gift from Michael Davidson: Addgene plasmid #54935, #54219) were used for imaging, however there were transiently transfected into cells in the plasmid they were provided in, without any additional cloning. All mammalian expression vectors contained the ampicillin resistance gene for selection, with the exception of the paxillin vectors, which contained the

kanamycin resistance gene. See Figure 2.1 and Table 8.1 in the appendices for details of gene sequences, restriction sites and vectors maps.

2.1.3 General cloning procedure

All genes from Genscript were provided in either pJET or pUC plasmids. The genes were excised from the plasmids using the BamHI and HindIII restriction sites for the *E.coli* expression genes, and AgeI and XhoI for the mammalian expression genes. Agarose gel purification followed digestion, whereby the digestion products were run on a 0.5-1% agarose gel at 90 V for 60-90 minutes. The band corresponding to the gene of interest was excised from the gel and purified using the GE gel purification kit (Illustra^{TM} GFX^{TM} PCR DNA and Gel Band Purification Kit, GE Healthcare). The gene of interest was then ligated into a pre-cut vector, for either bacterial or mammalian expression, and incubated with T4 DNA ligase (New England Biolabs) overnight at room temperature, following standard ligation protocol. The ligation reaction mixture was transformed into either DH5α or Top10 *E.coli* strains, and plated onto a 100 μg/mL ampicillin Luria Broth (LB) agar plate. Agar plates were incubated overnight at 37 °C. Following incubation, overnight ligation cultures were picked and 5 mL LB with 100 μg/mL ampicillin was inoculated for all plasmids, except SgK269 pMIH plasmids which were grown in Super Broth (SB). All cultures were incubated for 16-18 hours at 37 °C, shaking 180-220 RPM. DNA isolation through miniprep (GenElute^{TM} Plasmid Miniprep Kit, Sigma) was carried out following culture growth, as per miniprep kit protocol.

All the DNA cloned were PCR sequenced, using big dye terminator (BDT, Micromon, Monash University), and sent to Micromon for analysis.

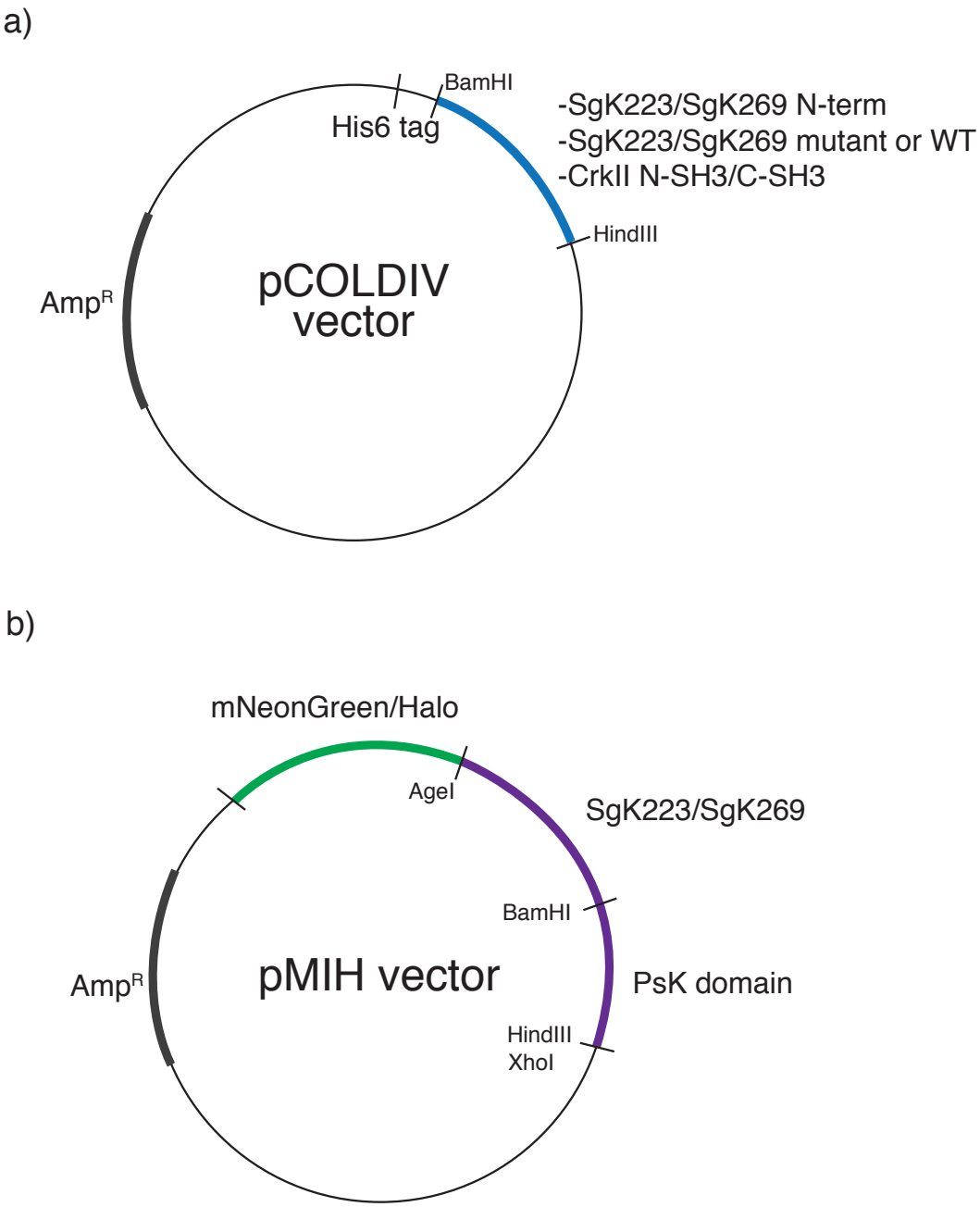


Figure 2.1. Vector maps of all the genes expressed in this study.
a) pCOLDIV vector map for *E.coli* expression of SgK223, SgK269 and CrkII constructs. All constructs have an N-terminal His6 tag and vectors contain an ampicillin resistance (Amp^R) gene.
b) pMIH vector map for mammalian cell expression of SgK223 and SgK269 constructs. SgK223 has an N-terminal mNeonGreen fusion fluorophore and SgK269 has an N-terminal Halo fusion tag. Vectors contain an ampicillin resistance gene.

2.2 Recombinant proteins and synthetic peptides

2.2.1 E.coli protein expression

For *E.coli* transformation, the plasmid containing the gene of interest was incubated with C41 (DE3) *E.coli* cells for 30 minutes on ice. Post incubation, C41 cells were subjected to heat shock for 45 seconds at 42 °C, to incorporate the plasmids into the cells. Following heat shock, the cells were placed back on ice before being plated out onto LB Agar with 100 μg/mL ampicillin for colony selection. Plates were incubated at 37 °C for 16-18 hours.

Following the transformation, overnight cultures were set up. For the IPTG induction method, 10 mL of SB for every 1 L of final culture volume was inoculated. 100 μg/mL ampicillin was added to the overnight culture and then inoculated with a single colony from the transformed plate. The culture was incubated for 16-18 hours at 37 °C, shaking at 200 RPM.

For every 10 mL of SB overnight culture, 1 L of SB was inoculated following incubation. 100 μg/mL of ampicillin was added to the SB and the cultures were incubated at 37 °C, shaking at 180 RPM, until their optical density (OD) reached 0.8 AU at 600 nm. The cultures were then moved to 16 °C incubation, shaking at 180 RPM, and induced with IPTG, 0.25 mM for SgK223 and SgK269 N-terminal expressing cultures or 0.5 mM IPTG for all other cultures. Cultures were incubated at 16 °C for 16-18 hours, shaking at 180 RPM, before cells were harvested. Cell pellets were harvested by centrifugation at 5000 RPM for 20 minutes at 4 °C. Lysates were discarded and cell pellets flash frozen at -80 °C.

The autoinduction method followed a similar protocol, however for the overnight culture, LB is used instead of SB, with a number of additives (see Table 2.1). Following the overnight culture, LB is mixed with the additives for autoinduction and is inoculated directly with the overnight culture. The expression culture was incubated at 16 °C, 180 RPM shaking, for 72 hours. Following incubation, the cells were harvested as described for the IPTG induction method.

2.2.2 Purification of recombinant protein

For purification of *E.coli* expressed proteins (SgK269 C-terminal WT and mutants, SgK223 C-terminal WT and mutants, SgK269 N-terminal and SgK223 N-terminal domains), cells were resuspended in resuspension buffer (see Table 2.1) treated with DNase (Sigma), 10 mg lysozyme (Sigma) and ½ a cOmplete EDTA free protease inhibitor cocktail tablet per 1 L of cells (Roche) and incubated for 10 minutes at room temperature. Resuspended cells were sonicated to release cell lysate and then centrifuged at 18,000 RPM for 30 minutes at 4 °C to pellet cell debris. Cell lysate was filtered (0.45 μm) and then loaded onto a Nickel-NTA column (1 mL) (Roche), washed once with resuspension buffer and then again without detergent and eluted with 250 mM Imidazole. Fractions containing protein were subjected to size exclusion chromatography (SEC) (Superdex HiLoad 75/200 16/600 GE column) in SEC buffer (see Table 2.1). Fractions were analysed on Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and pure fractions pooled and concentrated, as specified in each Chapter, and flash frozen at -80 °C until later use.

Following SEC, the SgK223 and SgK269 N-terminal domains required further purification, using ion exchange chromatography. SgK223 N-terminal domain was subjected to MonoQ (anion exchange, GE Healthcare) and SgK269 was subjected to MonoS (cation exchange, GE Healthcare). Following ion exchange chromatography, the proteins were concentrated to ~3 mg/mL and flash frozen at -80 °C until later use.

An additional step was added to cleave the His6 tags on SgK223 and SgK269 N-terminal domains for ion binding studies, and to cleave the His6 tag on CrkII N-SH3 domain for crystallography. After SEC, the proteins were incubated with Tobacco Etch Virus (TEV) protease in a ratio of 1 mg TEV: 15 mg protein to be cleaved, for 14-18 hours on ice and at 4 °C. After incubation, the TEV-protein mixture was incubated for 2 hours at room temperature with a further aliquot of TEV protease (~ ¼ of the original volume added) to enhance and complete His6 tag cleavage. Following incubation, the proteins were subjected to ion exchange

chromatography to separate the cleaved protein from TEV protease and any uncleaved protein. TEV protease does not bind to the MonoS or MonoQ column and elutes in the injection and column wash phase of the chromatography. Eluted fractions were analysed, concentrated and stored as previously described.

2.2.3 Synthetic peptides

Synthetic proline-rich region peptides of the PEAK family were purchased from Mimotopes, purified to 95% purity (see Chapter 4, Table 4.2.2 for peptide sequences). The N-terminus of the peptides were either biotinylated or acetylated, for use in Surface Plasmon Resonance (SPR) and crystallography experiments. The stock solution concentration of the peptides was 100 mM for the acetylated peptides and 50 mM for the biotinylated peptides. For SPR peptide immobilised experiments the peptides were immobilised at 100 μM concentration, diluted from the 50 mM accordingly, and for protein immobilised experiments the peptides were serially diluted from 100 μM to 0.098 μM or 100 μM to 1.56 μM.

Table 2.1. Buffer compositions.

Buffer	Components	Protein
Resuspension	20 mM Tris pH 8.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 2 mM EDTA, 10 mM DTT	SgK223 C-term Mutants and WT
	20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 2 mM EDTA, 10 mM DTT	SgK269 C-term Mutants and WT
	20 mM Tris pH 8.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 5 mM DTT	SgK223 N-terminal domain
	20 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 5 mM DTT	SgK269 N-terminal domain
	20 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 5 mM Imidazole	CrkII N-SH3
	10 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 5 mM Imidazole	CrkII N-SH3/C-SH3 (SPR)
	20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 5 mM Imidazole	CrkII C-SH3
Nickel affinity elution	20 mM Tris pH 8.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 2 mM EDTA, 150 mM DTT	SgK223 C-term Mutants and WT
	20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 2 mM EDTA, 250 mM DTT	SgK269 C-term Mutants and WT
	20 mM Tris pH 8.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 250 mM DTT	SgK223 N-terminal domain
	20 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 0.1% thesit, 250 mM DTT	SgK269 N-terminal domain
	20 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 250 mM Imidazole	CrkII N-SH3
	10 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 250 mM Imidazole	CrkII N-SH3/C-SH3 (SPR)

	20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, 5 mM DTT, 0.1% Thesit, 250 mM Imidazole	CrkII C-SH3
Size exclusion	20 mM Tris pH 8.5, 200 mM NaCl, 5% glycerol, 1 mM TCEP	SgK223 C-term Mutants and WT
	20 mM Tris pH 8.0, 200 mM NaCl, 5% glycerol, 1 mM TCEP	SgK269 C-term Mutants and WT
	20 mM Tris pH 8.5, 250 mM NaCl, 5% glycerol, 0.5 mM TCEP	SgK223 N-terminal domain
	20 mM Tris pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM TCEP	SgK269 N-terminal domain
	20 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, 0.5 mM TCEP	CrkII N-SH3
	10 mM Hepes pH 7.5, 200 mM NaCl, 0.5 mM TCEP	CrkII N-SH3/C-SH3 (SPR)
	20 mM Tris pH 8.0, 200 mM NaCl, 5% glycerol, 0.5 mM TCEP	CrkII C-SH3
Analytical size exclusion	20 mM Tris pH 8.5, 5% glycerol, 200 mM NaCl, 1 mM TCEP	SgK223 C-term Mutants and WT
	20 mM Tris pH 8.5, 5% glycerol, 200 mM NaCl, 1 mM TCEP	SgK269 C-term Mutants and WT
	20 mM Tris pH 8.5, 200 mM NaCl, 0.5 mM TCEP	SgK223 N-terminal domain
	20 mM Tris pH 8.5, 200 mM NaCl, 0.01 mM $ZnSO_4$, 0.5 mM TCEP	SgK223 N-terminal domain + zinc
	20 mM Tris pH 7.5, 200 mM NaCl, 0.5 TCEP	SgK269 N-terminal domain
	20 mM Tris pH 7.5, 200 mM NaCl, 0.01 mM $ZnSO_4$, 0.5 mM TCEP	SgK269 N-terminal domain + zinc
Analytical ultracentrifugation	20 mM Tris pH 8.5, 200 mM NaCl, 1 mM TCEP	SgK223 and SgK269 C-term Mutants and WT
	20 mM Tris pH 7.0, 200 mM NaCl, 0.5 mM TCEP	SgK269 N-terminal domain
	20 mM Tris pH 7.0, 200 mM NaCl, 0.5 mM TCEP	SgK223 N-terminal domain
TSA	20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT	All
Circular dichroism buffer	20 mM Tris pH 8.5, 200 mM NaCl	SgK223 N-terminal domain
	20 mM Tris pH 8.5, 200 mM NaCl, 0.01 mM $ZnSO_4$	SgK223 N-terminal domain + zinc
	20 mM Tris pH 7.5, 200 mM NaCl	SgK269 N-terminal domain

	20 mM Tris pH 7.5, 200 mM NaCl, 0.01 mM $ZnSO_4$	SgK269 N-terminal domain + zinc
SPR HBS-EP buffer	10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 1 mM TCEP, 0.005% (v/v) Tween20, pH 7.4	CrkII C-SH3 and N-SH3
Roswell Park Memorial Institute (RPMI) media	RPMI 1640 powder, 110 mg/mL Sodium pyruvate, 2 g/L Sodium bicarbonate, 1 x Penicillin/streptomycin solution	U2OS cells
Dulbecco's Modified Eagle's (DME) medium	DMEM powder, 1 g/L D-Glucose, L-Glutamine, 110 mg/mL Sodium pyruvate, 1 x Penicillin/streptomycin solution	HEK293, MDA-MB-231, DLD1 cells
Optimem media	Optimem powder, Hepes, 2.4 g/L Sodium bicarbonate, L-Glutamine	
Phosphate buffered saline	20 mM Na_2HPO_4 4.5 mM NaH_2PO_4 150 mM NaCl	
Autoinduction overnight culture media	Luria Broth with 1 mM $MgCl_2$, 0.8% Glucose, 1x NPS	SgK223 C-term WT and mutants, SgK269 C-term WT and mutants
Autoinduction expression media	Luria Broth with 1 mM $MgCl_2$, 1 x 5052, 1x NPS	SgK223 C-term WT and mutants, SgK269 C-term WT and mutants
20 x NPS	0.5 M $(NH_4)SO_4$, 1 M KH_2PO_4, 1 M Na_2HPO_4	
50 x 5052	0.5% glycerol, 0.05% glucose, 0.2% α-lactose	

Table 2.2. Construct details of all proteins expressed in this study.

CONSTRUCT	pI	MW (kDa)	EXPRESSION SYSTEM
pCOLD 6xHis SgK223 N-term 1-202	7.7	24.6	*E.coli*
pCOLD 6xHis SgK269 N-term 1-199	8.8	24.3	*E.coli*
pCOLD 6xHis SgK223 C-term WT 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term F992A 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term C996R 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term P999G 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term V1089R 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term A1367D, L955E 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK223- C-term A1367D, L955E, F1271A 932-1406	7.0	51.8	*E.coli*
pCOLD 6xHis SgK269- C-term WT 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term I1290A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term L1301A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term F1706A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term W1722A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term F1327A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term S1331R 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term P1334G 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term A1707D, L1290E 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis SgK269- C-term A1707D, L1290E, F1609A 1267-1746	6.1	56.8	*E.coli*
pCOLD 6xHis CrkII N-SH3, 134-191	5.8	9.1	*E.coli*

pCOLD 6xHis CrkII C-SH3, 218-299	9.3	11.2	*E.coli*
pMIH mNeonGreen SgK223-WT	6.9	176	Human cells
pMIH mNeonGreen SgK223 L955E, F1271A, A1367D	6.9	176	Human cells
pMIH Halo SgK269-WT	6.4	194	Human cells
pMIH Halo SgK269 I1290E, F1609A, A1707D	6.4	194	Human cells

2.3 Biochemical assays

2.3.1 Thermal stability assays

Purified SgK223 and SgK269 protein variants, at 7 μg for SgK223 and 14 μg for SgK269 per reaction mix, were diluted in TSA buffer and combined with 1 μL of 1:20 dilution (in DMSO) of Sypro Orange dye (Life Sciences) in a reaction well for a total reaction mixture of 25 μL. The wells were added to a Qiagen Rotor-Gene Q PCR machine and the temperature was raised in 1 °C intervals from 25 °C to 80 °C, with fluorescence read at each step. Fluorescence was detected at 530 nm [65]. Data were plotted as temperature vs. fluorescence and the Tm (midpoint of protein unfolding) was calculated through fitting the sigmoidal melt curve to the Boltzmann equation using GraphPad Prism version 8.0.2 for Mac, Graphpad Software, La Jolla California USA, www.graphpad.com.

2.3.1 Circular dichroism

Samples were prepared for Circular Dichroism (CD) directly after protein purification. Samples were diluted approximately 10-fold, to dilute the TCEP in the purification buffer as TCEP can cause interference in CD, and then concentrated to the required concentration of 0.15 mg/mL in 300 μL volume. CD spectra were recorded on the Aviv model 410 spectropolarimeter (Aviv Biomedical, Lakewood, NJ, USA). A quartz cell of 0.1 cm path length was used to record the spectra over the wavelength range 260-190 nm with a bandwidth of 1.0 nm. Measurements were performed in buffers as listed in Table 2.1, using a protein concentration of 0.15 mg/mL. Each spectrum was baseline corrected for the solvent contribution. Data were exported and imported into Prism (Graphpad), where spectra of wavelength versus molar ellipticity were constructed. CD signal was recorded from the instrument as milli absorbance (mAbs) and this was converted to molar ellipticity through the equation $\frac{mAbs \cdot MRW \cdot 3.298}{0.1 \cdot l}$ whereby MRW is the mean residual MW of the protein ($MRW = \frac{MW}{number\ of\ amino\ acids - 1}$), l is the path length, in this case 0.1 cm. Data were analysed using Dichroweb, using a mean

residue weight of 123 Da for SgK223 N-terminal domain and SgK269 N-terminal domain. The algorithm used for data fitting was K2D [66, 67].

2.3.2 Analytical size exclusion

All SEC analyses were performed on a Superdex^{TM} 200 Increase 10/300 GL or Superdex^{TM} 75 Increase 10/300 GL (GE Healthcare) (see Table 2.1 for buffers). Recombinant protein was used at concentrations of 30 μM for SgK223 and SgK269 pseudokinase domain mutants or 50-80 μM for SgK223 and SgK269 N-terminal domains, in final volume of 110 μL for loading onto the column. For complexation, samples were pre-incubated for 10 minutes on ice, then all samples were centrifuged (12000 RPM, 10 minutes, 4 °C) to sediment any debris prior to loading onto the SEC column. Eluted fractions of protein were analysed by SDS-PAGE.

To calculate predicted MW based on elution profiles, MW standards (GE Healthcare) were loaded onto the column at 1 mg/mL each in a total volume of 500 μL. Using Prism (Graphpad), the elution profiles were plotted and the peak elution volume of each protein standard was calculated. Next, the partition coefficient, K_{av}, was calculated. The equation for calculating K_{av} is $Kav\frac{Ve-Vo}{Vc-Vo}$, whereby V_e is the elution volume of the protein, V_o is the void volume of the column, determined by the peak elution volume of blue dextran, and V_c is the geometric column volume. A plot of the logMr (molecular mass) on the x axis versus the K_{av} on the y axis was generated. The K_{av} of SgK223 and SgK269 N-terminal domains was calculated, and using the logMr versus K_{av} plot, the logMr, and thus the MW, of SgK223 and SgK269 N-terminal domains was calculated.

2.3.4 Analytical ultracentrifugation

Sedimentation velocity experiments were carried out using an XL-I analytical ultracentrifuge (Beckman Coulter) with UV-Vis scanning optics. Reference buffer (see Table 2.1) and samples in 12 mm double-sector cells with quartz windows and were placed in an An-60Ti 4-hole rotor. Sample concentrations were 2.5, 5

and 10 μM for SgK223 and SgK269 N-terminal domains, 3 μM for SgK269 dimerisation domain mutants and 20 μM for SgK223 and SgK269 pseudokinase domain oligomerisation mutants. Experiments were carried out at 50,000 RPM (201,600 xg) at 20 °C, and radial absorbance data were collected at appropriate wavelengths in continuous mode. Data were fitted to a continuous sedimentation distribution model [c(s)] model. Molecular weights were calculate using the frictional ratio obtained from the fit to the sedimentation velocity data.

AUC runs were carried out at Bio21 Institute by Dr. Michael Griffin (University of Melbourne).

2.4 Measuring binding affinity - surface plasmon resonance

2.4.1 Immobilisation of CrkII SH3 domains

Experiments were carried out on the Biacore S200, GE Healthcare. CrkII SH3 domains were immobilised to a Series S Sensor Chip CM5 (GE Healthcare) at 25 °C using amine coupling according to the manufacturer's instructions (GE Healthcare). The chip surface was pre-equilibrated in HBS-EP buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 1 mM TCEP, 0.005% (v/v) Tween20, pH 7.4), then activated with a 1:1 ratio of EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) and NHS (N-Hydroxysuccinimide) for 420 seconds at a 10 μL/min flow rate. CrkII SH3 domains were immobilised using either manual injection (10 μL/min) or the automated wizard from the Biacore S200 software (GE Healthcare). The N-SH3 domain was immobilised onto two flow cells, at two different densities (1380 RU and 300 RU), and C-SH3 was immobilised onto one flow cell at a density of 625 RU. Deactivation followed immobilisation, using 1 M ethanolamine pH 8.5. The reference cell was also activated/deactivated as the CrkII SH3 domain flow cells, to serve as a control.

2.4.2 Interaction experiments with PEAK family PRMs

Peptides were synthesised commercially to 95% purity and prepared at either 100 mM or 50 mM in sterile 0.1% acetic acid in water. The stock solutions were diluted to 100 μM and then serially diluted (11-point, two-fold serial dilution from 100 μM to 0.098 μM for the consensus binding peptides and 7-fold, two-fold serial dilution from 100 μM to 1.56 μM for the two non-consensus peptides of SgK223 and SgK269) in HBS-EP buffer. 200 μL of each peptide concentration was required in duplicate for each run. Multicycle runs, with a 90 s injection, 140 s dissociation at a flow rate of 70 μL/min, were carried out with each peptide concentration in duplicate and three independent experiments were performed. A syringe wash was included after each injection with 50% DMSO. Residual binding to the chip was observed for SgK223$^{809\text{-}821}$ so a regeneration step was

required, consisting of an injection of 100 μM SgK269$^{1150\text{-}1162}$ for 12 s followed by 30 s dissociation. This successfully returned the baseline to zero.

2.4.3 Immobilisation of biotinylated PRMs

The chips surface was pre-equilibrated in HBS-EP buffer as described above. Biotinylated peptides were conjugated to the chip using the automated wizard with extra immobilisation to reach the desired RU (100 RU) using manual injection (10 μL/min). Two chips were coated with peptides, Chip 1 (FC1= blank reference, FC2= PEAK3$^{54\text{-}66}$, FC3= SgK269$^{1150\text{-}1162}$, FC4= SgK269$^{1117\text{-}1129}$) and Chip 2 (FC1= blank reference, FC2= blank, FC3= SgK223$^{709\text{-}721}$, FC4= SgK223$^{809\text{-}821}$).

2.4.4 Interaction experiments with CrKII SH3 domains

CrkII N-SH3 and C-SH3 domains in a 10 mM HEPES pH 7.4, 150 mM NaCl buffer were diluted from 100 μM to 0.098 μM in an 11-point titration and flowed over the PEAK family PRM peptides. 200 μL of each protein concentration was required in duplicate for each run. Multicycle kinetic runs, with a 35 s injection, 60 s dissociation at a flow rate of 70 μL/min, were carried out with each protein concentration in duplicate and three independent experiments were performed.

2.4.5 SPR data analysis

Data were analysed using the Biacore S200 evaluation software (GE Healthcare). Data from the reference flow cells were subtracted from the raw data as well as the blank injections. The data were solvent corrected for analysis. Steady-state affinity fitting was applied to all data to calculate the K_D of interaction using a 1:1 binding model, with all affinity values reported in Tables and Figures determined using steady-state fitting. Similar values were observed using kinetic fitting, assuming a 1:1 binding model, however, as K_{on} and K_{off} rates were so fast, this data is unreliable.

2.5 X-ray crystallography

2.5.1 Crystallisation

CrkII N-SH3 domain was cleaved of the His6 tag for crystallography. Initially, CrkII N-SH3 domain at 27 mg/mL was incubated with excess SgK269$^{1150\text{-}1162}$ peptide (Table 4.2.2, Chapter 4) in a 1:1.25 ratio and PEAK3 peptide for 30 minutes on ice, prior to drop setting of the formed complex. Crystallisation screening was carried out at CSIRO C3 facility in the Shotgun screen, a broad screen devised by Janet Newman at C3 (96-well, 150 nL protein and 150 nL reservoir solution) [68]. As this resulted in only clear drops, indicating the concentration of protein was not high enough, CrkII N-SH3 domain was concentrated to 40 mg/mL and screened in complex with SgK269$^{1150\text{-}1162}$ peptide in the PI_PI2_C3 (based on Peg/Ion and Peg/Ion II screens originally devised by Hampton Research), C3_2 (C3 devised screen) and Proplex (based on Molecular Dimensions screen) screens. Based on these screens, a hit condition was identified (0.1 M Na Cacodylate, 45% Peg 2000 MME) and further in-house fine screening was performed, based on this condition. Protein was further concentrated to 66 mg/mL to enhance crystal formation and the ratio of protein to reservoir in the drop was set up at 1:2 (0.5 μL: 1 μL) as this resulted in protein precipitation and eventually crystals. The crystals used for data collection were crystallised in 43% PEG 2000 MME, 0.1 M Na Cacodylate pH 6.3 at 20 °C; crystals were harvested directly from the drop and flash frozen in N_2 (liq.) without the need for additional cryoprotection.

2.5.2 Data collection

X-ray diffraction data were collected on the MX2 beamline at the Australian Synchrotron by Onisha Patel and Michael Roy. The data were collected at a wavelength of 0.9537 Å.

2.5.3 Data processing

Diffraction data for the CrkII N-SH3 domain were processed in XDS and Buster to 1.5 Å for the apo data and 2.1 Å for the CrkII N-SH3:SgK269$^{1150\text{-}1162}$ complex data, by Michael Roy [69, 70]. CrkII N-SH3 structures were solved by molecular replacement in PHASER, using the CrkII N-SH3 domain (PDB: 5IH2) as a search model [71]. The Matthew's probability calculator (http://www.ruppweb.org/mattprob) MATTPROB was used to calculate the number of CrkII N-SH3 molecules per asymmetric unit based on estimated solvent content, using the unit cell parameters [72, 73].

2.5.4 Refinement

The initial structures were improved by multiple rounds of building in Coot and refinement in Phenix [74, 75]. The CrkII N-SH3 domain apo crystallised in the space group $P\,4_1\,2_1\,2$ with unit cell dimensions 33.1, 33.1, 43.8 Å and angles 90, 90, 90 °, with one molecule in the asymmetric unit. The CrkII N-SH3:SgK269$^{1150\text{-}1162}$ complex crystallised in the space group $P\,4_1\,2_1\,2$ with unit cell dimensions 74.3, 74.3, 31.2 Å and angles 90, 90, 90 °, with one molecule in the asymmetric unit. The data collection and refinement statistics are listed in Table 4.2.2 in Chapter 4. Figures of the structures were made in Pymol or Chimera [76].

2.6 Imaging

2.6.1 Cell culture

MDA-MB-231, HEK293 and DLD1 cells were cultured at 37 °C, 5% CO_2 in Dulbecco's modified eagle medium (DMEM) media (Life Technologies) supplemented with 10% fetal calf serum (FCS). U2OS cells were cultured in RPMI 1640 media supplemented with 10% FCS in the same incubation conditions. For regular cell passaging, cells were washed with phosphate buffered saline (PBS) prior to adding trypsin to detach cells from the dish. After cells had detached they were resuspended in media (approximately 10 x the volume of the trypsin) and centrifuged at 5000 RPM for 5 mins to pellet cells. Media/trypsin supernatant was aspirated and the cells were resuspended in fresh media before diluting them into a new flask. Cell passaging was carried out three times a week for regular cell maintenance.

2.6.2 Cell plating for imaging experiments

For imaging experiments, cells were seeded into either a μ-Slide 8-well Ibidi chamber, an 8-well cover glass II Sarstedt chamber or μ-Dish 35 mm Ibidi chamber. Cell densities for seeding varied based on experiment, cell type and dish type. For experiments in the 8-well chambers, cells were seeded between 2.5-9 x 10^4 cells per well. For experiments in the 35-mm dish, cells were seeded between 6-9 x 10^5 cells per dish. Cell seeding followed a similar protocol to regular passaging, however after cells were pelleted, they were resuspended in 1-2 mL media. An aliquot of cells was mixed with a 1:1 of cells: trypan blue dye. Cells were counted using the Countess II FL Automated cell counter (Thermo Fisher). Cells were then seeded into the appropriate chamber/dish for imaging.

2.6.3 Transient transfections of SgK223 and paxillin for imaging

24-34 hours after cell plating, cells were treated with reagents for lipid transient transfection. For the majority of experiments, cells were transfected with Lipofectamine 3000 (Thermo Fisher), following reagent protocols. Experiments

in 8-well chambers (Ibidi or Sarstedt) were treated with 0.2 μg of DNA (either paxillin, SgK223 or both) per well and experiments in 35 mm dishes (Ibidi) were treated with 5 μg DNA per well. Lipofectamine 3000 reagent was incubated in Optimem reduced serum media (Gibco) for 5 minutes, P3000 reagent and the DNA were incubated in Optimem for 5 minutes. The Lipofectamine 3000 mixture was then mixed with the DNA/P3000 mix in Optimem and incubated for 15 minutes at room temperature. Cells to be transfected were aspirated of serum containing media and Optimem was added to these wells. The Lipofectamine mixture was then added to each well, dropwise, and incubated for 12 hours. Following the 12 hours, media containing the Lipofectamine was aspirated and media replaced with either L-15 media (Gibco) supplemented with 10% FCS or fluorobrite media (Gibco) supplemented with 10% FCS. Imaging was carried out within 6 hours of the media change.

2.6.4 Image acquisition

Live cell imaging was carried out on either the Laser scanning confocal microscope Zeiss LSM 880 or the Super Resolution Nikon dSTORM with Adaptive Optics. Airyscanning processing was used on the LSM 880 and total internal fluorescence microscopy (TIRF) on the Nikon dSTORM. Slides were mounted in the imaging chamber fitted with an incubation chamber maintained at 37 °C and 5% CO_2. Images were captured with a 63x objective collecting images beginning at the coverslip, where the focal adhesions are found, and subsequently ~20 Z-slices through the cell. Cells were excited with 488 nm and 561 nm lasers. The images of cells show the slice from the Z-stacks that is representative of the focal adhesion plane (usually closest to the coverslip).

2.6.5 Image analysis

Images were analysed in Imaris, using the co-localisation function. Pearson's coefficients were calculated based on paxillin vs. SgK223 fluorescence and values were imported into Prism for graphing. Cell area was calculated in Fiji, by initially thresholding the image based on paxillin fluorescence and then utilising the measurement function to calculate cell area.

3.0 Biochemical and functional characterisation of SgK223 and SgK269

SgK223 and SgK269 are homologous multidomain proteins, comprised of an N-terminal domain of unknown structure and function, a PEST region containing critical tyrosine residues for signalling, and a C-terminal pseudokinase domain flanked by regulatory helices (Fig. 3.1.1a) [6-8].

This Chapter is divided into two distinct parts. The first part aims to investigate the function of the N-terminal domains of SgK223 and SgK269, using biochemical and biophysical approaches. The second part of this Chapter aims to characterise the homo- and hetero-associations of SgK223 and SgK269, through mutational and biochemical analyses of their dimerisation and oligomerisation interfaces.

3.1 The N-terminal domain of SgK223 and SgK269

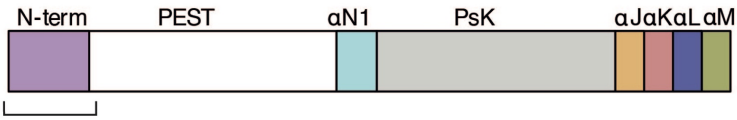
The N-terminal domains of SgK223 and SgK269 make up the first ~200 residues with ~28% sequence identity between them, however, this region does not have similarity to any known protein (Fig. 3.1.1b). The biological role of the N-terminal domains of SgK223 and SgK269 is poorly understood. N- and C-terminal truncation mutants of SgK269 fused to GFP were expressed in NIH 3T3 fibroblast cells to investigate their localisation to the actin cytoskeleton and focal adhesions. The construct of the first 300 amino acids, including the N-terminal domain and a small region of the PEST linker, was unable to localise to the cytoskeleton or focal adhesion [3]. Additionally, overexpression of SgK269 in mammary epithelial cells (MCF-10A) increased cell migration and cell elongation, which was unaffected by deletion of the first 300 residues. Thus, it seems that this domain does not play a role in cell migration or regulating cell morphology [1].

The first 50 residues of the N-terminal domains of SgK223 and SgK269 are highly conserved and display a number of cysteine and histidine residues (Fig. 3.1.1, residues highlighted in purple). Whether or not this cysteine/histidine motif (X_2-

Cys-$X_{2,4}$-Cys-X_{12}-His-$X_{3,4,5}$-His) represents a putative ion binding site will be investigated [77]. Another motif that is highly conserved between SgK223 and SgK269 N-terminal domains lies at ~80 residues into the N-terminal domains. The sequence is 11 residues long, however the importance of this motif is unknown. In addition, SgK223 N-terminal domain displays a proline-rich region that is likely to interact with SH3 domain containing proteins (Fig. 3.1.1, residues coloured orange).

In this Chapter I aim to biochemically characterise the N-terminal domains of SgK223 and SgK269.

a)



1-202 SgK223

1-199 SgK269

b)

```
SgK223   MHQTLCLNPESLKMSACSDFVEHIWKPGSCKNCFCLRSDHQLVAGPPQPRAGSLPPPPRL   60
SgK269   --------------SACNTFTEHVWKPGECKNCFKPKSLHQLPPDPEKAPITHGNV-KTN   45
                      ***. *.**:****.*****  :* ***  .* :

SgK223   PPRPENCRLEDEGVN-SSPYSKPTIAVKPTMMSSEASDVWTEANLSAE----VSQVIWRR   115
SgK269   ANHSNNHRIRNTGNFRPPVAKKPTIAVKPTMIVADGQSICGELSIQEHCENKPVIIGWNR   105
           : :* *:.: *       .**********: ::...:  * .:. .       : *.*

SgK223   APGKLPLPKQEDAPVVYLGSFRGVQKPAGPSTSPDGNSRC-----PPAYTMVGLHNLE-P   169
SgK269   NRAALSQKPLNNNNE-DDEGISHVPKPYGNNDSAKKMSDNNNGLTEVLKEIAGLDTAPQI   164
           . *     ::       .:  * ** * . * .  *            :.**..

SgK223   RGERNIAFHPVSFPEEKAVHKEKPSFPYQDRPS-----     202
SgK269   RGNETN--SRETFLGRI-NDCYKRSLERKLPPSCMIGG     199
         **:..      :*  .   .  * *:  :  **
```

c) X_2-Cys-$X_{2,4}$-Cys-X_{12}-His-$X_{3,4,5}$-His

Figure 3.1.1. SgK223 and SgK269 N-terminal domains.
a) Domain layout of SgK223 and SgK269. The N-terminal domains are in purple and domain boundaries are highlighted.
b) Sequence alignment of SgK223 and SgK269 N-terminal domains. Histidine and cysteine residues highlighted in purple that may contribute to an ion binding motif. Residues in orange that form a proline-rich motif.
c) Cysteine/histidine consensus motif.

3.2 Results: Biochemical characterisation of SgK223 and SgK269 N-terminal domains

3.2.1 Expression and purification of SgK223 and SgK269 N-terminal domains

SgK223 and SgK269 N-terminal domains (domain boundaries 1-202 and 1-199, respectively) were fused N-terminally to a His6 tag with a TEV protease cleavage site after the tag and expressed in a pCOLDIV vector in *E.coli* (see Chapter 2, Table 2.2 for details). Purification of the domains was conducted by nickel affinity chromatography (Roche) and size exclusion chromatography (SEC) (Superdex HiLoad 75 16/600 column from GE healthcare) (Figure 3.2.1). As the SgK223 and SgK269 N-terminal domains were prone to proteolytic cleavage, an additional purification step using ion exchange chromatography was included after SEC (MonoS cation exchange chromatography for SgK269 and MonoQ anion exchange chromatography for SgK223) to remove the proteolytic products (Fig. 3.2.1). For ion binding studies, the His6 tag was cleaved using TEV protease to prevent binding of the divalent ions to the tag and the cleaved proteins were further purified using ion-exchange chromatography as described above. SDS-PAGE confirmed protein purity to ~100% and pure proteins were concentrated to ~ 3 mg/mL and stored at -80°C.

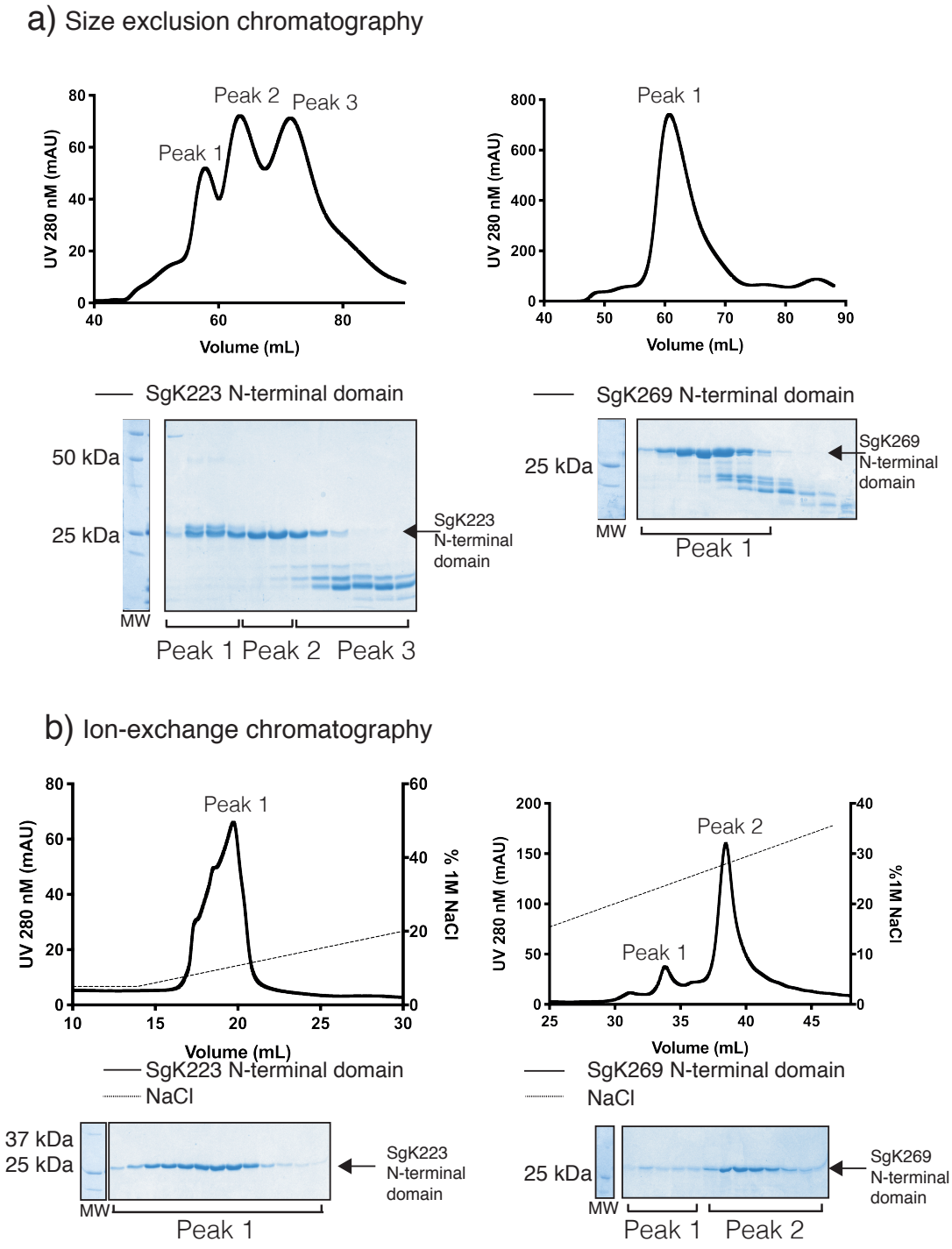


Figure 3.2.1. Purification of SgK223 and SgK269 N-terminal domains.
a) Size exclusion chromatography elution profiles of SgK223 and SgK239 N-terminal domains. SDS-PAGE of fractions from SEC elution, with peaks labelled correspondingly.
b) Ion-exchange chromatography elution profile of SgK223 and SgK269 N-terminal domains. SDS-PAGE of fractions from SEC elution, with peaks labelled correspondingly. MW= Molecular Weight marker (BioRad Precision Plus unstained protein standards).

3.2.2 Analytical Size Exclusion Chromatography of SgK223 and SgK269 N-terminal domains

Based on the elution profiles from SEC (Fig. 3.2.1a), it was evident that SgK223 and SgK269 N-terminal domains eluted at ~ 62 mL, earlier than the expected elution volume of ~ 70 mL (according to molecular weight standards from GE Healthcare) corresponding to a larger molecular weight (MW) than predicted based on their sequence. Thus, investigations into the oligomeric state of SgK223 and SgK269 N-terminal domains were carried out through analytical size exclusion chromatography (A-SEC), using a Superdex 75 10/300 column (GE Healthcare). SgK223 and SgK269 N-terminal domains eluted at 11.6 mL and 11.5 mL, respectively (Fig. 3.2.2a). Based on MW standards, the MWs of the proteins were calculated to be ~58 kDa (see Chapter 2 for calculation details). As their expected MWs are ~25 kDa, these data suggest these domains may undergo dimerisation (Fig. 3.2.2a).

3.2.3 Analytical Ultracentrifugation of SgK223 and SgK269 N-terminal domains

To further investigate the dimeric state of SgK223 and SgK269 N-terminal domains, we carried out Analytical Ultracentrifugation (AUC). AUC is an ultracentrifugation technique that sediments particles in solution, while an optical detection system monitors the sedimentation. This technique can quantitatively measure protein association, protein mass and provide information on protein shape [78].

SgK223 and SgK269 N-terminal domains were buffer exchanged for AUC through SEC (S75 10/300 column, GE Healthcare) into a glycerol free buffer (see Table 2.1 in Chapter 2 for buffer recipes). Samples were concentrated after SEC and sedimentation velocity experiments were carried out at three concentrations, 2.5 μM, 5 μM and 10 μM.

The AUC data were analysed using the software SedFit [79]. SedFit analysis produced sedimentation coefficient *(c(s))* distributions that showed a singular

peak with a sedimentation coefficient of ~1.5 S for SgK223 N-terminal domain and ~1.7 S for SgK269 N-terminal domain (Fig. 3.2.2b). These coefficients corresponded to molecular masses of ~20 kDa for SgK223 N-terminal domain and ~22 kDa for SgK269 N-terminal domain, suggesting a monomeric structure. Frictional ratios (f/f_0) of 1.8, 1.8, and 1.6 for SgK223 and 1.7, 1.5 and 1.8 for SgK269 were calculated from the fit data for 10 μM, 5 μM and 2.5 μM concentration of SgK223 and SgK269 N-terminal domains, respectively (Table 3.2.1). A perfect sphere would have a frictional coefficient of 1 and a globular protein would usually fall in the range of 1.05-1.35 [80]. Thus, these proteins are most likely elongated and the distribution suggests they are not forming dimeric species in solution. For data fitting see Figures 8.1 and 8.2 in the Appendices.

3.2.4 Circular Dichroism of SgK223 and SgK269 N-terminal domains

To investigate the secondary structure of the N-terminal domains, we carried out circular dichroism (CD) (Fig. 3.2.3). CD is a technique that uses polarised light directed at protein samples. The protein may absorb left or right polarised light, resulting in a molar ellipticity spectra of the protein. Fitting the data generated from the spectra of the protein of interest, using online tools, can provide insight into the secondary structural elements that make up the protein of interest [81].

Protein preparation for CD was carried out similarly to that for AUC; the proteins were buffer exchanged and concentrated to 0.15 mg/mL (see Chapter 2 Table 2.1 for buffers). Experiments were carried out using the Aviv model 410 (Aviv Biomedical, Lakewood, NJ) instrument. Figure 3.2.3a shows the spectra of SgK223 N-terminal (purple) and SgK269 N-terminal (salmon) domains. The data acquired were fitted to the K2D online database using the online application Dichroweb [67]. We found that the data fitting using the online software was unreliable and the fitted curve had large deviations from the actual data. Thus, the online software was unable to generate an accurate analysis of the secondary structural content of the N-terminal domains. Comparing the N-terminal domain curves to those of known secondary structural elements (Fig. 3.2.3a and b), the

N-terminal domains appear to lack secondary structure. The minima of the N-terminal domain on the CD spectra is at ~200 nm wavelength, similar to the minima in the spectra of the disordered data set from the literature curves (Fig. 3.2.3b, shown in green) [81]. Taking the A-SEC, AUC and CD data into account, it is likely that these domains, due to lack of secondary structure, are elongated and thus elute at a higher than predicted MW on A-SEC.

Secondary structure prediction, using the XtalPred Server (Fig. 3.2.4), demonstrated that SgK223 N-terminal domain is predicted to have four main helices, two short (~3 residues) and two longer (~8 residues) and a number of short beta strands (Fig. 3.2.4a). The cysteine/histidine motif is predicted to be partly helical and the other conserved motif is predicted to be a loop. SgK269 N-terminal domain is predicted to have two main helices (11 and 18 residues in length), a number of smaller helices and four predicted beta strands (Fig. 3.2.4b). The cysteine/histidine binding motif is predicted to be two small helices and the second conserved motif is predicted to form two beta strands [82]. Thus, these predictions indicate that SgK223 and SgK269 N-terminal domains could be structured, however the areas of sequence conservation do not seem to have predicted structural conservation between SgK223 and SgK269 (Fig. 3.2.4).

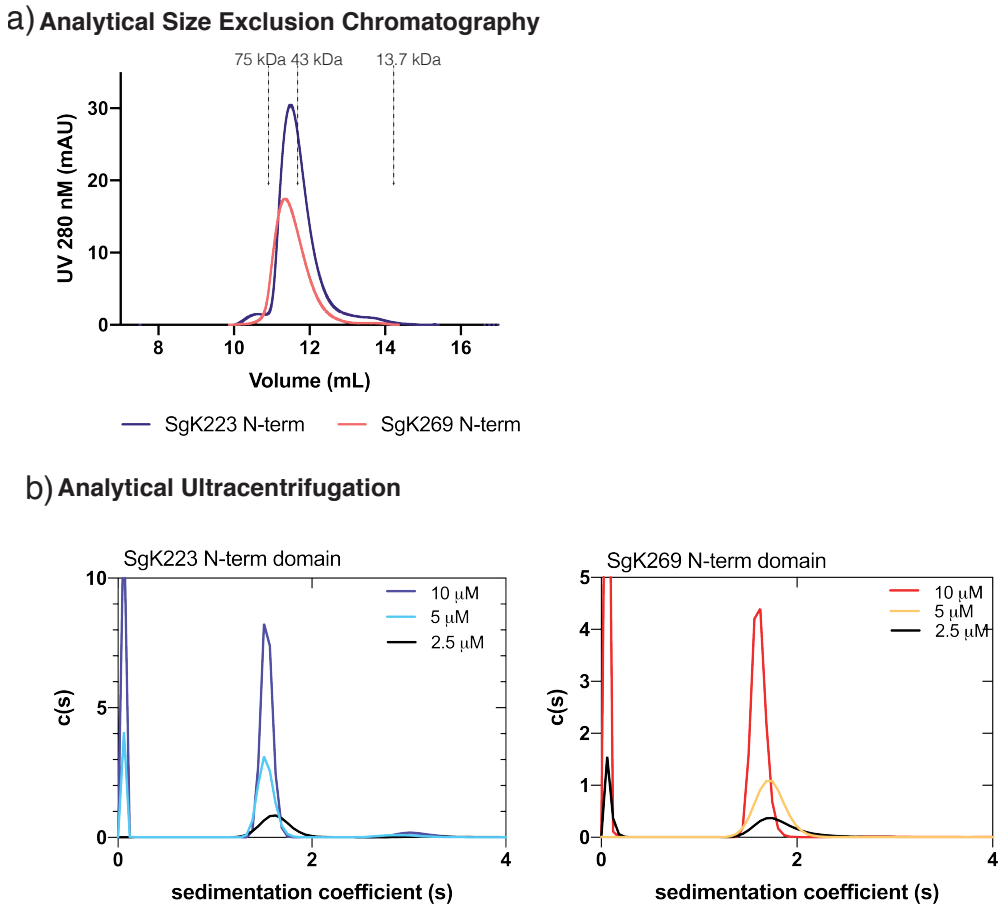


Figure 3.2.2. Biochemical analysis of SgK223 and SgK269 N-terminal domains.
a) Analytical SEC elution profile of SgK223 N-terminal domain and SgK269 N-terminal domains. SgK223 is in purple and SgK269 is in salmon. MW standards are shown in dotted lines and the MWs are listed above the lines.
b) Analytical ultracentrifugation sedimentation coefficient *c(s)* distribution for SgK223 N-terminal domain and SgK269 N-terminal domain.

Table 3.2.1. Molecular weight and frictional ratio of SgK223 and SgK269 N-terminal domain, as measured by AUC.

	SgK223 N-terminal domain			**SgK269 N-terminal domain**		
Concentration	2.5 μM	5 μM	10 μM	2.5 μM	5 μM	10 μM
Molecular weight (kDa)	20	21	21	26	20	22
Frictional Ratio	1.6	1.8	1.8	1.8	1.5	1.7

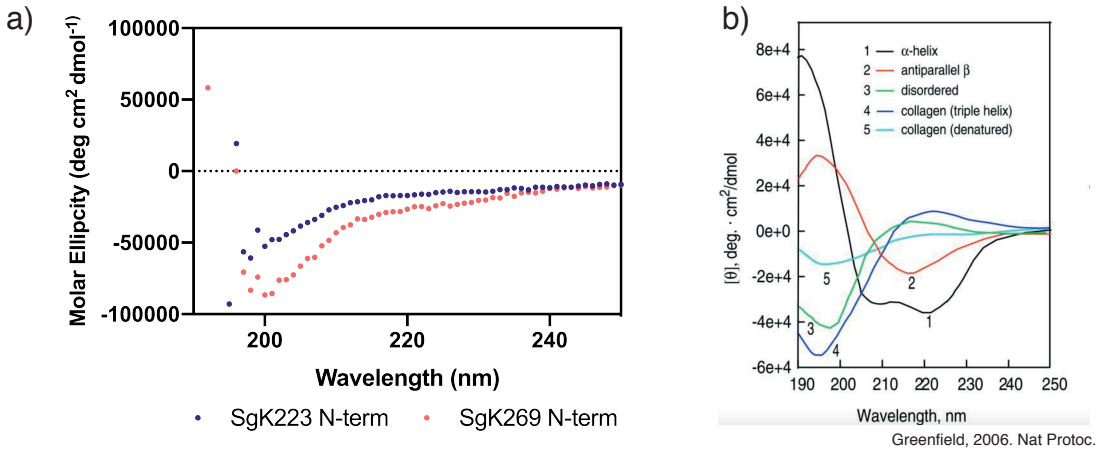
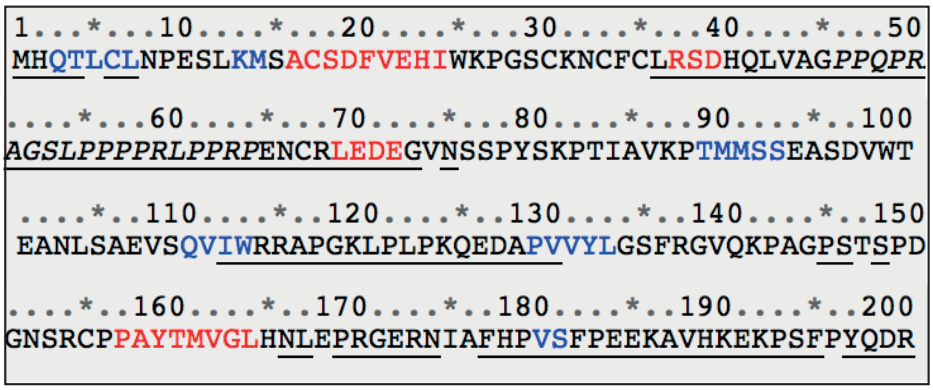


Figure 3.2.3. Circular dichroism of SgK223 and SgK269 N-terminal domains.
a) Circular dichroism spectra of SgK223 N-terminal domain and SgK269 N-terminal domain.
b) Literature CD spectra, green line demonstrates disordered data set.

a) SgK223 N-terminal domain



b) SgK269 N-terminal domain

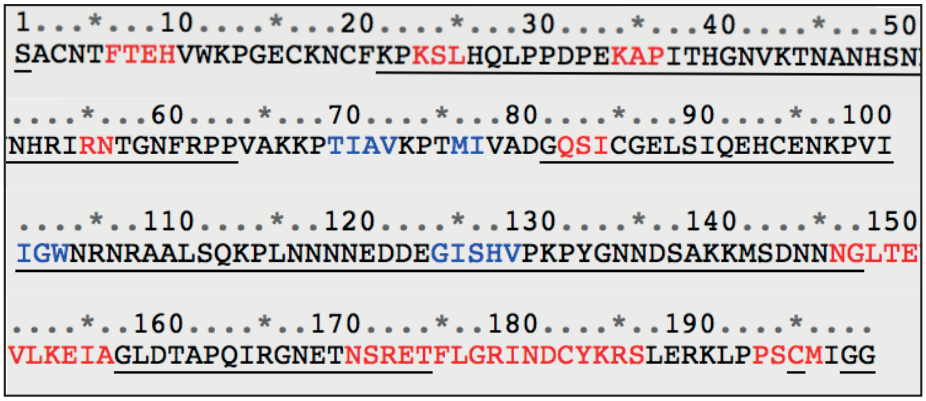


Figure 3.2.4. SgK223 and SgK269 N-terminal domain secondary structure prediction.
a) Secondary structure prediction of SgK223 N-terminal domain from XtalPred. Red font indicates a helix, blue font indicates a beta strand, black font indicates a loop, underline indicates disorder of the protein and italics indicates low protein complexity.
b) Secondary structure prediction of SgK269 N-terminal domain from XtalPred.

3.2.5 Ion binding studies

As previously mentioned, SgK223 and SgK269 have a region at the beginning of the N-terminal domain with cysteine/histidine residues that may resemble an ion binding motif (Fig. 3.1.1b). To investigate if these domains can bind an ion, in this case zinc, and if ion binding induces a significant conformational change and impacts folding, we carried out CD and A-SEC in the presence of zinc (Fig. 3.2.5).

The N-terminal domains were incubated with 0.02 mM zinc sulfate and then CD was carried out on these domains in the presence of 0.01 mM zinc sulfate (Fig. 3.2.5a) [83]. As demonstrated in Figure 3.2.5a, the addition of zinc to SgK223 (in purple) and SgK269 (in salmon) N-terminal domains did not alter the CD spectra compared to the spectra of the N-terminal domains in the absence of zinc (shown in black). This indicates that zinc is not inducing secondary structural changes in SgK269 and SgK223 N-terminal domains and these domains most likely do not bind zinc.

Additionally, A-SEC was carried out on the N-terminal domains pre-incubated with 0.02 mM zinc sulfate and in the presence of 0.01 mM zinc sulfate, to investigate if the addition of zinc altered their elution profiles and thus, conformations. SgK223 N-terminal domain pre-incubated with zinc sulfate eluted at 11.9 mL and SgK269 N-terminal domain pre-incubated with zinc sulfate eluted at 12.5 mL (corresponding to MWs of ~53 kDa and ~45 kDa, respectively) (Fig. 3.2.5b). Compared to the SgK269 N-terminal domain elution profile in the absence of zinc, there is a small increase in the elution volume upon incubation with zinc (~1 mL). However, this experiment was only carried out once, so no conclusions can be made about the cause of the increase in elution volume.

3.2.6 Future directions

Further experiments need to be carried out to investigate the structure and function of the N-terminal domains of SgK223 and SgK269. Although the current data suggest these domains do not bind zinc, these experiments were only

carried out once, so we cannot draw any conclusions until these are repeated. Future experiments could involve testing the binding of a variety of divalent ions, such as manganese or magnesium. CD and A-SEC could be used as a readout of ion binding with the N-terminal domains, demonstrating if ion binding impacts the secondary structure of these domains.

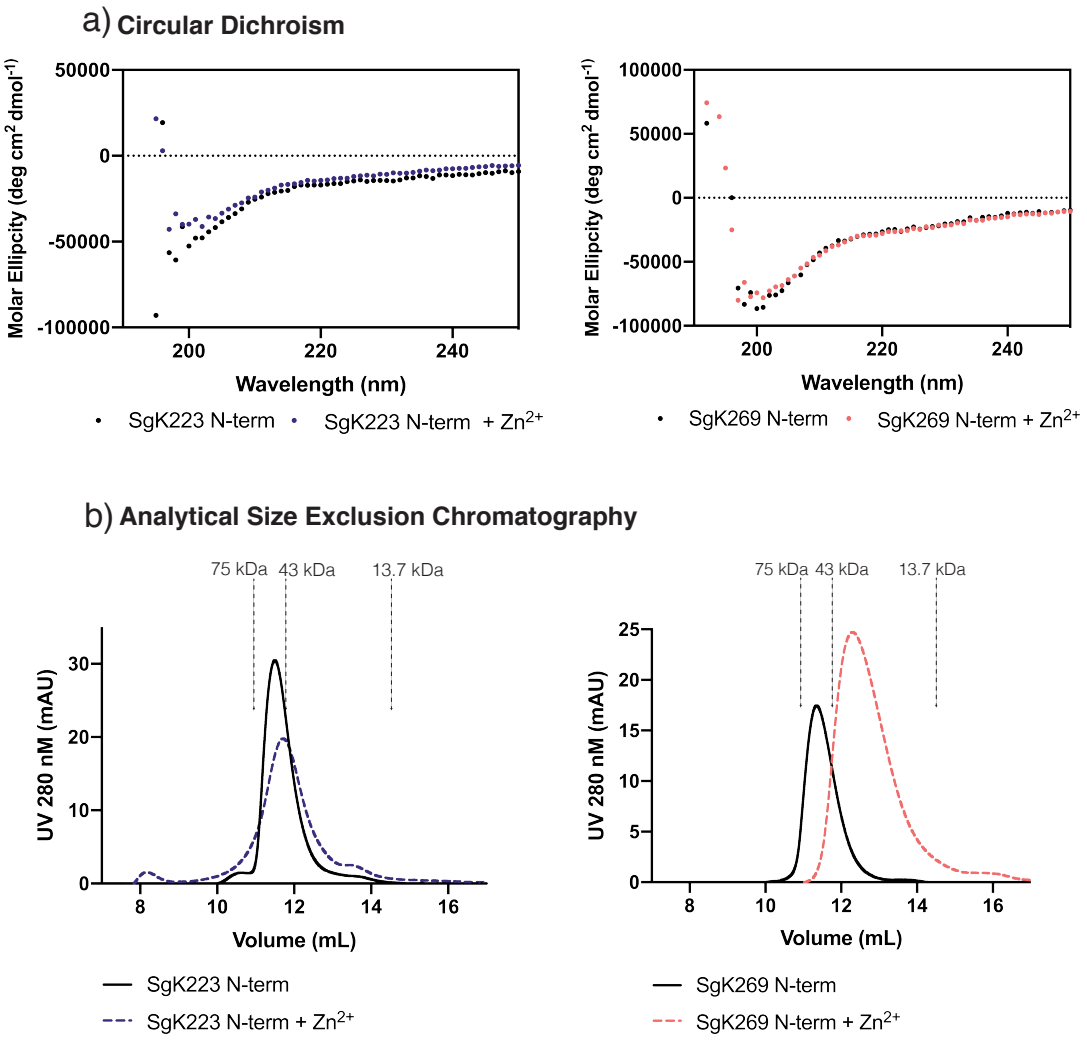


Figure 3.2.5. SgK223 and SgK269 N-terminal domain ion binding studies.
a) Circular dichroism spectra of SgK223 N-terminal domain and SgK269 N-terminal domain in the presence and absence of zinc.
b) Analytical SEC elution profile of SgK223 N-terminal domain and SgK269 N-terminal domain in the presence and absence of zinc. MW standards are shown in dotted lines and the MWs are listed above the lines.

3.3 Characterisation of SgK223 and SgK269 homo- and heterotypic associations

3.3.1 The dimerisation interface of SgK223 and SgK269 is formed by the N- and C-terminal regulatory helices

Recent studies from our lab and others have provided structural insight into the C-terminal pseudokinase domain and flanking regulatory helices of SgK223 and SgK269 (Fig. 3.3.1) [6-8]. These structures highlighted a conserved mechanism of dimerisation that drives homo- and hetero-association of SgK223 and SgK269. The novel dimerisation domain assembles in an 'XXL' formation through the helix N-terminal to the pseudokinase domain, the αN1, and the four helices C-terminal to the pseudokinase domain, αJ, αK, αL and αM (Fig. 3.3.1b) [6-8]. The two longest helices of the dimerisation domains, the αN1 and the αK, from one monomer interact with the αN1 and αK helices from another monomer (Fig. 3.3.1b, αN1 in blue and αK in salmon) [6-8]. When the αN1 helix is deleted from SgK269 *in vitro*, SgK269 can no longer dimerise, thus confirming the importance of this helix for dimerisation [1]. Furthermore, as described in Table 3.3.1, various single mutations of the dimerisation domain helices and pseudokinase domain have identified hot spots of residues that drive SgK223 and SgK269 dimerisation [6-8]. The residues that upon mutation were demonstrated to most successfully disrupt dimerisation of SgK223, L955A and L966A of the αN1 and A1367E from the αK helix (mapped onto Figure 3.3.1b), form important contacts in the core of this large hydrophobic interface [6]. These residues are conserved in SgK269 as Ile1290, Leu1301 and Ala1707, occupying the same positions in the SgK269 dimerisation interface (Table 3.3.1, Fig. 3.3.1b) [6]. Additionally, it was demonstrated that mutation of Ala1707, to an aspartate residue, disrupted dimerisation of SgK269 (Table 3.3.1, Fig. 3.3.1b)[7].

The three shorter C-terminal helices, αJ, αL and αM, form the junction connecting the dimerisation interface to the pseudokinase domain, anchoring and orienting the dimerisation domain in respect the pseudokinase domain (Fig. 3.3.1b, αJ was unresolved in the SgK223 structure, αL is shown in purple and αM is shown in

yellow). There are a number of critical residues that stabilise these helices. These residues are conserved between SgK223 and SgK269. Phe1366 of the αK of SgK223, which upon alanine mutation disrupts SgK223 dimerisation, forms critical contacts with hydrophobic residues of the αL to stabilise this interface and is conserved as Phe1706 in SgK269 (Table 3.3.1) [6, 7]. Trp1382 of SgK223 αL helix stacks between a number of aromatic residues from the αN1 helix and the loop connecting the αC helix to the β4 strand of the pseudokinase N-lobe, and alanine mutation of this residue also disrupts dimerisation [6]. Similarly, the crystal structure of SgK269 demonstrated that the equivalent Trp1722 lies in the same position and stacks between the conserved aromatic residues of the αN1 helix, stabilising the pseudokinase domain with respect to the dimerisation domain [6, 7].

3.3.2 The oligomerisation interface of SgK223 and SgK269 involves the pseudokinase domain αG helix and A-loop

In addition to dimerisation, SgK223 and SgK269 can homo and hetero-oligomerise to form higher molecular weight complexes, as evident through size exclusion chromatography (SEC) experiments [1]. However, until recently the interface driving these interactions was unknown. The SgK223 crystal structure indicated the presence of another interface consisting of residues from the αG, αEF and end of the A-loop of the C-lobe that pack against the N-lobe of the symmetry related molecule (Fig. 3.3.1b) [6]. The conformation of the αG in SgK269 and SgK223 is remarkably similar in their structures and the sequence is highly conserved [6-8]. Analogous to the dimerisation interface, through alanine mutagenesis the role of the αG helix and A-loop of the oligomerisation interface was verified. Additionally, SgK223 αG and A-loop mutants, I1243A, F1271A and Y12812A, were incapable of forming higher molecular weight complexes compared to SgK223 WT, as verified through AUC (Table 3.3.1) [6]. Thus, these regions of SgK223 pseudokinase domain are critical for oligomerisation of SgK223 [6]. Unlike SgK223, crystal packing of SgK269 did not reveal the interface responsible for SgK269 oligomerisation. However, alanine mutagenesis

was carried out to the αG and A-loop, mutating residues Ile1581, Phe1609 and Tyr1620 based the SgK223 structure. These mutants showed a significant reduction of higher order oligomerisation, compared to SgK269 WT, demonstrated through AUC [6].

3.3.3 Functional implications of SgK223 and SgK269 homo- and hetero-association in cells

It is unknown what the importance of SgK223 and SgK269 dimerisation is for their function in cells. However, expression of SgK223 and SgK269 mutants that were unable to dimerise lead to decreased MCF-10A cell migration compared to overexpression of SgK223 and SgK269 WT [1, 6]. This has yet to be demonstrated for dimerisation of SgK269. Additionally, disrupting dimerisation of SgK223 reduced the ability of SgK223 to activate the kinase Csk, compared to SgK223 WT [8]. As mentioned in Chapter 1, SgK223 associates with the SH2 domain of Csk upon phosphorylation of Tyr411 in the PEST linker of SgK223. This suggests that the interaction of two Csk kinase domains, through their association with dimerised SgK223, could lead to trans-activation of Csk and enhanced kinase activity [8].

The functional relevance of SgK223 and SgK269 oligomerisation requires further characterisation. However, expression of an oligomerisation defective mutant of SgK223 resulted in a decrease in cell migration compared to expression of SgK223 WT, indicating oligomerisation of SgK223 could be important for its role in signalling pathways [6].

3.3.4 Investigations into the homo and hetero-associations of SgK223 and SgK269 in this Chapter

In this Chapter, to investigate the structural conservation between SgK223 and SgK269 pseudokinase and dimerisation domains, we mutated residues from the dimerisation domain of SgK269. Much like SgK223, we aimed to determine hot spots of residues that are critical for driving dimerisation of SgK269, to then investigate the importance of dimerisation for the role of SgK269 in cells.

Yet to be confirmed experimentally is whether the complementary interacting interface for SgK223 and SgK269 homo- and hetero-oligomerisation is the N-lobe of the pseudokinase domain of the two proteins. Thus, in this Chapter we will investigate the role of the N-lobe in SgK223 and SgK269 homo-and hetero-oligomerisation.

Finally, using our knowledge of the hot spots that drive dimerisation and oligomerisation of SgK223 and SgK269, we aim to combine the mutants that disrupted dimerisation and oligomerisation, to generate a mutant that is monomeric. We will use this mutant to investigate the role of homo-and hetero-association of SgK223 and SgK269 in cells in Chapter 5.

Table 3.3.1. Summary of SgK223 and SgK269 dimerisation and oligomerisation mutants in this study and from the literature.

SgK223				**SgK269**			
Literature mutants	Location	Mutants in this study	Location	Literature mutants	Location	Mutants in this study	Location
Dimerisation interface							
L955A [6]	αN1	L955E, A1367D	αN1, αK	A1707D [7]	αK	I1290A	αN1
L966A [6]	αN1	L955E, F1271A, A1367D	αN1, αG, αK			L1301A	αN1
A1367E [8]	αK					F1706A	αK
F1366A [6]	αK					W1722 A	αL
W1382A [6]	αL					I1290E, A1707D	αN1, αK
						I1290E, F1609A, A1707D	αN1, αG, αK
Oligomerisation interface							
I1243A [6]	A-loop	F992A	N-lobe (loop between αN1 and β1)	I1581A [6]	A-loop	F1327A	N-lobe (loop between αN1 and β1)
F1271A [6]	αG	C996R	N-lobe (loop between αN1 and β1)	F1609A [6]	αG	S1331R	N-lobe (loop between αN1 and β1)
Y1282A [6]	αG	P999G	N-lobe (β1)	Y1620A [6]	αG	P1334G	N-lobe (β1)
		V1089R	N-lobe (β5)			V1460R	N-lobe (β5)

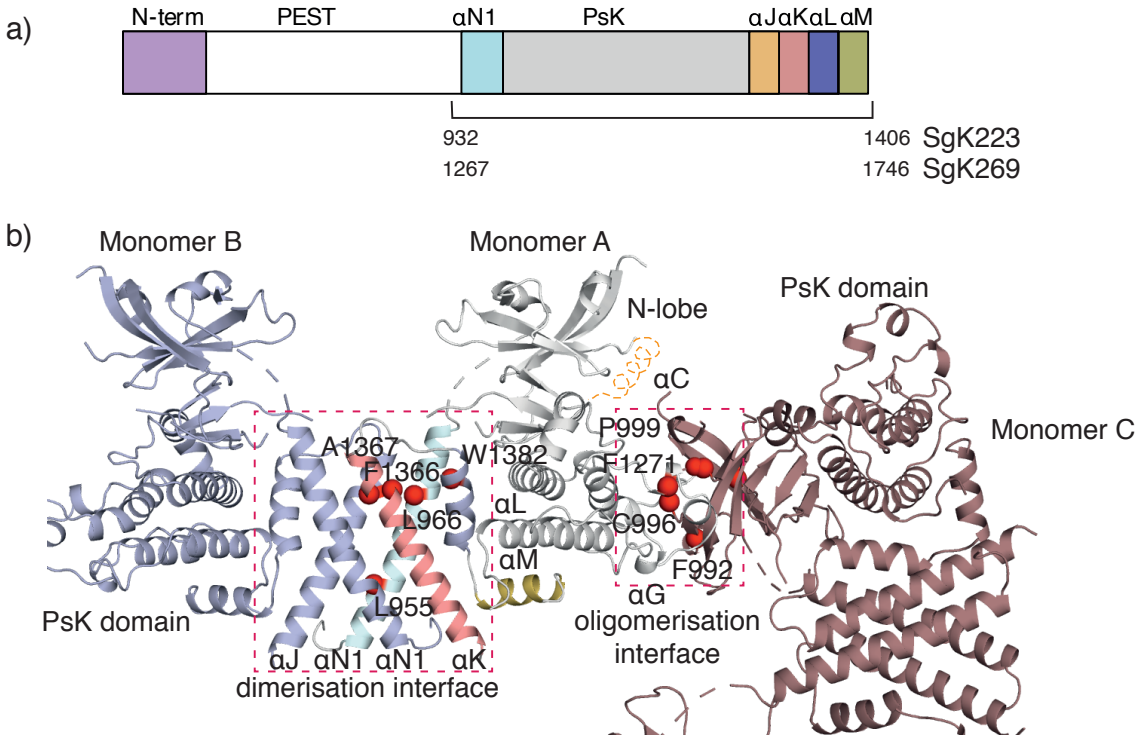


Figure 3.3.1. Mutagenesis of SgK223 and SgK269.
a) Domain layout of SgK223 and SgK269 demonstrating the domain boundaries of the constructs that were mutated below.
b) Structure of the pseudokinase domain and flanking helices of SgK223 (PDB: 5VE6) demonstrating residues that were mutated in red spheres and labelled.

3.4 Results: Biochemical characterisation of SgK223 and SgK269 homo- and hetero-associations

The mutants generated and analysed in this study are listed in the Table 3.3.1. Initially, we carried out single site mutagenesis of the SgK269 dimerisation domain, mutating two residues of the αN1 helix (I1290A, L1301A) that were demonstrated to disrupt dimerisation in SgK223 (L955A, L966A), as well as F1706A from the αK and W1722A from the αL, also demonstrated to disrupt dimerisation in SgK223 (F1366A, W1382A). Additionally, we carried out single site mutagenesis of the N-lobe of the pseudokinase domain of SgK223 and SgK269 to investigate if the N-lobe is involved in oligomerisation of SgK223 and SgK269. The mutations, F992A, C996R, P999G and V1089R (F1327A, S1331R, P1334G and V1460R in SgK269) are shown in Figure 3.3.1 and Table 3.3.1.

3.4.1 Cloning, expression, and purification of SgK223 and SgK269 pseudokinase domain and dimerisation domain mutants

SgK223 pseudokinase domain and dimerisation domain mutants (F992A, C996R, P999G, V1089R and L955E/A1367D/F1271A), residues 932-1406 (Fig. 3.4.1a,b and Table 3.3.1), and SgK269 pseudokinase domain and dimerisation domain mutants (I1290A, L1301A, F1706A, W1772A, F1327A, S1331R, P1334G, V1089R and I1290E/A1707D/F1609A), residues 1267-1746 (Fig. 3.4.2a,b and Table 3.3.1), generated by Genscript (for gene details see Table 8.1 in the appendices and for construct details see Table 2.2 in Chapter 2), were fused N-terminally to a His6 tag with a TEV protease cleavage site after the tag and cloned into a pCOLDIV vector for *E.coli* expression. Purification of the mutants was conducted by nickel affinity chromatography (Roche) and size exclusion chromatography (SEC) (Superdex HiLoad 200 16/600 column from GE Healthcare) (Fig. 3.4.1c, d and Fig. 3.4.2c, d). Each of the proteins eluted as a single peak (Fig. 3.4.1 c, d and Fig. 3.4.2c,d). SDS-PAGE confirmed protein purity to ~100% and pure proteins were concentrated to ~4 mg/mL and stored at -80°C.

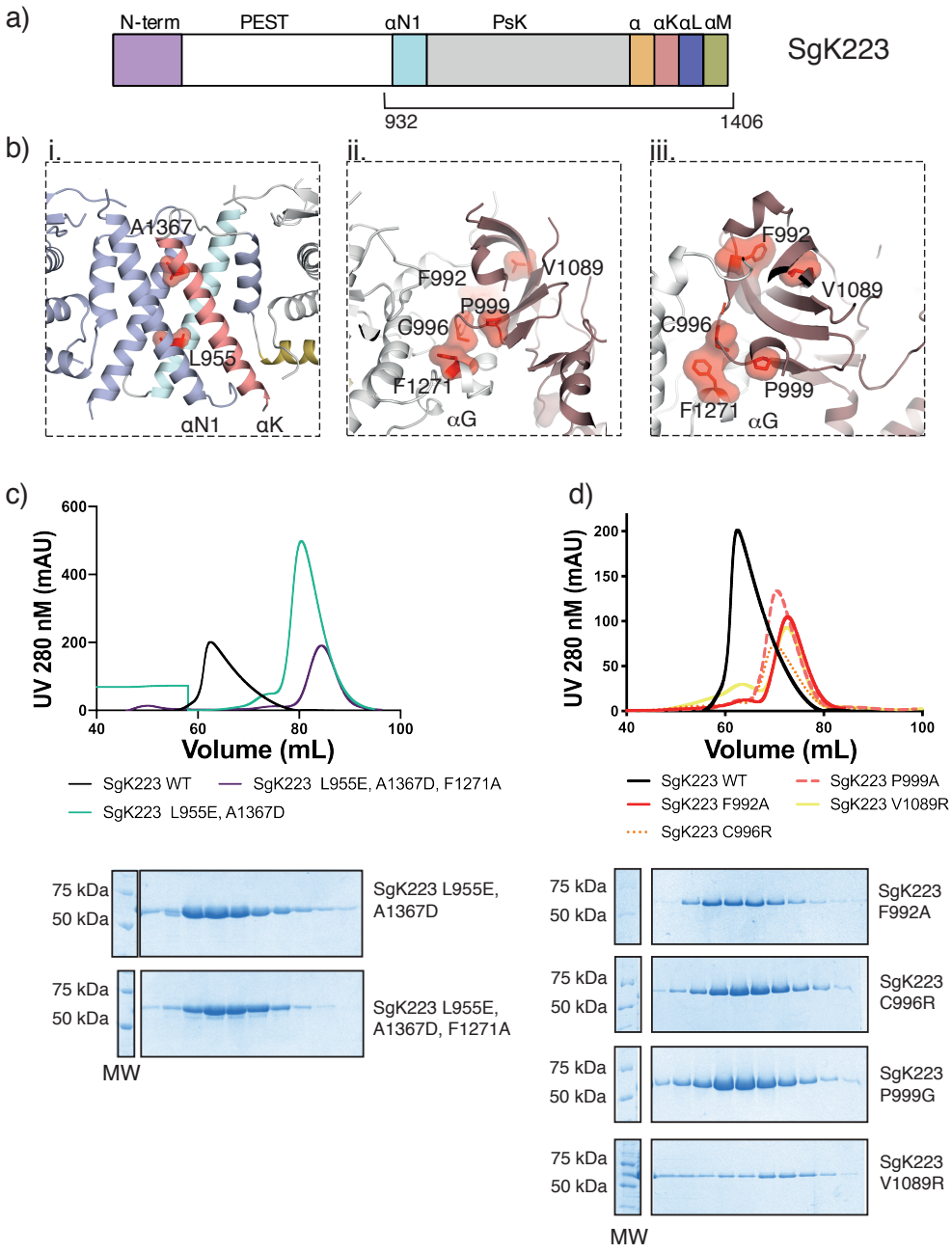


Figure 3.4.1. Purification of SgK223 pseudokinase and dimerisation domain mutants.

a) Domain layout of SgK223, the pseudokinase domain and dimerisation domain are underlined and domain boundaries are labelled below.

b) Crystal structure of SgK223 pseudokinase domain and flanking helices (PDB: 5EV6). Highlighted in red are the residues that were mutated. i. Dimerisation domain mutations and ii. and iii. Pseudokinase domain N-lobe mutations.

c) SEC elution profiles of the SgK223 dimerisation domain mutants. SgK223 WT is shown for reference in black. SDS-PAGE gels of peak fractions below.

d) SEC elution profiles of the SgK223 pseudokinase domain N-lobe oligomerisation mutants. SgK223 WT is shown for reference in black. SDS-PAGE gels of peak fractions below. MW= Molecular Weight marker (BioRad Precision Plus unstained protein standards).

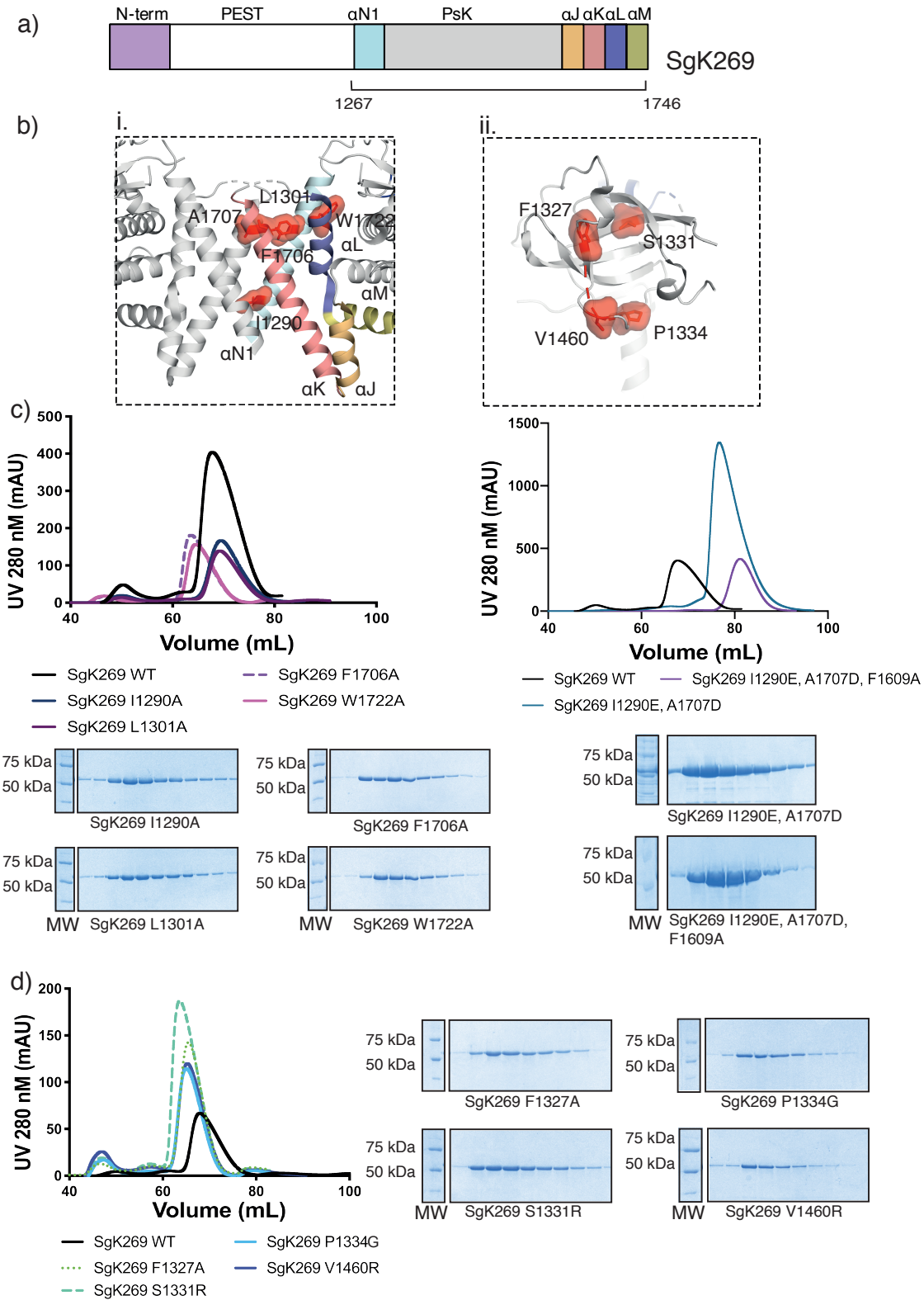

a)
N-term
PEST
αN1
PsK
αJ αK αL αM
SgK269
1267
1746
b)
i.
L1301
A1707
W1722
αL
F1706
αM
I1290
αN1
αK
αJ
ii.
F1327
S1331
V1460
P1334
c)
UV 280 nM (mAU)
Volume (mL)
SgK269 WT
SgK269 I1290A
SgK269 L1301A
SgK269 F1706A
SgK269 W1722A
SgK269 I1290E, A1707D, F1609A
SgK269 I1290E, A1707D
75 kDa
50 kDa
MW
SgK269 I1290A
SgK269 F1706A
SgK269 L1301A
SgK269 W1722A
SgK269 I1290E, A1707D
SgK269 I1290E, A1707D, F1609A
d)
UV 280 nM (mAU)
Volume (mL)
SgK269 WT
SgK269 F1327A
SgK269 S1331R
SgK269 P1334G
SgK269 V1460R
SgK269 F1327A
SgK269 P1334G
SgK269 S1331R
SgK269 V1460R

Figure 3.4.2. Purification of SgK269 pseudokinase and dimerisation domain mutants.
a) Domain layout of SgK269, the pseudokinase domain and dimerisation domain are underlined and domain boundaries are labelled below.
b) Crystal structure of SgK269 pseudokinase domain and flanking helices (PDB: 6BHC). Highlighted in red are the residues that were mutated. i. Dimerisation domain mutations and ii. Pseudokinase domain N-lobe mutations.
c) SEC elution profiles of the SgK269 dimerisation domain mutants. SgK269 WT is shown for reference in black. SDS-PAGE gels of peak fractions below.
d) SEC elution profiles of the SgK269 pseudokinase domain N-lobe oligomerisation mutants. SgK269 WT is shown for reference in black. SDS-PAGE gels of peak fractions on the right. MW= Molecular Weight marker (BioRad Precision Plus unstained protein standards).

3.4.2 Biochemical and mutational analysis of SgK269 dimerisation domain

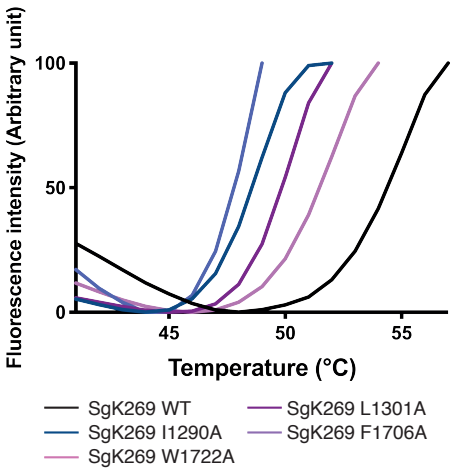
Initially, we characterised the SgK269 dimerisation domain mutants (I1290A, L1301A, F1706A, W1722A). A thermal stability assay (TSA) was carried out to assess the impact of the mutations on the stability of the SgK269 dimer. In this assay, the temperature of a protein solution is steadily increased from 25 °C to 80 °C. As the temperature increases and the protein unfolds, hydrophobic residues in the protein are exposed and SYPRO orange dye can bind to these residues and fluoresce, allowing for fluorescence quantitation [84]. The temperature of the midpoint of protein melting (Tm) is calculated from the sigmoidal curve of temperature vs. SYPRO fluorescence (Fig. 3.4.3a). All mutants demonstrated a reduction of 2-6 °C in their Tm and thus reduction in stability (Fig. 3.4.3a), which may indicate a transition from dimer to monomer, as a dimer is more stable than a monomer.

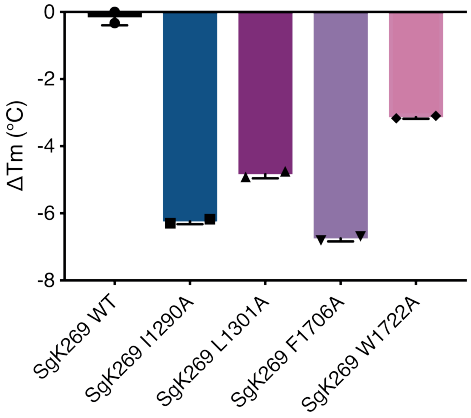
To confirm if the mutants of the SgK269 dimerisation domain are disrupting dimerisation, analytical size exclusion chromatography (A-SEC) was carried out (Fig. 3.4.3b). If the elution volume of the mutants is increased compared to the WT protein, this could indicate the mutants are no longer dimerising, as an increase in SEC elution volume correlates to a decrease in MW of the protein. WT or mutant proteins were loaded at 30 μM concentration in a volume of 110 μL onto a Superdex 200 10/300 column (GE Healthcare) for analysis. Eluted protein was collected and analysed by SDS-PAGE. The SgK269 WT elution profile, carried out as a control, had a major peak eluting at 12 mL and a minor peak eluting just after the void volume at ~8.5 mL (Fig. 3.4.3b). This minor peak represents higher order oligomers, as previously demonstrated by AUC experiments carried out at 20 μM [6]. The elution profiles of the dimerisation domain mutants of SgK269 were very similar to that of the WT elution profile, eluting at 11.9 mL, with a minor peak eluting just after the void volume. This indicates these single mutations are most likely unable to perturb dimerisation of SgK269 (Fig. 3.4.3b). The decrease in stability demonstrated through TSA could be due to these mutations exposing hydrophobic regions in the dimerisation

interface, rather than completely disrupting dimerisation. Table 3.4.1 summarises the behaviour of SgK269 mutants and WT on SEC (Superdex 200 10/300) and AUC. To confirm the dimeric state of the mutants, we carried out AUC, using SgK269 WT protein as a control, to verify if the mutant proteins are capable of forming dimers.

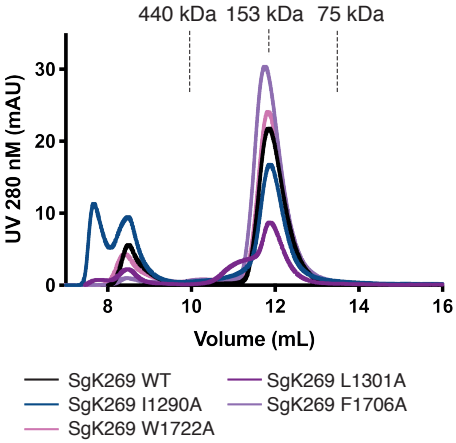
Protein samples were prepared as previously described for sedimentation velocity experiments (see section 3.2.3), all subjected to SEC to remove glycerol from the buffer. To analyse the dimerisation mutants, SgK269 WT, SgK269 I1290A, SgK269 L1301A and SgK269 W1722A were concentrated to 3 μM, as Patel *et al*. previously found that at 3 μM SgK269 forms stable dimers [6]. If these mutations disrupted dimerisation, we would expect a shift in the sedimentation coefficient *(c(s))* distributions to a lower sedimentation coefficient (S), compared to the WT protein. At 3 μM SgK269 WT formed a single peak on the sedimentation coefficient *(c(s))* distribution, at 5.15 S (Fig. 3.4.3c). The mutants also formed a single peak on the sedimentation coefficient *(c(s))* distribution, at ~5-5.15 S (Fig. 3.4.3c). Thus, as the sedimentation coefficients and peaks on the distribution were very similar to that of SgK269 WT, these mutations could not disrupt dimerisation of SgK269 (for data fitting see appendices, Fig. 8.3b).

a)

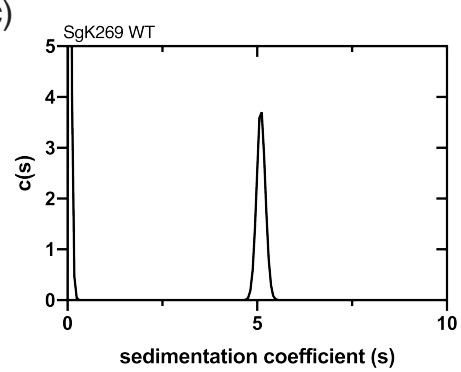

Fluorescence intensity (Arbitrary unit)
100
50
0
45
50
55
Temperature (°C)
SgK269 WT
SgK269 I1290A
SgK269 W1722A
SgK269 L1301A
SgK269 F1706A

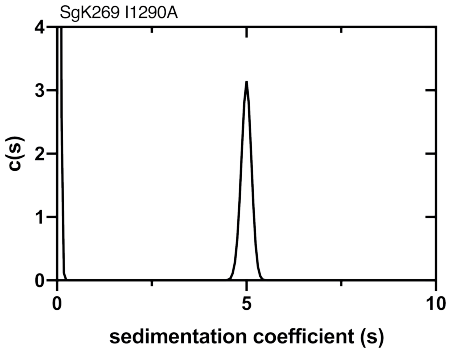

ΔTm (°C)
0
-2
-4
-6
-8
SgK269 WT
SgK269 I1290A
SgK269 L1301A
SgK269 F1706A
SgK269 W1722A

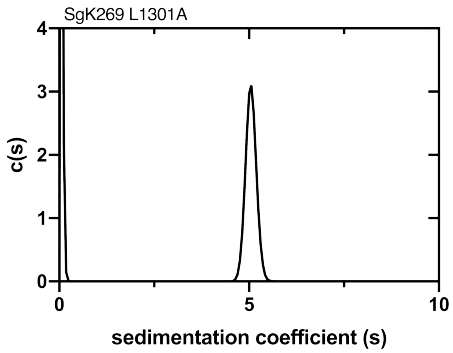
b)


440 kDa
153 kDa
75 kDa
UV 280 nM (mAU)
30
20
10
0
8
10
12
14
16
Volume (mL)
SgK269 WT
SgK269 I1290A
SgK269 W1722A
SgK269 L1301A
SgK269 F1706A

c)


SgK269 WT
c(s)
sedimentation coefficient (s)


SgK269 I1290A
c(s)
sedimentation coefficient (s)


SgK269 L1301A
c(s)
sedimentation coefficient (s)

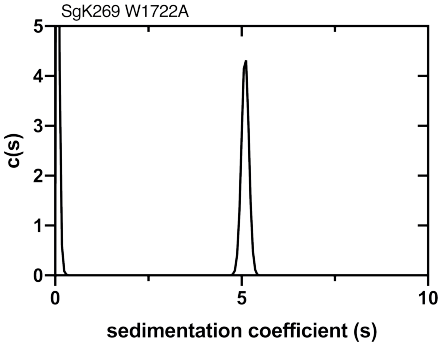

SgK269 W1722A
c(s)
sedimentation coefficient (s)

Figure 3.4.3. Biochemical analysis of SgK269 dimerisation domain mutants.
a) Thermal stability assay (TSA) of SgK269 dimerisation domain mutants. Tm's of the mutants are normalised to the Tm of SgK269 WT.
b) Analytical SEC elution profile of SgK269 dimerisation domain mutants. MW standards are represented by dotted lines and the MWs are listed above the lines.
c) Analytical ultracentrifugation sedimentation coefficient (*c(s))* distribution of SgK269 dimerisation domain mutants.

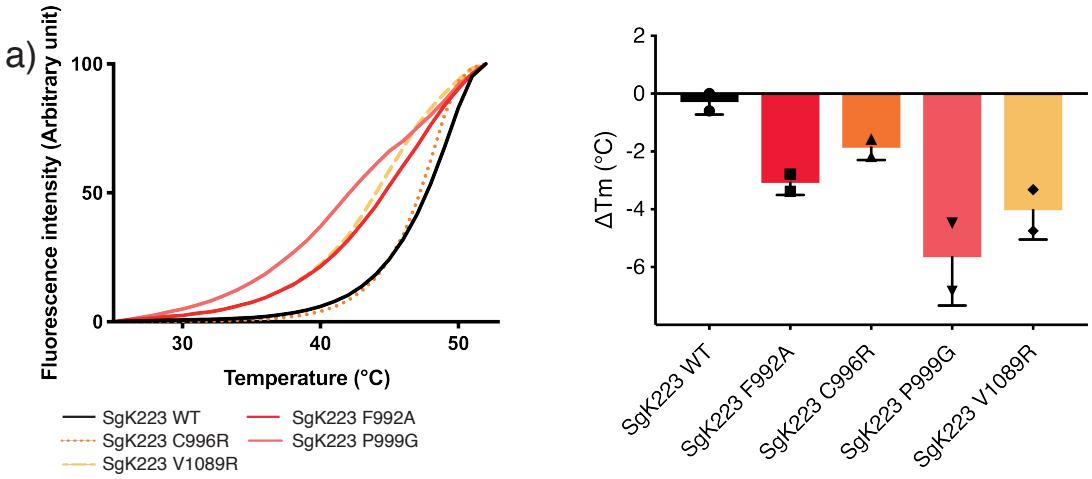
3.4.3 Investigating the role of the SgK223 N-lobe in homo-oligomerisation and hetero-oligomerisation with SgK269

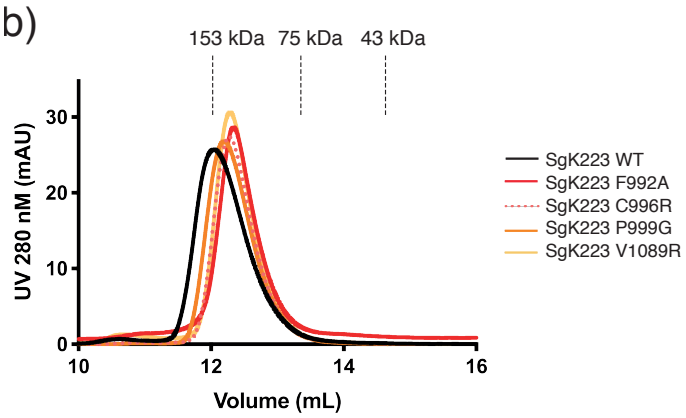
Next, we characterised the SgK223 N-lobe mutants (F992A, C996R, P999G, V1089R) to investigate their role in homo- oligomerisation of SgK223 and hetero-oligomerisation of SgK223 and SgK269. TSA analysis revealed mutations within the N-lobe of SgK223 decreased the overall stability of the protein compared to SgK223 WT, reducing the Tm between 2- 6 °C (Fig. 3.4.4a). This could indicate that the mutations have altered the oligomerisation state and in doing so, the stability of the protein.

To confirm the role of the N-lobe in oligomer formation, A-SEC was carried out (Fig. 3.4.4b). WT or mutant SgK223 proteins were loaded at 30 μM concentration in a volume of 110 μL onto a Superdex 200 10/300 column (GE Healthcare) for analysis. Eluted protein was collected and analysed by SDS-PAGE. A-SEC of the SgK223 mutants demonstrated a sharper elution profile with an elution volume of ~12.3 mL, suggesting a reduction in oligomeric species compared to SgK223 WT (Fig. 3.4.4b). Table 3.4.1 summarises the behaviour of SgK223 mutants and WT on SEC (Superdex 200 10/300) and AUC, mutants highlighted in blue demonstrated disrupted oligomerisation.

To further investigate the oligomeric state of the SgK223 N-lobe mutants, sedimentation velocity experiments were carried out using SgK223 WT as a control. Mutant and WT proteins were buffer exchanged into a glycerol free buffer and concentrated to 20 μM (AUC was only carried out on F992A, C996R and V1089R for technical reasons) (see Chapter 2 Table 2.1 for buffer recipes). If

these mutations disrupted oligomerisation, we would expect a shift in the sedimentation coefficient *(c(s))* distributions compared to the WT protein. The mutants of the N-lobe of SgK223, F992A, C996R, and V1089R, all formed a singular peak with sedimentation coefficients of 5.15 S, 5.3 S and 5.15 S, respectively (Fig. 3.4.4c). Comparatively, the SgK223 WT protein formed two peaks on the sedimentation coefficient *(c(s))* distribution, the first peak at ~5.6 S and the second peak at ~7 S (Fig. 3.4.4c). The second peak of the WT protein is oligomers, and the loss of this peak in all the mutants, as well as a decrease in the sedimentation coefficients, demonstrated that these mutations disrupt homo-oligomerisation of SgK223 [6] (for data fitting see appendices, Fig. 8.3a).


a)
Fluorescence intensity (Arbitrary unit)
100
50
0
30
40
50
Temperature (°C)
SgK223 WT
SgK223 C996R
SgK223 V1089R
SgK223 F992A
SgK223 P999G
ΔTm (°C)
2
0
-2
-4
-6
SgK223 WT
SgK223 F992A
SgK223 C996R
SgK223 P999G
SgK223 V1089R


b)
153 kDa
75 kDa
43 kDa
UV 280 nM (mAU)
30
20
10
0
10
12
14
16
Volume (mL)
SgK223 WT
SgK223 F992A
SgK223 C996R
SgK223 P999G
SgK223 V1089R

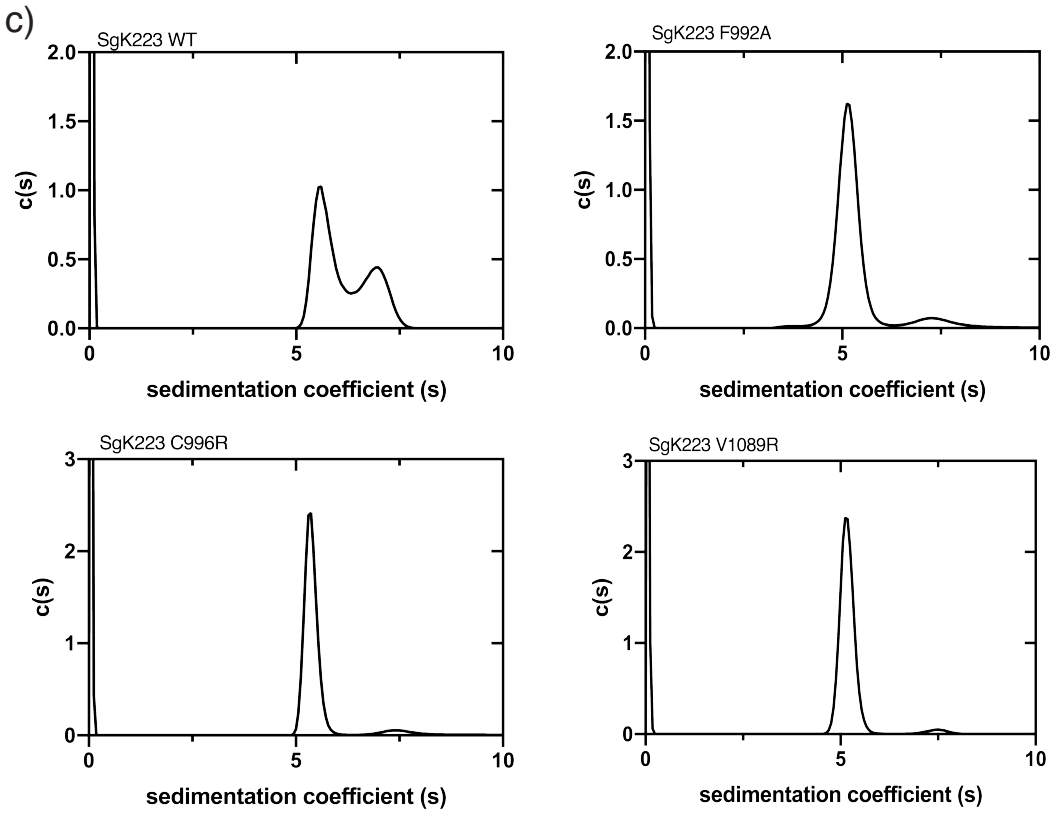

c)
SgK223 WT
SgK223 F992A
SgK223 C996R
SgK223 V1089R
c(s)
sedimentation coefficient (s)
0
5
10

Figure 3.4.4. Biochemical analysis of SgK223 pseudokinase and dimerisation domain mutants.
a) TSA of SgK223 N-lobe oligomerisation mutants. Tm's of the mutants are normalised to the Tm of SgK223 WT.
b) Analytical SEC elution profiles of SgK223 N-lobe oligomerisation mutants. MW standards are represented by dotted lines and the MWs are listed above the lines.
c) Analytical ultracentrifugation sedimentation coefficient *c(s)* distribution of SgK223 N-lobe oligomerisation mutants.

Table 3.4.1. Summary of the behaviour of SgK223 and SgK269 mutants on A-SEC and AUC.

Mutant	Elution volume on A-SEC (mL)	Sedimentation coefficient (S)	Monomer/dimer /oligomer
SgK223			
WT	12	5.6, 7	Dimer/oligomer
F992A	12.4	5.15	Dimer
C996R	12.3	5.3	Dimer
P999G	12.3	5.15	Dimer
V1089R	12.3	-	Dimer
L955E, A1367D	12.5, 14	-	Monomer
L955E, A1367D, F1271A	12.5, 14.1	-	Monomer
SgK269			
WT	8.6, 11.9	5.15	Dimer/oligomer
I1290A	7.7, 8.5, 12	5	Dimer/oligomer
L1301A	8.6, 12	5	Dimer/oligomer
F1706A	8.6, 11.8	-	Dimer/oligomer
W1722A	8.4, 12	5.15	Dimer/oligomer
F1327A	8.6, 11.9	-	Dimer/oligomer
S1331R	8.6, 11.9	-	Dimer/oligomer
P1334G	8.6, 11.9	-	Dimer/oligomer
V1460R	8.4, 12.1	-	Dimer/oligomer
I1290E, A1707D	13.9	-	Monomer
I1290E, A1707D, F1609A	13.8	-	Monomer

To investigate if the SgK223 N-lobe mutations have an impact on SgK223 and SgK269 hetero-oligomerisation, each mutant was pre-incubated with SgK269 WT to a final protein concentration of 30 μM and analysed through A-SEC (S200 10/300 GE Healthcare). Pre-incubation of SgK223 WT and SgK269 WT resulted in a marked shift in the elution profile to a larger elution volume, compared to the proteins analysed alone, indicating formation of higher order oligomers (Fig. 3.4.5). In comparison, the marked shift in elution profile towards higher order oligomers was not observed when the SgK223 N-lobe mutants were incubated with SgK269 WT. This indicates these mutants have a reduction in hetero-oligomerisation with SgK269 WT (Fig. 3.4.5a, b, c, d). The SgK223 V1089R mutant and SgK269 WT elution profile was the most similar to that of the single proteins eluting alone, highlighting this mutation to be the most effective in disrupting hetero-oligomerisation of SgK223 and SgK269 (Fig. 3.4.5d). The SDS-PAGE showed that over the elution peak, SgK223 mutants and SgK269 WT never appear in a 1:1 ratio (Fig. 3.4.5, SDS-PAGE below curves). It would be expected that if the mutants were forming a higher order complex with SgK269 WT, the SDS-PAGE would show equal amounts of both proteins over the elution peak, as present in the SgK223 WT-SgK269 WT SDS-PAGE (Fig. 3.4.5a).

These results indicate that the interacting interfaces for SgK223 homo oligomerisation is the αG and A-loop of the pseudokinase domain and the N-lobe of the pseudokinase domain. Additionally, these interfaces are also involved in SgK223 hetero-oligomerisation with the αG and A-loop of SgK269.

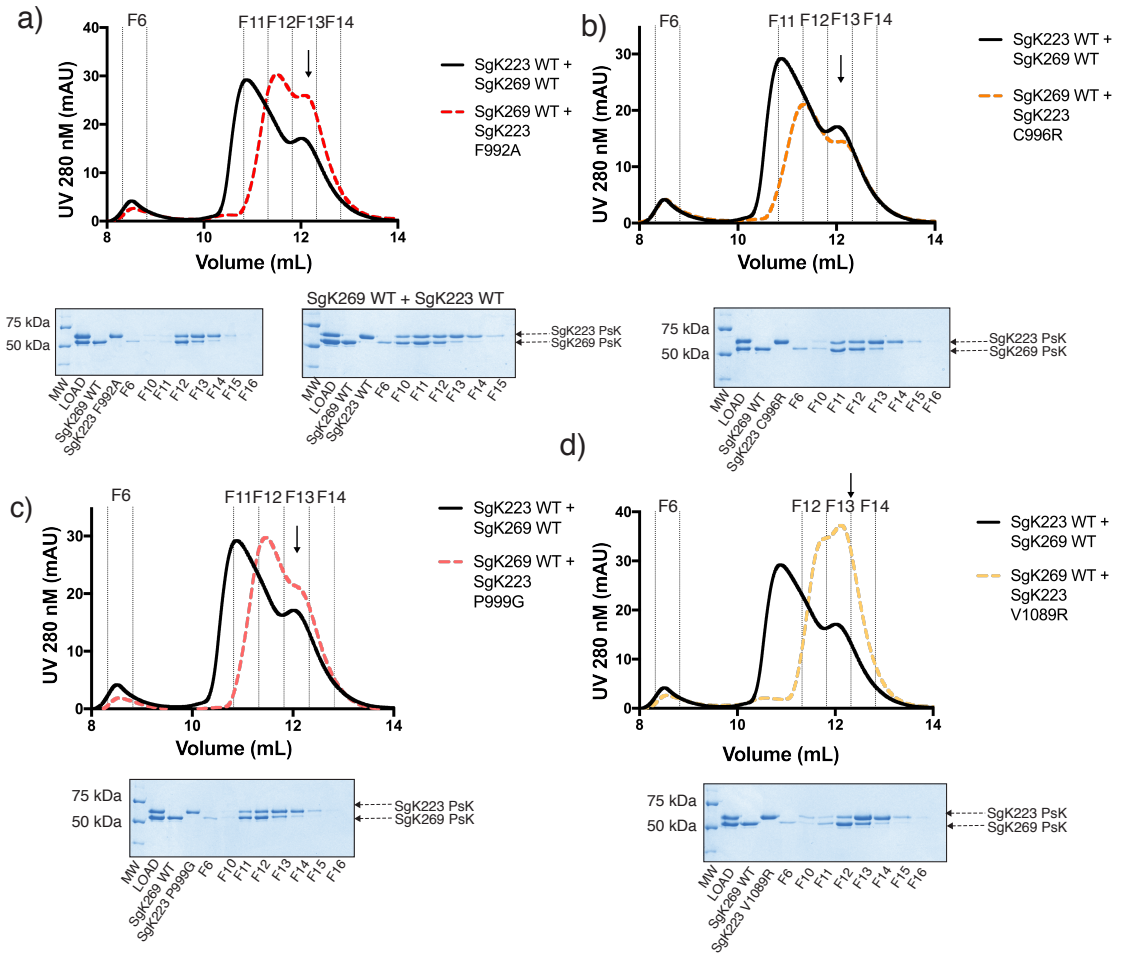


Figure. 3.4.5. Analytical SEC of the hetero-complexation of SgK223 N-lobe oligomerisation mutants with SgK269 WT.

a) SEC elution profile of the complex between SgK223 F992A mutant and SgK269 WT (red dashed line), compared to SgK223 WT and SgK269 WT complex (black line shown in all chromatograms). Black arrow indicates peak of SgK223 dimers, shown in all chromatograms. SDS-PAGE of fractions from SEC elution profiles (WT-WT gel included as a comparison).

b) SEC elution profile of the complex between SgK223 C996R mutant and SgK269 WT (orange dashed line). SDS-PAGE of fractions from SEC elution profiles.

c) SEC elution profile of the complex between SgK223 P999G mutant and SgK269 WT (pink dashed line). SDS-PAGE of fractions from SEC elution profiles.

d) SEC elution profile of the complex between SgK223 V1089R mutant and SgK269 WT (yellow dashed line). SDS-PAGE of fractions from SEC elution profiles.

3.4.4 Investigating the role of SgK269 N-lobe in homo-oligomerisation and hetero-oligomerisation with SgK223

To investigate homo- and hetero-oligomerisation of SgK269, mutants were designed at the N-lobe of SgK269 (F1327A, S1331R, P1334G, V1460R) corresponding to the mutants at the SgK223 N-lobe (Fig. 3.4.2b). TSA analysis demonstrated a reduction of 4-6 °C in Tm of the mutants, compared to SgK269 WT, indicative of a loss in stability of these mutants (Fig. 3.4.6a). Analysis of the mutants by A-SEC (Superdex 200 10/300 GE Healthcare) at 30 μM concentration demonstrated that the SgK269 N-lobe mutants behaved similarly to SgK269 WT, with their elution peaks at 11.9 mL, suggesting the peaks are a mixture of dimers and higher order oligomers (Fig. 3.4.6b) [6]. As there was no change evident in the elution volume of the mutants, we did not carry out AUC, as we predicted that these mutants did not disrupt homo-oligomerisation. Table 3.4.1 summarises the behaviour of SgK269 mutants on SEC.

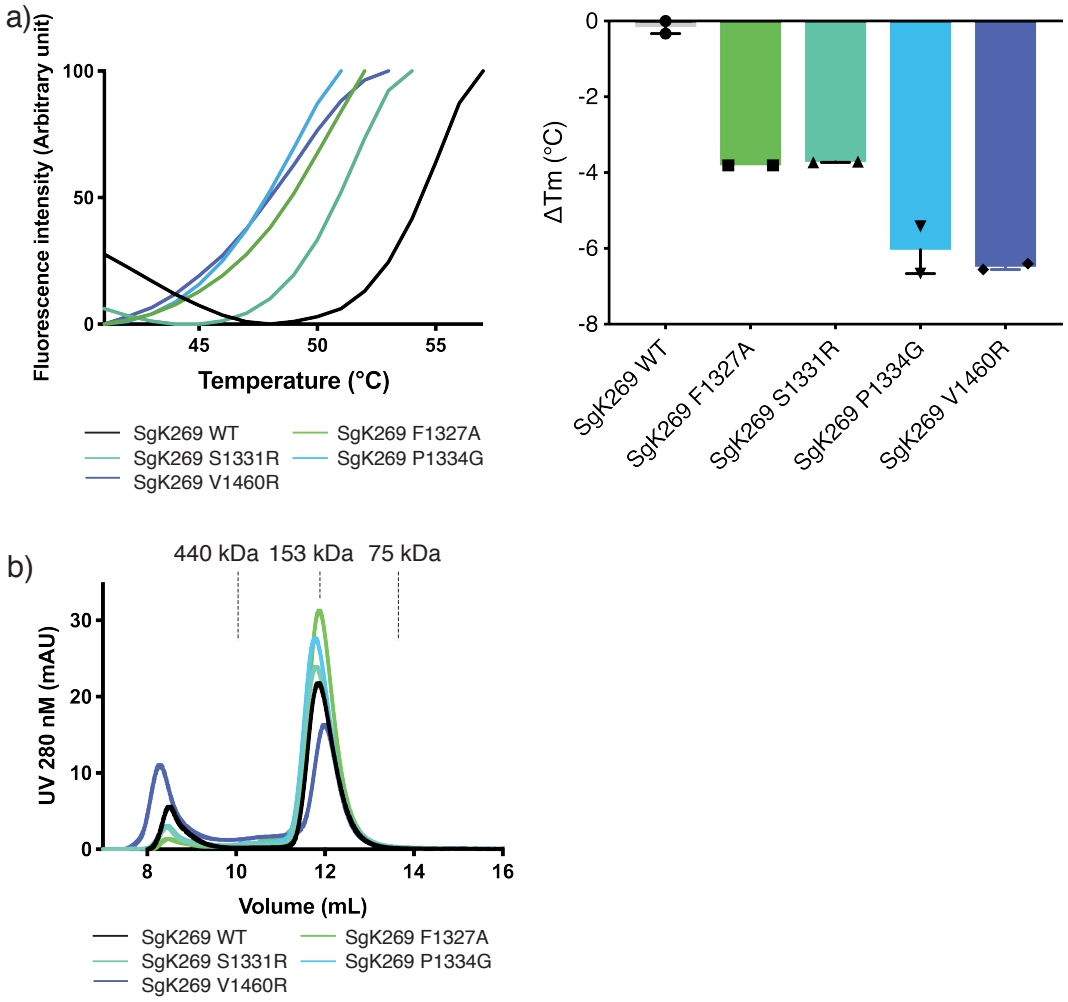


Figure 3.4.6. Biochemical analysis of SgK269 pseudokinase and dimerisation domain mutants.

a) Thermal stability assay (TSA) of SgK269 N-lobe oligomerisation mutants. Tm's of the mutants are normalised to the Tm of SgK269 WT.

b) Analytical SEC elution profile of SgK269 N-lobe oligomerisation mutants. MW standards are represented by dotted lines and the MWs are listed above the lines.

To investigate the role of the N-lobe in hetero-oligomerisation, SgK269 N-lobe mutants were incubated with SgK223 WT to a final protein concentration of 30 μM and analysed by A-SEC, as previously described (see section 3.4.3). The elution profiles of the N-lobe mutants in complex with SgK223 WT eluted as the SgK269 WT-SgK223 WT control (Fig. 3.4.7). The SDS-PAGE gels showed a 1:1 stoichiometry of SgK269 mutants to SgK223 WT, similar to the SgK269 WT-SgK223 WT SDS-PAGE, indicating the mutants formed higher order hetero-complexes as both proteins were present equally in the elution peak (Fig. 3.4.7a, b, c, d). Thus, the interface that drives oligomer formation of SgK269 remains unclear.

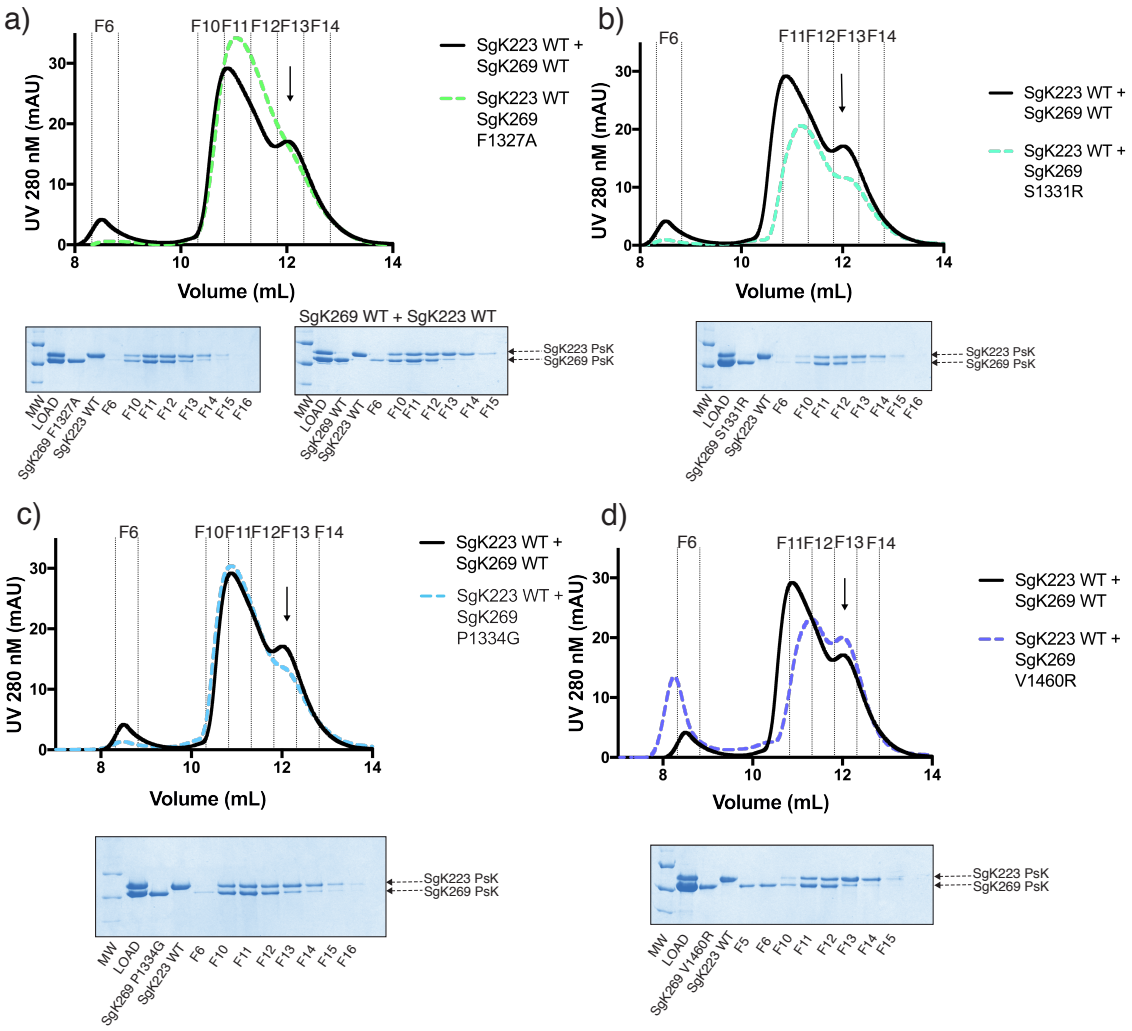


Figure 3.4.7. Analytical SEC analysis of the hetero-complexation of SgK269 N-lobe oligomerisation mutants with SgK223 WT.

a) SEC elution profile of the complex between SgK269 F1327A mutant and SgK223 WT (green dashed line), compared to SgK223 WT and SgK269 WT complex (black line, shown in all chromatograms). Black arrow indicates the peak of SgK269 dimers, shown in all chromatograms. SDS-PAGE of fractions from SEC elution profiles. (WT-WT gel included as a comparison).

b) SEC elution profile of the complex between SgK269 S1331R mutant and SgK223 WT (aqua dashed line). SDS-PAGE of fractions from SEC elution profiles.

c) SEC elution profile of the complex between SgK269 P1334G mutant and SgK223 WT (blue dashed line). SDS-PAGE of fractions from SEC elution profiles.

d) SEC elution profile of the complex between SgK269 V1460R mutant and SgK223 WT (purple dashed line). SDS-PAGE of fractions from SEC elution profiles.

3.4.5 Biochemical characterisation of SgK223 and SgK269 mutants that disrupt dimerisation and oligomerisation

As an overarching aim of our studies is to understand the importance of homo- and hetero-association of SgK223 and SgK269 for their role in cell signalling, we aimed to design mutants to abolish both dimerisation and oligomerisation of SgK223 and SgK269 that could be expressed in human cells for functional assays. To this end, we combined two mutations that have previously been found to individually disrupt dimerisation in recombinant protein studies, and a mutant of the αG found to disrupt oligomerisation [6-8].

The SgK223 double mutant combines two dimerisation domain mutations, one from the αN1 helix and one from the αK helix, L955E and A1367D, respectively (Fig. 3.4.1b, c). The SgK223 triple mutant combines these two dimerisation domain mutations with the oligomerisation mutant from the αG helix, F1271A, to form a completely monomeric SgK223 (Fig. 3.4.1b, c) [6].

The SgK269 double mutant combines the equivalent two dimerisation domain mutations from the αN1 helix and the αK helix, I1290E and A1707D, respectively (Fig. 3.4.2b, c). The SgK269 triple mutant combines these two dimerisation domain mutations with the oligomerisation mutant from the αG helix, F1609A, to form a completely monomeric SgK269 (Fig. 3.4.2b, c) [6].

Based on analytical SEC elution profiles of the proteins at 30 μM concentration, approximately 70% of the SgK223 double mutant eluted as a monomer with an elution volume of 14 mL and 30% eluted as the WT (dimeric) with an elution volume of 12.5 mL (Fig. 3.4.8a). The triple mutant of SgK223 eluted similarly, approximately 65% as a monomer and 35% as a dimer (Fig. 3.4.8a). Both SgK269 mutants eluted as monomers with an elution volume of ~14 mL, demonstrating a marked shift to a larger elution volume compared to the WT protein, with an elution volume of 12 mL, indicative of a decrease in MW (Fig. 3.4.8b). Table 3.4.1 summarises the behaviour of SgK223 and SgK269 mutants

on SEC (Superdex 200 10/300), mutants highlighted in green demonstrated disrupted dimerisation and oligomerisation.

AUC analysis was not required as it is evident that these mutations to the pseudokinase and dimerisation domains successfully disrupt dimerisation and oligomerisation as shown through A-SEC.

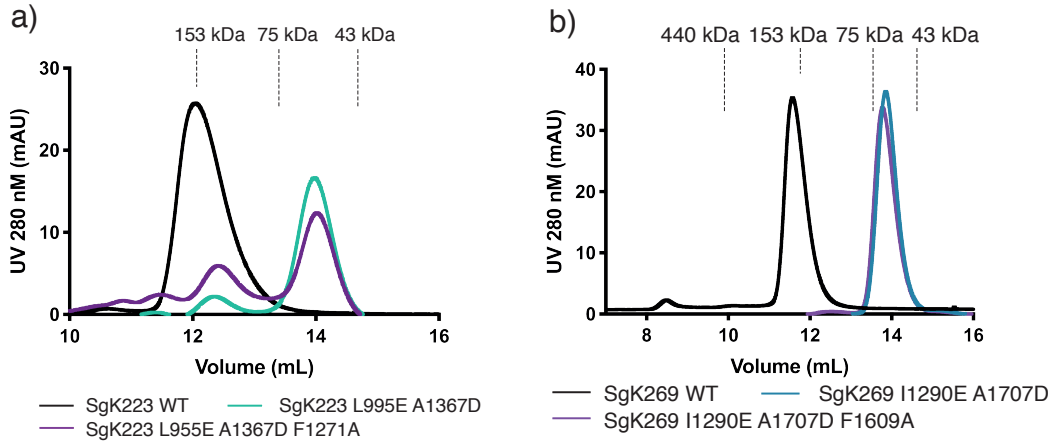


Figure 3.4.8. Analytical size exclusion chromatography of SgK223 and SgK269 dimerisation and oligomerisation mutants.
a) A-SEC of SgK223 double mutant (L955E, A1367D) (green) and SgK223 triple mutant (L955E, A1367D, F1271A) (purple). SgK223 WT in black as a comparison.
b) A-SEC of SgK269 double mutant (L1290E, A1707D) (green) and SgK223 triple mutant (L1290E, A1707D, F1706A) (purple). SgK223 WT in black as a comparison.

3.5 Discussion

In this Chapter, we found that the first 200 amino acids of SgK223 and SgK269, forming the N-terminal domains, were likely to be unstructured. The N-terminal domains eluted from A-SEC at an elution volume corresponding to approximately double their predicted MW, however we demonstrated through AUC that these domains are monomeric and most likely elongated in their structure, rather than globular.

It is possible that these proteins require a post-translational modification (PTM) or interaction with another protein to induce secondary structure of these domains. Based on the online tools Scansite and Phosphosite plus - that use the literature to predict interaction sites/PTM sites in proteins - the SgK269 N-terminal domain has no predicted PTMs or interaction sites, whereas, SgK223 N-terminal domain has a predicted SH3 domain binding site and two tyrosine residues (Tyr132 and Tyr159) that are predicted to undergo phosphorylation.

To investigate if phosphorylation of these residues induces secondary structure of SgK223 N-terminal domain, we could carry out expression of these proteins in insect cells. Insect cell expression can recapitulate the PTM's that occur upon expression of proteins in human cells, possibly leading to the folding of these domains. Using CD to measure secondary structure, we could investigate if phosphorylation leads to conformational changes of SgK223. Additionally, using techniques developed in Chapter 4, we could characterise the SH3 domain containing proteins that interact with the SgK223 N-terminal domain proline-rich region. The SH3 domain containing protein could then be co-expressed with SgK223 N-terminal domain to investigate if the interaction of SgK223 with an interaction partner induces secondary structure and conformational changes.

Another technique we could use to understand the overall shape of the N-terminal domains is small angle X-ray scattering (SAXS). This technique uses X-rays directed at a protein in solution to determine the average particle shape and size,

revealing an envelope of the protein [85]. This technique could provide information on the shape of the N-terminal domains, to visualise if they do have an elongated shape, as AUC suggested.

It is difficult to discern the function of these domains as we do not know their structure. However, to investigate their interaction partners, we could express tagged N-terminal domain constructs in human cell lines and carry out immunoprecipitation experiments from the cell lysate, coupled with mass spectrometry. This would demonstrate if there is an enrichment of proteins that interact with either SgK223 or SgK269 N-terminal domains. If we can characterise potential interactors of the N-terminal domains, this could provide information on their function.

In the second part of this Chapter we focused on the homo-and hetero-associations of SgK223 and SgK269. Previous research found that the αG and A-loop of SgK223 and SgK269 are responsible for driving homo- and hetero-oligomerisation [6]. We found that the N-lobe of the SgK223 pseudokinase domain forms the SgK223 complementary interacting oligomerisation interface (Fig. 3.5.1a), however the N-lobe of SgK269 pseudokinase domain does not contribute to the SgK269 complementary interacting oligomerisation interface. Thus, as illustrated in Figure 3.5.1b, the SgK269 homo- oligomerisation and hetero-oligomerisation interface involves the αG and A-loop and another unknown region of the pseudokinase domain or the flanking helices. The αG and A-loop of SgK269 may form the interface alone, interacting with the αG and A-loop of another SgK269 molecule (Fig. 3.5.1b).

Numerous signalling molecules can homo- and hetero-associate in the cell, such as the HER family of tyrosine kinase receptors or cytokine receptors, and this has been demonstrated to lead to different signalling outputs based on the subunits that interact [86, 87]. It is evident that SgK223 and SgK269 undergo homo- and hetero-dimerisation, a factor that could contribute to the diversification of signalling output [6]. Additionally, SgK223 and SgK269 can homo- and hetero-

oligomerise, and this occurs through different interacting interfaces in SgK223 compared to SgK269. Their oligomerisation is another factor that can contribute to the diversification of the signalling output by SgK223 and SgK269. Based on their dimeric and oligomeric arrangement, the different interactors of SgK223 and SgK269 will be localised together, leading to signalling outputs.

As we generated a suite of mutations capable of disrupting homo- and hetero-association of SgK223 and SgK269 in cells, we aim to use these mutations to understand the importance of SgK223 and SgK269 associations for signalling. As demonstrated in Patel *et al.*, disrupting dimerisation and oligomerisation of SgK223 lead to a decrease in cell migration, compared to overexpression of SgK223 WT [6]. Similarly, we could investigate the impact the N-lobe mutants of SgK223 have on cell migration, to verify if disrupting the N-lobe oligomeric interface has the same impact as the mutations at the A-loop and αG.

Interestingly, the alanine scanning mutations to the dimerisation domain that disrupted dimerisation of SgK223, failed to disrupt dimerisation in SgK269 [6]. Considering the helices that make up the dimerisation interface of SgK223 and SgK269 are highly conserved in sequence and structure, this result was not expected. The reason these mutants do not disrupt dimerisation in SgK269 could be attributed to other structural features of SgK269 that are not present in SgK223. SgK269 has a ~50 residue insertion in the N-lobe of the pseudokinase domain, which was absent from the SgK269 crystal structure. Additionally, there are regions where these proteins differ in sequence, which could play a role in stabilising the dimerisation interface of SgK269. Thus, although the sequences and structures of SgK223 and SgK269 are very similar, the differences that do occur between the two proteins may contribute to differences in the stability of their dimerisation interfaces.

The mutants of SgK223 and SgK269 that incorporate charged residues into their dimerisation interfaces both, as expected, disrupted dimerisation. Thus, these SgK269 dimerisation mutants could be expressed in cells to investigate the role

dimerisation plays in cell migration and morphology, as carried out for SgK223 [6]. Yet to be investigated is the signalling output upon disrupting dimerisation and oligomerisation of SgK223 and SgK269. Thus, these mutants could be expressed in cells and we could carry out immunoprecipitation coupled to mass spectrometry to investigate their binding partners when they cannot dimerise and oligomerise compared to WT protein.

Additionally, in Chapter 5 the mutants of SgK223 and SgK269 that disrupt dimerisation and oligomerisation are used to investigate the importance of SgK223 and SgK269 associations for their localisation. The localisation of SgK223 and SgK269 is critical for their role in signalling as they are scaffold proteins. Thus, understanding the role their associations play in their localisation could shed light on the importance of their associations in promoting signalling.

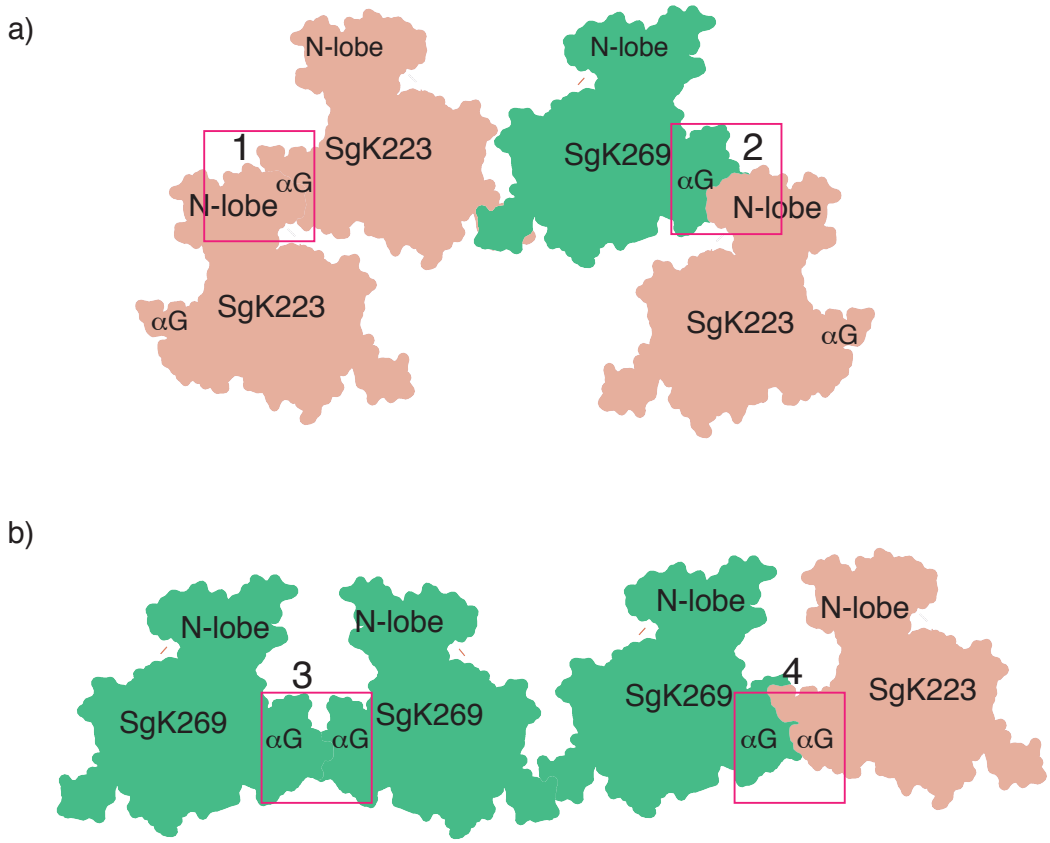


Figure 3.5.1. SgK223 and SgK269 homo- and hetero-oligomerisation schematic.
a) Homo- and hetero-oligomerisation interfaces of SgK223. SgK223 (pink) in a heterodimer with SgK269 (green). SgK223 can form 1. A homo-oligomerisation interface through the interacting interfaces of the αG and A-loop and the N-lobe of the pseudokinase domain. 2. A hetero-oligomerisation interface between the αG and A-loop of SgK269 and the N-lobe of SgK223.
b) Homo- and hetero-oligomerisation interfaces of SgK269. SgK269 (green) in a homodimer with SgK269 (green). SgK269 can form 3. A homo-oligomerisation interface between the αG and A-loop of two SgK269 pseudokinase molecules. 4. A hetero-oligomerisation interface between the αG and A-loop of SgK269 and the αG and A-loop of SgK223.

4.0 Biophysical and structural characterisation of the interactions between CrkII and the PEAK family of pseudokinases

4.1 Introduction

This Chapter focuses on the structural and biophysical characterisation of the interactions between the PEAK family members (SgK223 and SgK269, as well as the least studied member of this family, PEAK3) and the protein, CT-10 regulator of kinase II (CrkII). CrkII is an important signalling adaptor protein that localises to focal adhesions and in response to cell stimuli, such as growth factor signalling and cell adhesion, can promote cell migration signalling and changes to cell morphology [88-91]. The *c-CRK* gene encodes two splice variants, CrkI and CrkII, which differ in the domains they contain. CrkII consists of an N-terminal SH2 domain, followed by an N-terminal SH3 domain and a C-terminal SH3 domain (Fig. 4.1.1a). CrkI lacks the C-terminal SH3 domain [90]. CrkII is upregulated in a number of cancers. Specifically, studies of primary lung adenocarcinoma tumours demonstrated that there was a significantly higher expression of *CRK* mRNA (from CrkI and CrkII splice variants) in advanced stage and more invasive tumours, compared to early stage and less invasive tumours [89]. The mRNA levels of *CRK* correlated to the levels of protein present in tumours in the later and more invasive stages of lung adenocarcinoma [89]. Additionally, levels of CrkII protein were shown to be elevated in primary breast tumours [90]. Knockdown of *CRK* using siRNA decreased cell migration significantly in a number of human cancer cell lines including the breast cancer cell lines MDA-MB-231/435 [90]. As previously mentioned in Chapter 1, overexpression of SgK223 and SgK269 was also demonstrated to increase cell migration, and knockout of SgK223 in MCF-10A cells returns cell migration to rates comparable with the control cells expressing an empty vector [1, 6]. Previous studies have demonstrated that SgK269 can immunoprecipitate with CrkII, and this interaction was predicted to occur between a proline rich motif (PRM) in SgK269 and CrkII N-SH3 domain [3]. This suggests that the interaction of CrkII and the PEAK family could play a role in cell migration signalling.

Additionally, it was demonstrated recently that PEAK3 can immunoprecipitate CrkII and this interaction occurs through a PRM located ~70 residues upstream of the PEAK3 pseudokinase domain [52]. Mutating this PRM abolished the interaction of PEAK3 and CrkII, and functional data indicated that their interaction plays an important role in cell morphology signalling [52]. Specifically, PEAK3 interaction with CrkII antagonises the role of CrkII in inducing membrane ruffling [52]. As CrkII and CrkI are very similar, it is predicted that CrkI can also interact with the PEAK family proteins.

Co-immunoprecipitation experiments between CrkII and SgK269 may not demonstrate a direct interaction, and other than the PEAK3 - CrkII interaction, a direct interaction between CrkII and the PEAK family has not been demonstrated, nor the affinity of these interactions determined. As the PEAK family of proteins have many proline rich motifs and thus potential SH3 domain binding sites; we were aimed to better characterise these interactions *in vitro.*

4.1.1 Domain organisation of CrkII

Most adaptor proteins, like CrkII, have a number of protein interaction domains, such as SH2, SH3 or Pleckstrin homology (PH) domains [92]. This enables them to simultaneously interact with different protein domains or motifs and in doing so, localise multiple proteins to a specific site within a cell for regulation of signalling. As mentioned, CrkII has an N-terminal SH2 domain, followed by two SH3 domains referred to as the N-terminal SH3 domain and the C-terminal SH3 domain (Fig.4.1.1a) [93].

The Src homology 2 (SH2) domain was originally identified as a 100-amino acid region found in proteins related to c-Src, and was known to be an important module for interaction with receptor tyrosine kinase cytoplasmic domains [94]. SH2 domains were shown to specifically interact with phosphorylated tyrosine residues, dictated by the sequences that surround the tyrosine [94]. The interactions of the CrkII SH2 are specific and the consensus binding motif of the

domain is pYxxP, whereby the ‘pY’ represents the phosphorylated tyrosine residue [95].

The Src homology 3 (SH3) domain was first identified as a ~60-amino acid region based on sequence similarity of a number of proteins such as Crk and Src tyrosine kinase. Subsequently, it was demonstrated that its function in cells was to interact with proline rich motifs (PRMs) of interacting partners [96]. Typically, these PRMs adopt a polyproline II helix (PPII) on binding to the SH3 domain [97, 98]. SH3 domains have been classified into nine sub-classes based on the consensus sequence of the PRMs with which they are able to interact (Table 4.1.1) [97]. The first two classes are defined as the canonical SH3 domains. For example, as shown in Figure 4.1.2, class I SH3 domains interact with PRMs that have the consensus sequence RxxPxxP, whereby x is any amino acid, annotated with a plus (+) indicating the peptide binding orientation (Fig. 4.1.2a). Class II interacts with the consensus sequence PxxPxR, in the minus (-) peptide binding orientation (Fig. 4.1.2b). In addition to the ‘canonical’ binding modes, a number of PRM/SH3 interactions have been shown to adopt other non-canonical binding modes [98]. Classes III-VIII are the non-canonical PRM binding SH3 domains and class IX is termed the atypical class as these SH3 domains can interact with either non-canonical PRMs or multiple different classes of PRM sequences (Table 4.1.1) [97].

Of the two SH3 domains of CrkII, the N-SH3 domain is the best characterised. It belongs to the class II SH3 domains and is known to interact with a number of focal adhesion associating proteins such as Abl kinase and DOCK180 [95, 97]. The consensus PRM sequence for the CrkII N-SH3 domain is Px(L/Y)PxK, which differs slightly from the class II SH3 domain consensus motif of PxxPxR, whereby the CrkII binding motif has a lysine in place of the C-terminal arginine. In contrast, less is known about the function of the C-SH3 domain, however it is suggested to play a role in the negative regulation of CrkII [99]. The PRM binding region of CrkII N-SH3 is quite hydrophobic. Comparatively, the predicted proline-rich motif

binding region of CrkII C-SH3 domain is hydrophilic with no known interactors, suggesting it may not bind PRMs [100-103].

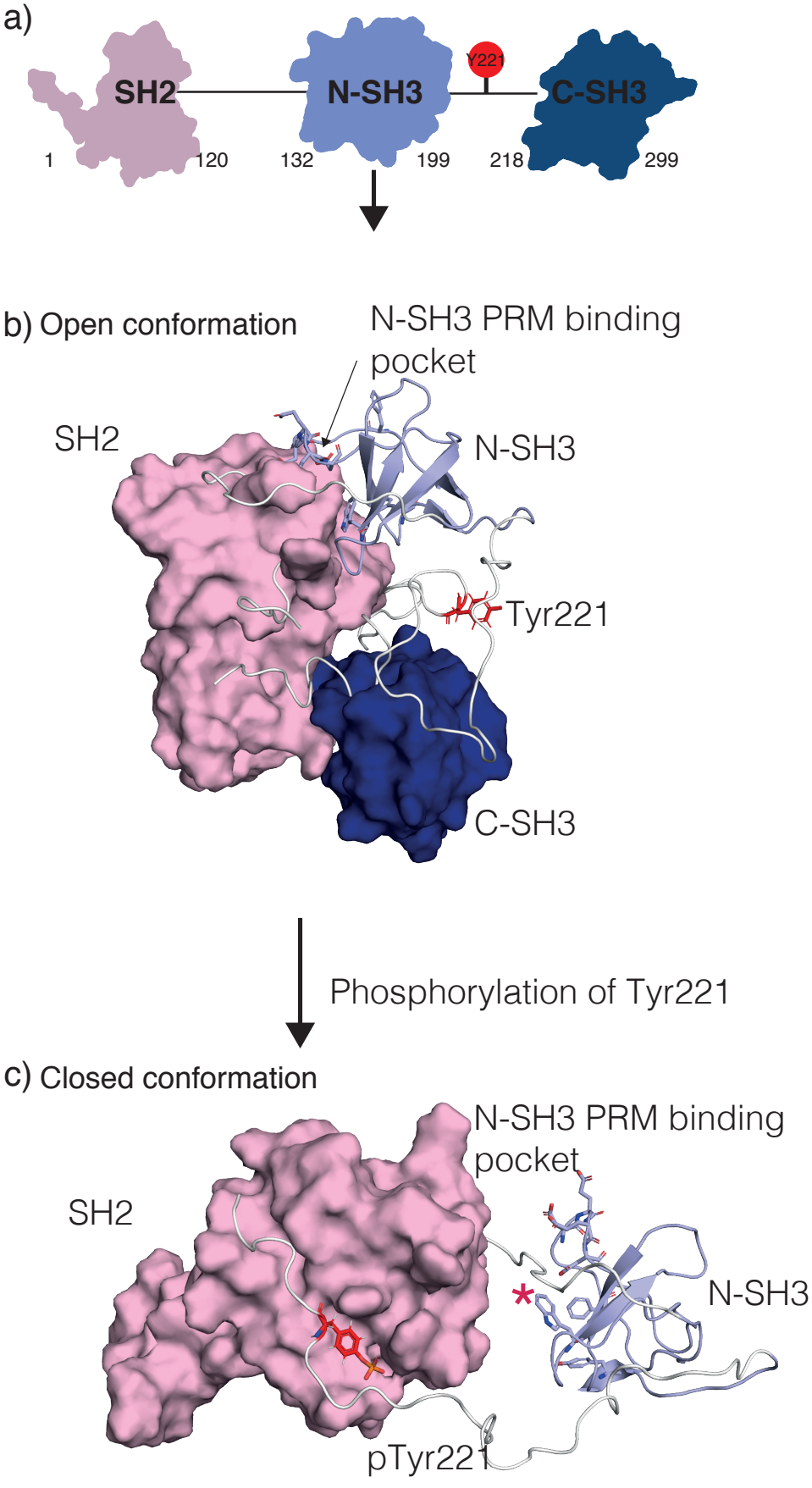

a)
SH2
N-SH3
C-SH3
Y221
1
120
132
199
218
299
b) Open conformation
N-SH3 PRM binding pocket
SH2
N-SH3
Tyr221
C-SH3
Phosphorylation of Tyr221
c) Closed conformation
N-SH3 PRM binding pocket
SH2
N-SH3
pTyr221

Figure 4.1.1. Regulation of CrkII N-SH3 interactions.
a) Domain layout of CrkII, tyrosine 221 (in red) between N- and C-SH3 domains. N-SH3, light blue; C-SH3, dark blue; SH2, pink.
b) Structure of CrkII in the active conformation with tyrosine 221 surface exposed (PDB: 2EYZ). Colours as above.
c) Structure of CrkII SH2 and N-SH3 domains in the inactive conformation, with phosphorylated tyrosine 221 (PDB: 2DVJ). The loop between the SH2 domain and the SH3 domain blocks N-SH3- PRM interactions, highlighted by a red asterisk.

Table 4.1.1. SH3 domains classed according to the consensus PRM sequence they recognise.
[97]

Type of Motif	Class of Motif	Consensus Sequence	Example of protein with the PRM
Canonical	I	$^{-3}RxxP^{0}xxP^{+3}(+)$	YES1-1/1, NCK21/3-SV2/2, SRC1/2
	II	$^{0}PxxP^{+3}xR^{+5}(-)$	NCK1-2/3, CrK-1/2, CD2AP1/1
	I/II	-	FYN-1/1, GRB2-1/2
Non-Canonical	III	$^{-6}RxxxxxP^{0}(+)$	STAC2-1/1, BTK-1/1
	IV	$^{0}PxxxxR^{+5}(-)$	MAP3K9-1/2, CD2AP-1/3
	V	$^{-3}RxxP^{0}xxx^{+3}(+)$	TXK-1/1, ASAP2-1/1
	VI	$^{-3}xxxP^{0}xR^{+2}(-)$	ASAP1-1/1, ASAP2-1/1
	VII	$^{0}PxK^{+2}P^{+3}(-)$	BCAR1-1/1, NEDD9-1/1
	VIII	$^{0}PxxP^{+3}xxP^{+6}(-)$	SORBS2-1/3, SORBS2-2/3,
	IX	Atypical	GRB2-2/2, SRC-1/1

a) Class I: $^{-3}RxP^{0}xxP^{+3}(+)$
PBD: 1QWF

b) Class II: $P^{0}xxP^{+3}xR^{+5}(-)$
PBD: 3U23

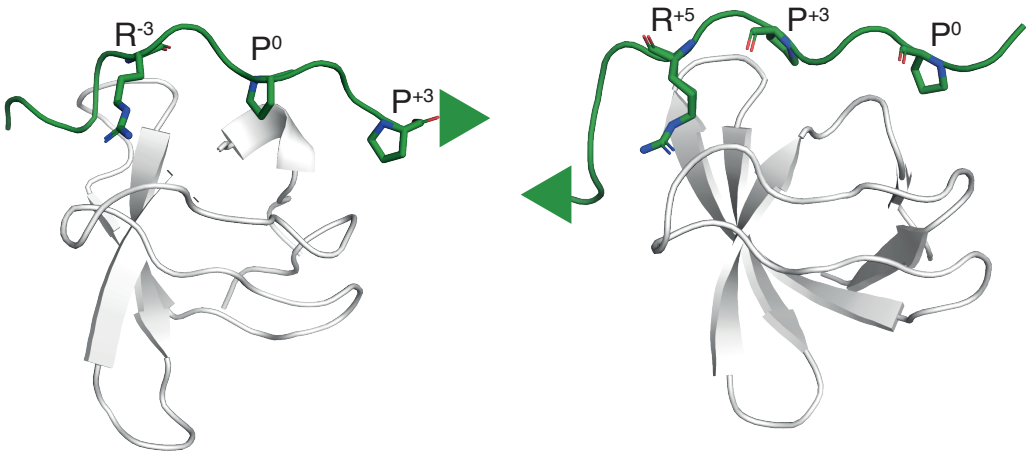


Figure 4.1.2. Class I SH3 domain compared to Class II SH3 domain.
a) Representative structure of a class I SH3 domain (Src SH3 domain in complex with a PRM, PDB: 1QWF) bound to a PPII proline-rich peptide. The peptide is shown in green and the arrowhead indicates the C-terminus. The consensus sequence residues are shown as sticks and labelled.
b) Representative structure of a class II SH3 domain (CD2AP SH3 domain in complex with a PRM, PDB: 3U23) bound to a PPII proline-rich peptide. The peptide is shown in green and the arrowhead indicates the C-terminus. The consensus sequence residues are shown as sticks and labelled.

4.1.2 CrkII interactions

CrkII has been reported to influence cell migration by interacting with the focal adhesion proteins paxillin and p130Cas. Upon integrin stimulation, paxillin and p130Cas are tyrosine phosphorylated, leading to their interactions with CrkII through its SH2 domain, promoting cell migration [93, 104]. It has been reported that upon mutation of paxillin's two main phosphorylation sites (Tyr31 and Tyr118) from a tyrosine to a phenylalanine, the rate of migration of Nara bladder tumour II cells was decreased, as paxillin could no longer interact with CrkII [104].

Similarly, CrkII interacts with DOCK180, through its N-SH3 domain, to alter cell morphology [105]. Upon integrin stimulation in NIH 3T3 fibroblast cells and expression of p130Cas and CrkII, DOCK180 localises to focal adhesions and the cell membrane [106]. It is suggested that p130Cas, CrkII and DOCK180 form a complex that is necessary for the localisation of DOCK180 to the membrane to induce changes to cell morphology [90, 93, 107].

4.1.3 Regulation of CrkII interactions

Unlike interactions of SH2 domains, in which binding to interacting motifs occurs following phosphorylation of the motifs, SH3 domains can constitutively interact with their target proteins to promote signalling. Thus, SH3/PRM interactions are often heavily regulated by other mechanisms, such as alterations in protein conformation. CrkII can adopt variable conformations to regulate its interaction potential. For example, Abl kinase, a focal adhesion associating protein, binds to the CrkII N-SH3 domain via a PRM. Upon interaction with the N-SH3 domain, Abl kinase phosphorylates tyrosine 221 of CrkII, located between the N-SH3 and C-SH3 domains (Fig.4.1.1a) [103]. This event induces a conformational change in CrkII, such that phosphorylated Tyr221 becomes a binding site for the CrkII SH2 domain, leading to a closed conformation of CrkII. Because of this conformational change, the N-SH3 domain is no longer accessible to PRM containing substrates, such as Abl and DOCK180, and is unable to carry out subsequent signalling [108, 109] (Fig. 4.1.1b,c). Dephosphorylation of CrkII Tyr221 by the phosphatase

PTP1B releases the closed conformation, promoting both SH2 and N-SH3 binding activity of CrkII (Fig. 4.1.1b) [93]. Interaction of a cell with the extracellular matrix (ECM) fibronectin leads to signalling that induces the dephosphorylation of CrkII. Subsequently, CrkII interacts with its substrate paxillin, increasing cell migration [93]. CrkII can therefore switch between a closed and open state depending on the local environment and extracellular signalling.

4.1.4 Proline rich motifs in the PEAK family

SgK223, SgK269 and PEAK3 have proline rich motifs (PRMs) within their PEST linker. The PRMs of SgK223 (residues 809-821), SgK269 (residues 1150-1162) and PEAK3 (residues 54-66) follow the consensus sequence (Px(L/Y)PxK) predicted to bind the CrkII N-SH3 domain (Fig. 4.1.3b). In addition, SgK223 and SgK269 each have another PRM in their PEST linker (SgK223 residues 709-721 and SgK269 residues 1117-1129), however these PRMs do not stringently follow the CrkII N-SH3 consensus motif, or follow the consensus motifs of any of the other classes of SH3 domains (Fig. 4.1.3b).

We aim to demonstrate a direct interaction between the PEAK family of proteins and CrkII N-SH3 domain and to determine the affinities of their interaction. Additionally, we aim to provide structural insights into CrkII recognition by the PEAK family.

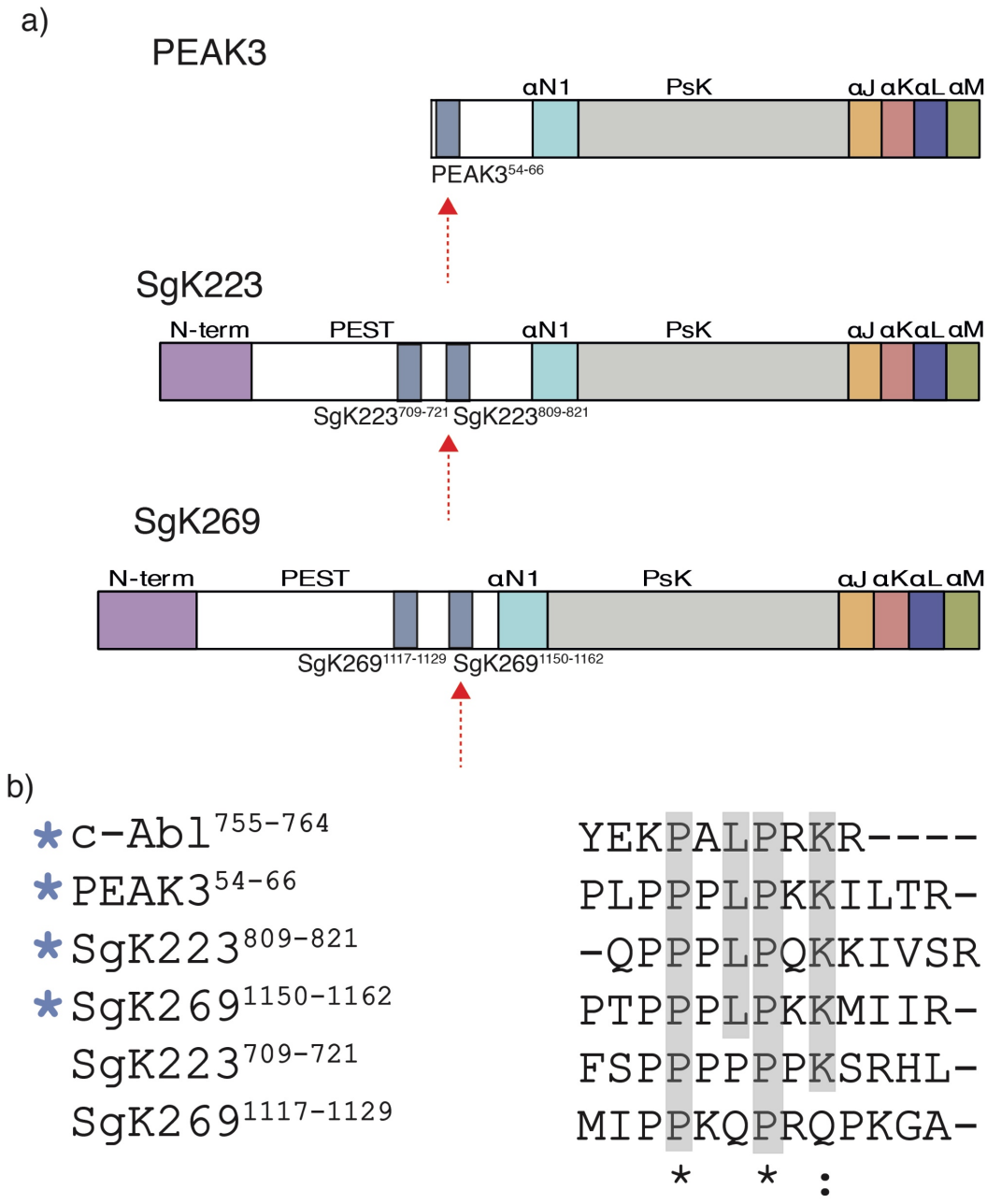


Figure 4.1.3. Diagram of PEAK family proteins and their PRMs selected for study.

a) Domain layout of PEAK3, SgK223 and SgK269, showing the PRMs in the PEST region selected for study. PRM residue numbers are labelled below each site. PRMs with the consensus sequence for CrkII N-SH3 domain indicated with a red arrow below the domain layout.

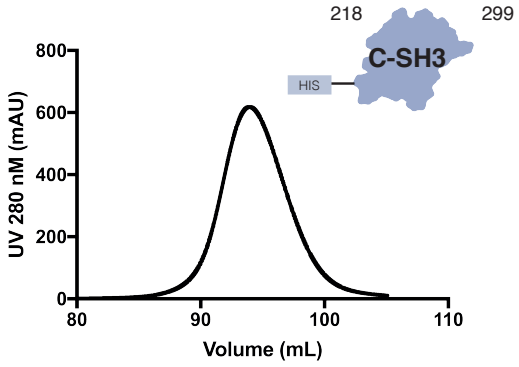
b) Sequence alignment of the PEAK family PRM peptides selected for study. C-Abl is a known interactor of CrkII and is used as a reference sequence. Blue asterisks indicate the PRMs containing the recognised consensus sequence for binding to the CrkII N-SH3 domain; consensus sequence residues are highlighted in grey.

4.2 Results

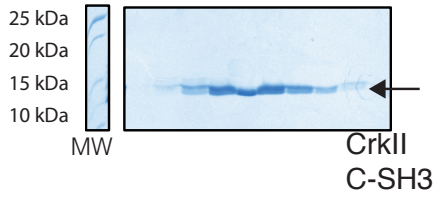
4.2.1 Expression and purification of CrkII SH3 domains

To characterise the interaction between CrkII and the PEAK family of proteins, we cloned the N-SH3 and C-SH3 domains into pCOLD IV *E.coli* expression vectors as N-terminal His6 fusion proteins with a TEV protease cleavage site C-terminal to the His6 tag (see Chapter 2, Figure 2.1 and Table 8.1 in the Appendices). CrkII N-SH3 and C-SH3 domains were purified using nickel affinity chromatography followed by size exclusion chromatography (SEC) (Fig. 4.2.1a, b) [110]. Each of the N-SH3 and C-SH3 domains eluted as a single peak with a retention volume of approximately 92 mL on a Superdex HiLoad 75 16/600 column (GE Healthcare), suggesting they eluted as monomers with the expected MW of ~9-11 kDa. SDS-PAGE confirmed protein purity to ~100%. Purified His-tagged proteins were concentrated to 100 μM for Surface Plasmon Resonance (SPR) experiments (see section 4.2.2). For crystallisation of CrkII N-SH3 domain, the His6 tag was first cleaved using TEV protease, then the protein was further purified using anion exchange chromatography (MonoQ, GE Healthcare). Final protein purity was assessed by SDS-PAGE (Fig. 4.2.1c). Fractions containing protein were concentrated to 5-11.5 mM for crystallography (see section 4.2.3). Proteins were all flash frozen and stored at -80 °C or used directly after purification.

a)

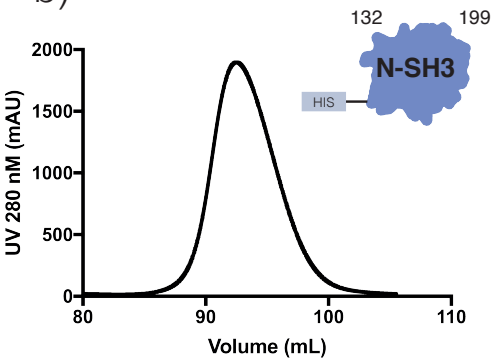


CrkII C-SH3

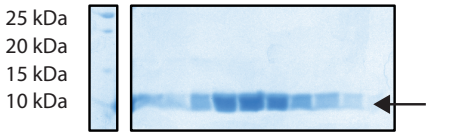


MW: 11220 kDa

b)

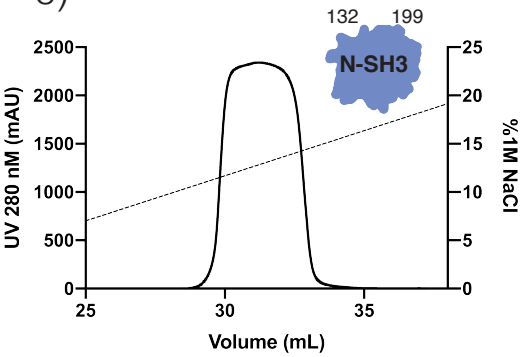


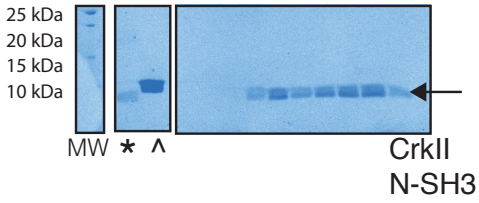
CrkII N-SH3



MW: 9129 kDa

c)





MW: 6962 kDa

CrkII N-SH3

%1M NaCl

Figure 4.2.1. Purification of CrkII SH3 domains.
a) Size exclusion chromatogram of CrkII N-SH3 domain and SDS-PAGE of fractions from SEC elution.
b) Size exclusion chromatogram of CrkII C-SH3 domain and SDS-PAGE of fractions from SEC elution.
c) Anion- exchange chromatogram of CrkII N-SH3 domain, salt concentration indicated by the dashed line. SDS-PAGE of fractions from Anion exchange elution profiles. *Indicates protein after TEV protease cleavage and ^indicates protein prior to TEV protease cleavage.
MW= Molecular Weight marker (BioRad Precision Plus unstained protein standards).

4.2.2 SPR analysis of binding interactions between CrkII and PRMs from the PEAK family proteins

To better characterise the interactions between the CrkII N-SH3 domain and the PEAK family proteins, we undertook binding studies using Surface Plasmon Resonance (SPR). This is a biophysical technique that can sensitively measure binding interactions between an analyte in solution and an immobilised ligand.

The principle of SPR involves light (p-polarised light) that is directed at a chip coated with a thin gold film. Gold is chosen as the metallic chip coating as it has conduction band electrons, the requirement for generating surface plasmon resonance, and is chemically inert [111]. Once the light hits the backside of the chip, the light is reflected off it at a specific incidence angle, known as the resonance angle. Light is also absorbed by the gold layer on the chip, causing electrons in the metal to resonate, creating surface plasmons [112]. This results in an intensity loss of the reflected beam and can be seen as a dip in SPR reflection intensity, which is used to convey information about the sensor chip surface. Once ligand is bound to the chips surface and an analyte is flowed over, changes in analyte binding to the ligand on the chip can be measured as changes in reflection angle or wavelength, over time [112, 113]. The changes in reflection angle or wavelength are converted to an arbitrary measurement of resonance units (RU) by the Biacore machine (GE Healthcare) and can then be used to

calculate the affinity and kinetics of analyte binding to the ligand. This is plotted as SPR response over time (Fig. 4.2.2) [113].

The PRMs selected for this study are listed in Figure 4.1.3 and Table 4.2.2. Two sets of peptides were designed corresponding to these sequences. Each peptide was designed as 1) a biotinylated peptide through the addition of a long-chain biotin tag (Lcbiotin) to the N-terminus, to allow efficient interaction with CrkII when bound to the SPR chip, and 2) as an acetylated peptide to be used as the analyte for SPR studies and crystallographic studies.

We first immobilised the N-SH3 domain of CrkII to a CM5 chip by random amine coupling. We injected 5 μg/mL of protein onto the chip over two separate flow cells, until we reached the desired response units (RU) (1380 RU and 443.5 RU) (see Chapter 2 for experimental details) (Fig.4.2.2a). CrKII C-SH3 was used as a control and immobilised using the same approach, injected at a concentration of 10 μg/mL until it reached 649.9 RU on the chip. Then, acetylate peptides over a concentration range of 100 μM-0.098 μM for the consensus peptides (PEAK3$^{54-66}$, SgK223$^{809-821}$ and SgK269$^{1150-1162}$) and 100 μM- 1.56 μM for SgK223$^{709-721}$ and SgK269$^{1117-1129}$ peptides, were flowed over the CrkII immobilised SH3 domains.

The SgK223$^{809-821}$ peptide did not dissociate completely from the chip during the experiments, most likely due to non-specific interactions, thus we included a regeneration step by injecting the SgK269$^{1150-1162}$ peptide to out-compete the interaction between SgK223$^{809-821}$ and CrkII N-SH3 domain. Complete dissociation of the analyte from the ligand between cycles in a multicycle experiment and return of the RU to baseline is important to correctly calculate the affinity and kinetics of interaction. The regeneration step lead to the complete dissociation of SgK223$^{809-821}$ and allowed us to accurately calculate the affinity of interaction between CrkII N-SH3 domain and the SgK223$^{809-821}$ peptide (Fig. 4.2.3b).

Data were analysed using the Biacore S200 software version 1.1 (GE Healthcare) applying a 1:1 binding model to calculate the K_D (Fig. 4.2.3). Qualitatively, the on and off rates of the peptides interacting with CrkII N-SH3 domains were very fast, precluding accurate kinetic fitting. Thus, we used a steady-state model to calculate the affinities of interaction. As shown in Table 4.2.2, all the affinities (K_D) between the consensus PEAK family peptides (PEAK3$^{54-66}$, SgK223$^{809-821}$ and SgK269$^{1150-1162}$) and CrkII N-SH3 domain are in the range 1-2 μM (Fig. 4.2.3). Both of the PRMs of SgK223 tested were found to interact with CrkII N-SH3; SgK223$^{809-821}$ showed a higher affinity with a K_D of 2.4 μM, as compared to 13 μM for SgK223$^{709-721}$. SgK269$^{1117-1129}$ did not interact with CrkII N-SH3 domain.

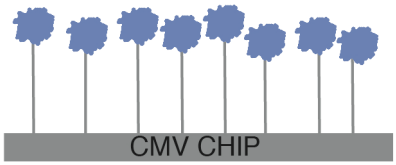
To further support the binding data, we then performed the reverse experiment, in which biotinylated PEAK family PRM peptides were immobilised on a streptavidin chip and CrkII SH3 domains were flowed over (Fig. 4.2.2b). Experiments were carried out as previously described and the CrkII N- and C-SH3 domains were titrated from 100 μM-0.098 μM. The affinities of interaction between the N-SH3 domain and the PEAK peptides were calculated at steady-state using a 1:1 binding model, as described previously. As for the earlier experiments in which CrkII was immobilised, the peptides, PEAK3$^{54-66}$, SgK223$^{809-821}$ and SgK269$^{1150-1162}$, were found to interact relatively tightly, in the range of 3-9 μM, with CrkII N-SH3 domain (Fig. 4.2.4 and Table 4.2.2). There was consistently an approximately a 3-fold weaker observed binding affinity between CrkII N-SH3 domain and the PEAK peptides in this reverse experiment as compared to when CrkII was immobilised to the chip (Table 4.2.2). Possible reasons for this difference may be the peptides being closer to the chip surface and thus leading to steric hindrance of the interaction with CrkII, or measurement differences in the determination of protein and peptide concentrations. Despite the small absolute difference (~3-fold) in the values for the experiment in the two formats, it is reassuring that the relative binding trends for the different peptides is maintained.

Together, these experiments indicated that two PRMs from SgK223 are capable of binding to CrkII N-SH3 domain, with the affinity of SgK223$^{809-821}$ to CrkII N-SH3 domain 5 –fold higher than the affinity of SgK223$^{709-721}$. This demonstrates that the SgK223$^{809-821}$ peptide of SgK223 is the most probable interaction site with the CrkII N-SH3 domain. As the SgK223$^{709-721}$ peptide can also interact with CrkII but with a weaker affinity (~13 μM), it may be important for a transient interaction at focal adhesions. Additionally, binding of the SgK223$^{709-721}$ peptide may only be permitted upon clustering of SgK223 with CrkII, increasing their local concentrations. This weaker interaction may be required for SgK223 signalling in a specific pathway at focal adhesions, however the details of this are yet to be elucidated. The SgK269$^{1150-1162}$ peptide interacted with the CrkII N-SH3 domain, however, SgK269$^{1117-1129}$ peptide showed considerably weaker interaction with the N-SH3 domain (70-fold weaker compared to SgK269$^{1150-1162}$). The SgK269$^{1117-1129}$ peptide had a K_D of >100 μM in the peptide immobilised experiment and no measurable interaction in the CrkII immobilised experiment (Table 4.2.2). The only PRM peptide of PEAK3, PEAK3$^{54-66}$ interacted with CrkII N-SH3 domain in both experiments, as expected.

Overall, these data support previous conclusions that PxLPxK (where x can be any amino acid) is the conserved motif for CrkII N-SH3 recognition of PRMs [110]. Thus, we can confirm that each member of the PEAK family has one PRM that binds with low micromolar affinity to CrkII N-SH3 domain (SgK223 residues 809-821, SgK269 residues 1150-1162 and PEAK3 residues 54-66).

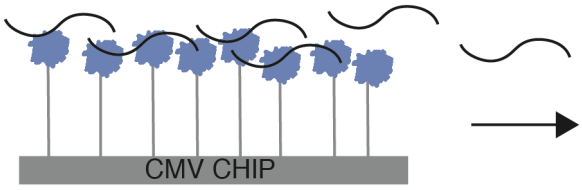
CrkII C-SH3 showed no measurable binding (>100 μM) to the PEAK family peptides.

a)

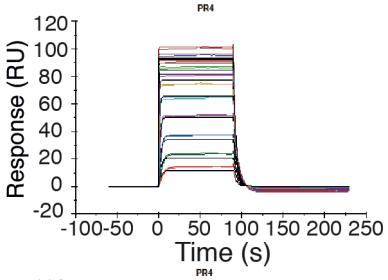


1. CrkII immobilised on the chip through random amine coupling

2. Peptides flowed over

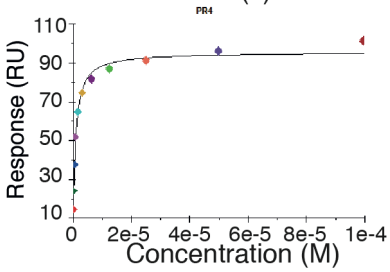


3. Peptides interact with CrkII SH3 domain

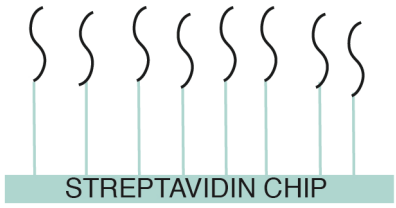


4. Sensorgram of peptide-CrkII SH3 domain interaction

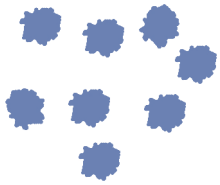
5. K_D of interaction calculated



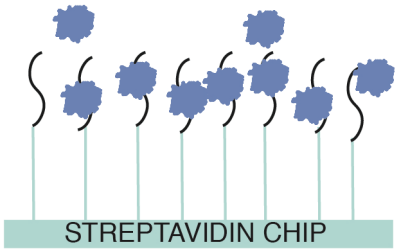
b)



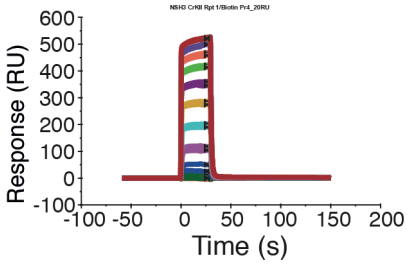
1. Biotinylated PRMs immobilised on a streptavidin chip



2. CrkII flowed over



3. Peptides interact with CrkII SH3 domain



4. Sensorgram of peptide-CrkII SH3 domain interaction

5. K_D of interaction calculated

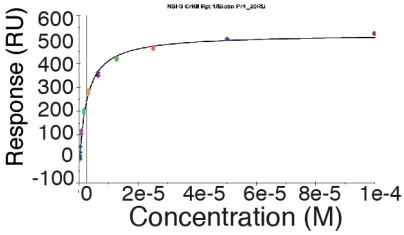


Figure 4.2.2. SPR experimental design.
a) Experimental design of CrkII immobilisation. 1. CrkII is immobilised on the CMV chip using random amine coupling. 2&3. The PEAK family peptides are flowed over, interacting with the protein immobilised on the chip. 4&5. Light is directed at the backside of the chip and the change in reflection intensity is measured and converted to response units. Depending on the analyte-ligand interactions, the reflection intensity of the light will change, plotted as a sensorgram showing time versus response, over the different concentrations of analyte flowed over the chips surface. Subsequently, the response values of each concentration at steady-state is calculated and the concentration of analyte versus response is plotted. These data points are fitted and the affinity values are calculated (K_D).
b) Experimental design of peptide immobilisation. 1. Biotinylated peptides are immobilised to the streptavidin chip. 2&3. CrkII SH3 domains are flowed over, interacting with the peptides on the chip. 4&5. A sensorgram is generated and affinity values calculated from the data generated as described in a).

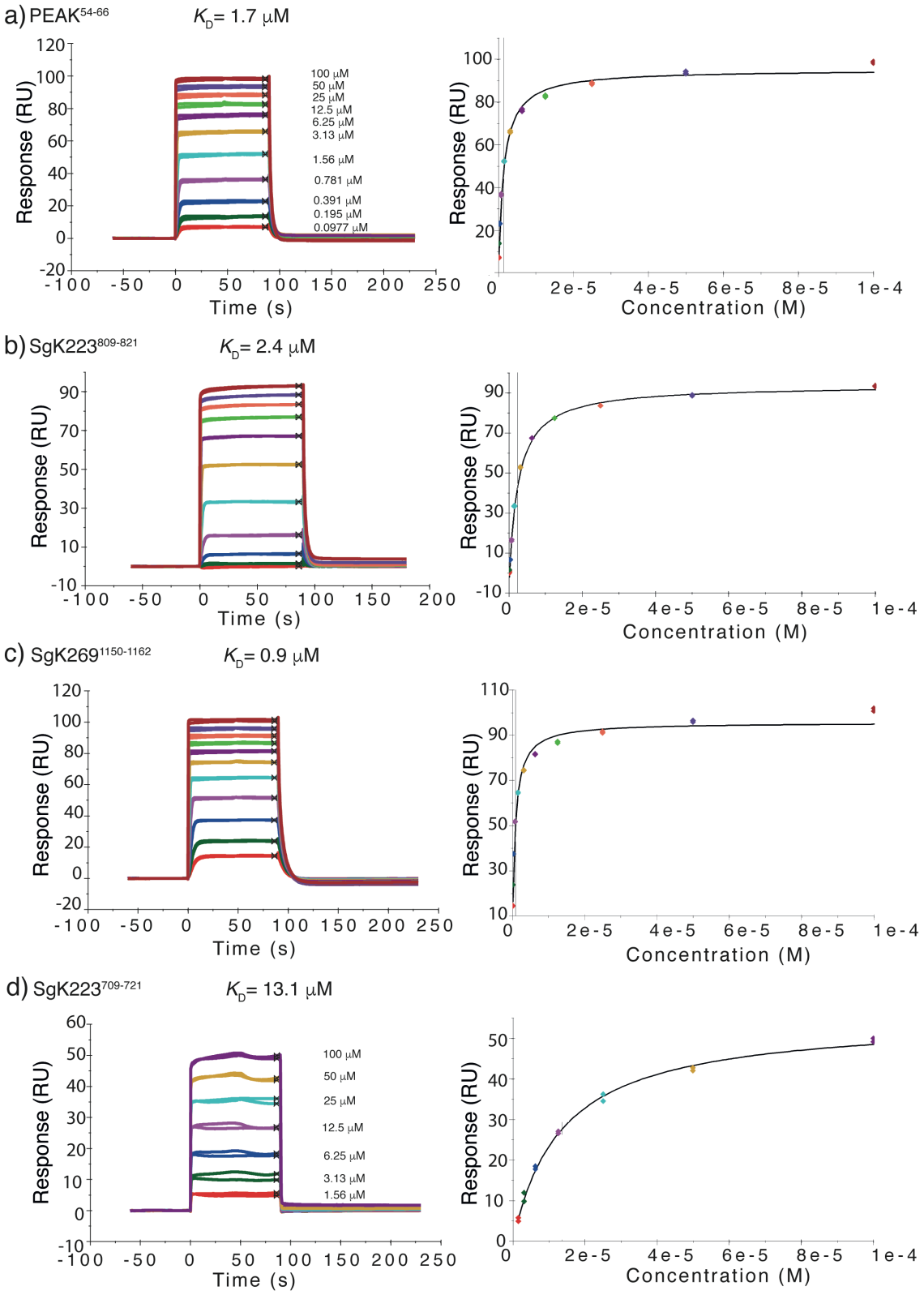

a) PEAK54-66
KD= 1.7 μM
Response (RU)
Time (s)
100 μM
50 μM
25 μM
12.5 μM
6.25 μM
3.13 μM
1.56 μM
0.781 μM
0.391 μM
0.195 μM
0.0977 μM
Concentration (M)
b) SgK223809-821
KD= 2.4 μM
c) SgK2691150-1162
KD= 0.9 μM
d) SgK223709-721
KD= 13.1 μM

Figure 4.2.3. Sensorgrams and steady state model fitting for CrkII immobilised SPR experiments.
a) PEAK3$^{54-66}$ peptide sensorgram and steady state model fitting, concentrations of peptides PEAK3$^{54-66}$, SgK223$^{809-821}$ and SgK269$^{1150-1162}$ listed next to the PEAK3$^{54-66}$ sensorgram curves.
b) SgK223$^{809-821}$ peptide sensorgram and steady state model fitting.
c) SgK269$^{1150-1162}$ peptide sensorgram and steady state model fitting.
d) SgK223$^{709-721}$ peptide sensorgram and steady state model fitting, concentrations of peptide SgK223$^{709-721}$ listed next to the sensorgram curves.

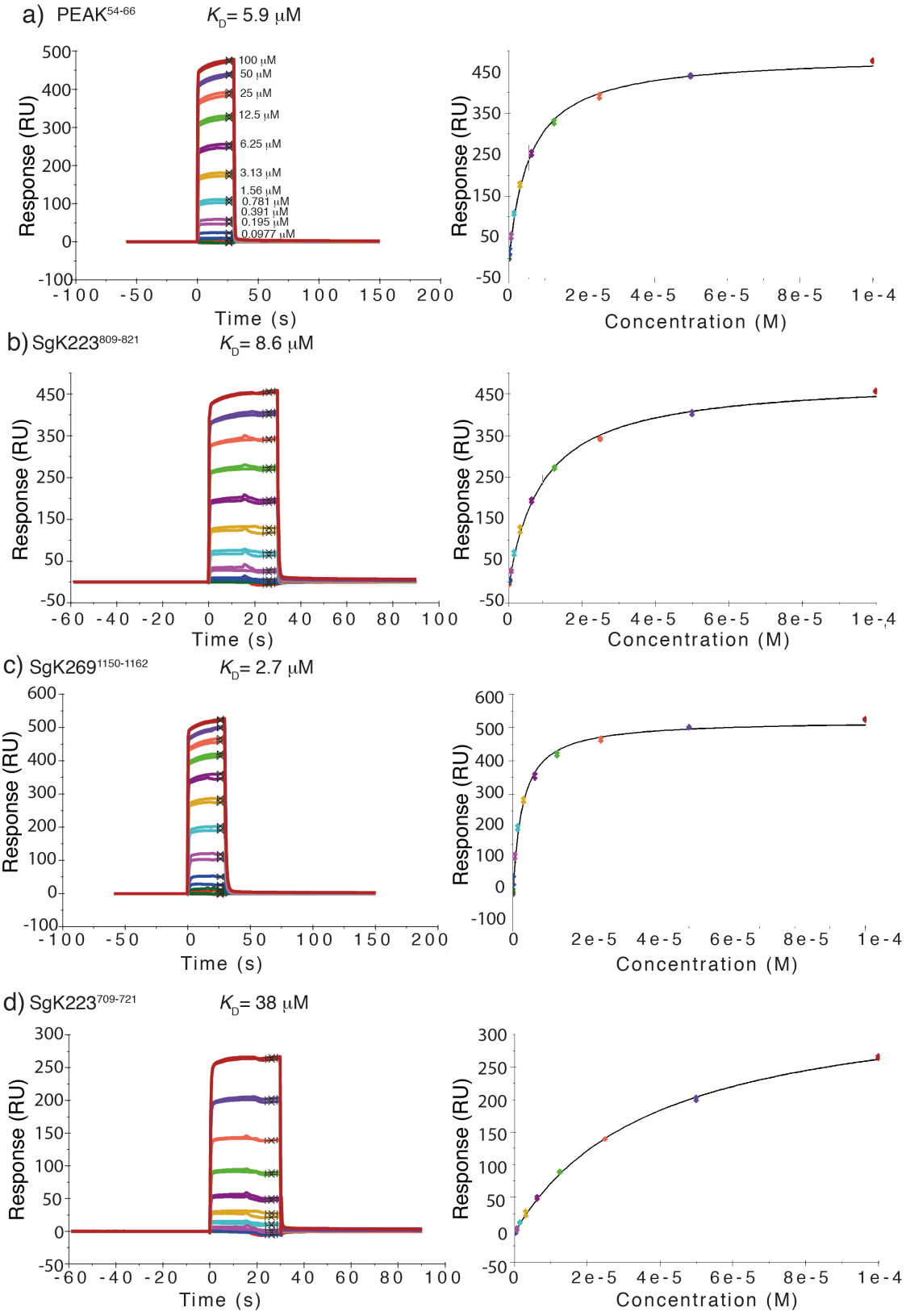

a) PEAK54-66 KD= 5.9 μM
100 μM
50 μM
25 μM
12.5 μM
6.25 μM
3.13 μM
1.56 μM
0.781 μM
0.391 μM
0.195 μM
0.0977 μM
Response (RU)
Time (s)
Concentration (M)
b) SgK223809-821 KD= 8.6 μM
c) SgK269 1150-1162 KD= 2.7 μM
d) SgK223709-721 KD= 38 μM

Figure 4.2.4. Sensorgrams and steady state model fitting for peptide immobilised SPR experiments.
a) PEAK3$^{54-66}$ peptide sensorgram and steady state model fitting, concentrations of peptides PEAK3$^{54-66}$, SgK223$^{809-821}$ and SgK269$^{1150-1162}$ listed next to the PEAK3$^{54-66}$ sensorgram curves.
b) SgK223$^{809-821}$ peptide sensorgram and steady state model fitting.
c) SgK269$^{1150-1162}$ peptide sensorgram and steady state model fitting.
d) SgK223$^{709-721}$ peptide sensorgram and steady state model fitting.

Table 4.2.2. Affinity values from the interactions of CrkII N-SH3 and the PEAK family PRMs, measured by SPR.

Ligand	Sequence	CrkII N-SH3 domain immobilised K_D +/- SEM	Biotinylated peptide immobilised K_D +/- SEM
PEAK$^{54-66}$	PLPPPLPKKILTR	1.7 +/- 0.36	5.9 +/- 0.15
SgK223$^{809-821}$	QPPPLPQKKIVSR	2.4 +/- 0.14	8.6 +/- 0.32
SgK269$^{1150-1162}$	PTPPPLPKKMIIR	0.9 +/- 0.050	2.7 +/- 0.12
SgK223$^{709-721}$	FSPPPPPPKSRHL	13 +/- 0.29	38 +/- 1.3
SgK269$^{1117-1129}$	MIPPKQPRQPKGA	-	$1.9 \cdot 10^2$ +/- 51

4.2.3 Structural investigations into PEAK family- CrkII N-SH3 interactions

To gain insight into the binding mode of the PEAK family peptides with CrkII N-SH3 domain, we next sought to crystallise the purified CrkII N-SH3 domain with each of the three peptides displaying the highest affinity for this domain by SPR (Table 4.2.2).

4.2.3.1 Crystallisation and data collection of CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$ peptide

The SH3/PRM complex was prepared by co-incubating the CrkII N-SH3 domain (5.7 mM, 40 mg/mL) with a slight molar excess of SgK269$^{1150-1162}$ peptide (6.5 mM) as described in Bhatt *et al.* 2016 [110]. Crystallisation trials were set up at CSIRO C3 facility (see Chapter 2 for the crystallisation screens used). Crystals grew in 45% PEG 2000 MME, 0.1 M Na Cacodylate pH 6.5 at 20 °C and had a long, thin needle morphology. Optimisation was carried out in house using the hanging drop method. The protein concentration was increased to 60 mg/mL to increase crystal size and thickness. The crystals used for data collection were crystallised in 43% PEG 2000 MME, 0.1 M Na Cacodylate pH 6.3 at 20 °C; crystals were harvested directly from the drop and flash frozen in N_2 (liq.) without the need for additional cryoprotection. Data were collected at the Australian Synchrotron by Onisha Patel and Michael Roy (MX2 beamline) and processed by Michael Roy.

4.2.3.2 Crystallisation and data collection of CrkII N-SH3 domain in apo form

Crystals of the apo form of CrkII N-SH3 were also obtained in a similar condition with a slight molar excess of SgK223$^{809-821}$ peptide, although due to the crystal packing, there was no peptide present in the X-ray crystal structure. The crystals were harvested, flash frozen and data collected as described above.

In addition, crystal screening was also performed using complexes of CrkII N-SH3 with the PEAK3$^{54-66}$ peptide and under the same conditions, however this has not resulted in any crystals at this stage.

4.2.3.3 Structure determination of CrkII N-SH3 domain apo

Diffraction data for the CrkII N-SH3 domain were processed in XDS and Buster to 1.5 Å [69, 70]. The structure was solved by molecular replacement (Phaser, Phenix Suite), using the CrkII N-SH3 domain (PDB: 5IH2) as a search model [110, 114]. The initial structure was improved by multiple rounds of building (Coot) and refinement in Phenix [74, 75]. The CrkII N-SH3 domain crystallised in the space group $P\,4_1\,2_1\,2$ with unit cell dimensions 33.1, 33.1, 43.8 Å and angles 90, 90, 90 °, with one molecule in the asymmetric unit. The data collection and refinement statistics are listed in Table 4.2.3.

The CrkII N-SH3 domain adopts the expected fold typical for a SH3 domain [115, 116]. The N-SH3 domain has two beta sheets (each consisting of three beta strands) that lie orthogonal to each other. One sheet consists of the strands b-a-e and the other b-c-d (Fig. 4.2.5). Additionally, there is a 3_{10} helix, RT-loop, n-Src loop and distal loop, with the loops named after regions of Src kinase (Fig. 4.2.5). The RT-loop and n-Src loop form an important part of the ligand binding interface (Fig. 4.2.5) [116-118].

Table 4.2.3. Data collection and refinement statistics.

	CrkII N-SH3 apo	**CrkII N-SH3:SgK269$^{1150-1162}$**
Data collection		
Wavelength (Å)	0.9537	0.9537
Space Group	$P\,4_14_12$	$P\,4_14_12$
Cell dimensions		
a, b, c (Å)	33.1 33.1 43.8	74.4 74.4 31.4
α, β, γ (°)	90 90 90	90 90 90
Resolution (Å)	43.8 - 1.55 (1.64 - 1.55)*	37.2-2.14 (2.25-2.13)*
Total reflections	93651 (14308)	133859 (20440)
Unique reflections	6917 (1102)	5306 (827)
Multiplicity	9.74	25.2
R_{merge}	6.80 (103)	10.1 (129)
I/σ I	20.7 (2.29)	25.1 (2.94)
Wilson B-factor	29.4	47.0
R-meas	4.6 (55.9)	10.4 (124)
Completeness (%)	98.9 (97.9)	99.9 (99.8)
Refinement		
Resolution (Å)	33.1-1.55 (1.60-1.55)*	28.8 - 2.13 (2.18 - 2.1)*
No. of reflections	6916 (686)	5455 (523)
R_{work}/R_{free}	17.4 (25.6)/ 21.2 (20.2)	21.5 (27.0)/26.1 (30.1)
No. of atoms		
Protein	475	551
Ligand/ion	-	-
Water	35	10
B-factors		
Protein	28.9	49.50
Ligand/ion	-	
Water	33.0	44.6
R.m.s deviations		
Bond lengths (Å)	0.006	0.013
Bond angles (°)	0.77	1.71
Ramachandran favoured (%)	96.4	92.2
Ramachandran allowed (%)	3.57	6.25
Ramachandran outliers (%)	0	1.56

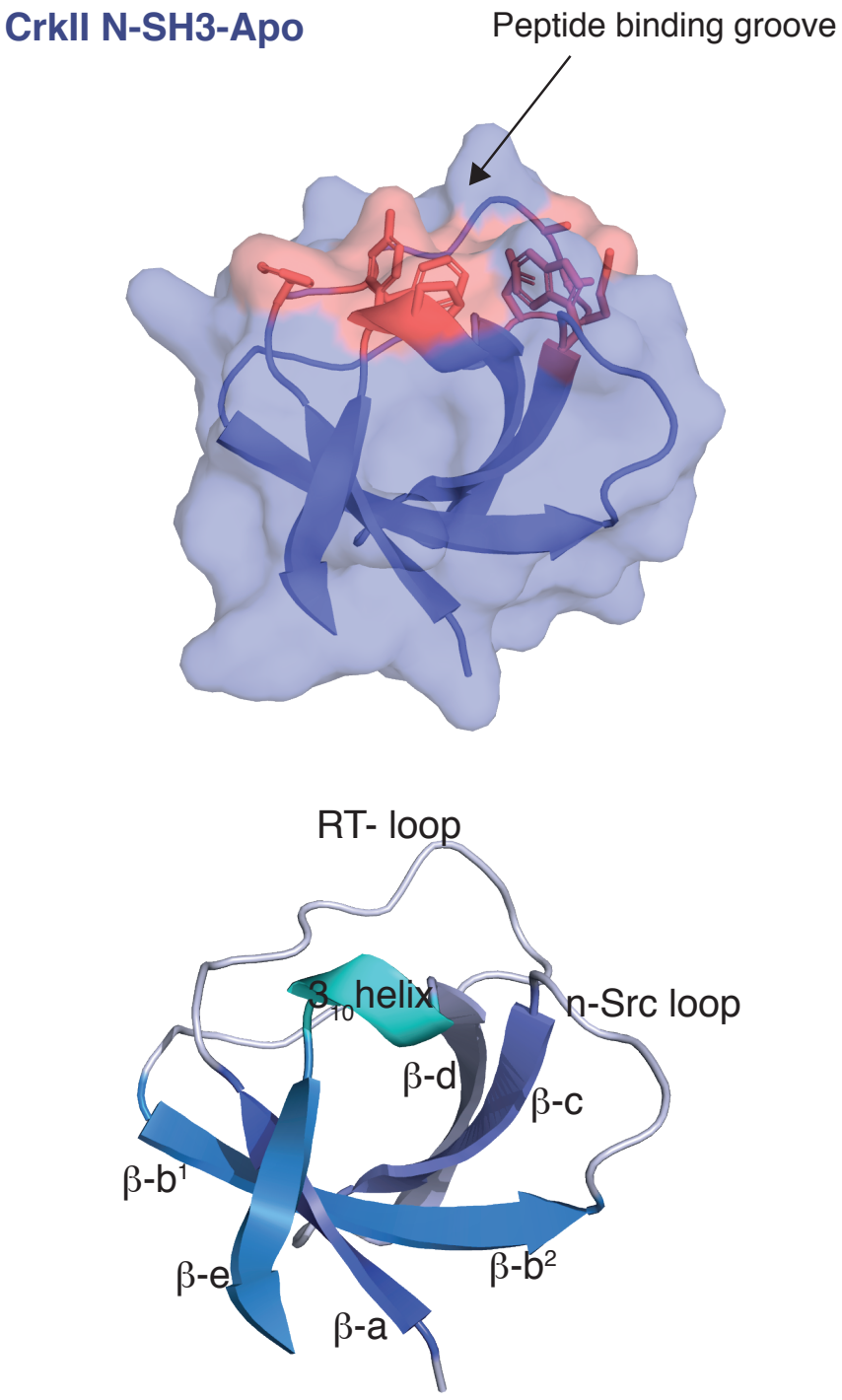


Figure 4.2.5. Structure of CrkII N-SH3 domain in apo form.
Crystal structure of CrkII N-SH3 in apo form. Surface representation of CrkII N-SH3 domain with the peptide binding groove highlighted in red and residues important for peptide binding represented as sticks. Secondary structural elements are mapped onto the domain, 3_{10} helix in green, β strands in shades of blue and loops in pale blue.

4.2.3.4 Structure determination of CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$ peptide

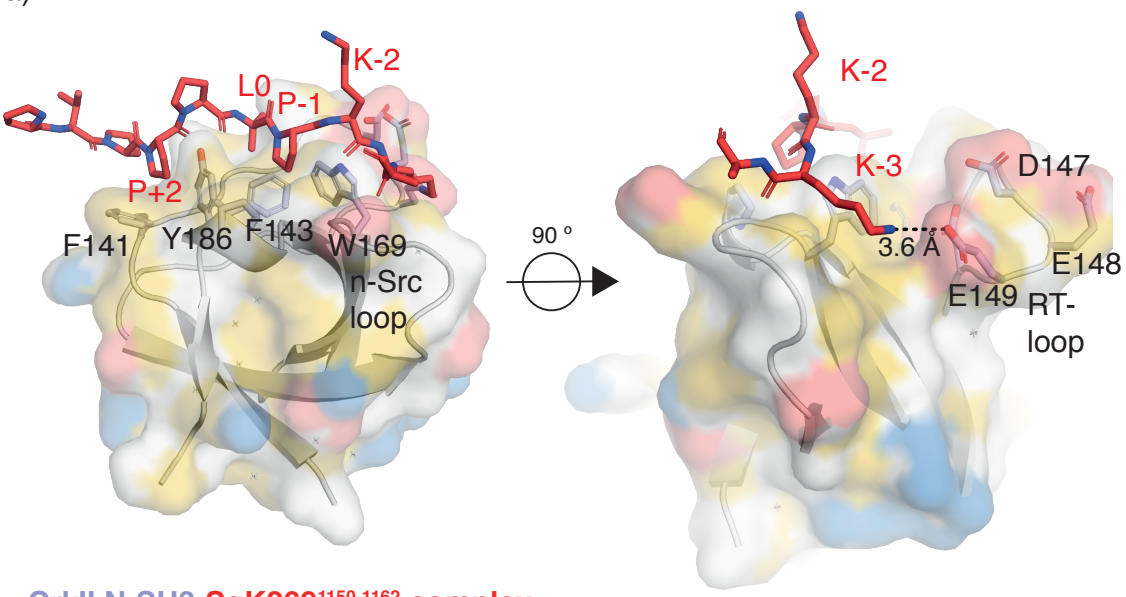
To understand the binding of the PEAK family PRM peptides to CrkII N-SH3 domain in greater detail, we solved the crystal structure of the CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$. Diffraction data for the CrkII N-SH3: SgK269$^{1150-1162}$ structure were processed in XDS and Buster to 2.1 Å [69, 70]. The structure was solved by molecular replacement (Phaser, Phenix Suite), using the CrkII N-SH3 domain (PDB: 5IH2) as a search model [110, 114]. The initial structure was improved by multiple rounds of building (Coot) and refinement in Phenix [74, 75]. The CrkII N-SH3 domain crystallised in the space group $P\,4_1\,2_1\,2$ with unit cell dimensions 74.3, 74.3, 31.2 Å and angles 90, 90, 90 °, with one molecule in the asymmetric unit. The data collection and refinement statistics are listed in Table 4.2.3.

The SgK269$^{1150-1162}$ peptide bound to the expected peptide binding site, forming contacts with residues from the RT-loop and n-Src loop of CrkII N-SH3 binding groove (Fig. 4.2.6) [118]. The peptide is labelled such that the central leucine in the motif PxLPxK is L0 and residues N-terminal of this are P+1, P+2 and so on and residues C-terminal are P-1, P-2. The residues of the SgK269$^{1150-1162}$ peptide that do not interact with the SH3 domain are P+1, T+4, P+5 and K-2, as they are not pointing towards the peptide binding groove. The P+2, L0 and P-1 residues of the SgK269$^{1150-1162}$ peptide interact with the residues of the hydrophobic binding interface of the CrkII N-SH3, Phe141, Phe143, Trp169, Pro183, Pro185 and Tyr186 (Fig. 4.2.6).

The peptide density beyond residue K-2 of the SgK269$^{1150-1162}$ peptide is not clear as the density extends into the symmetry related molecule (Fig. 4.2.6b). The crystal packing causes the symmetry related molecules to form close crystal contacts around a 2-fold rotational axis, at the K-2 and K-3 residues of the peptide (Fig. 4.2.6b). Thus, residue K-2 of SgK269$^{1150-1162}$ binds into the selectivity groove of the symmetry related CrkII N-SH3 domain. The selectivity groove is defined by a patch of acidic residues (Asp147, Glu149, Glu150) in the CrkII N-SH3 domain

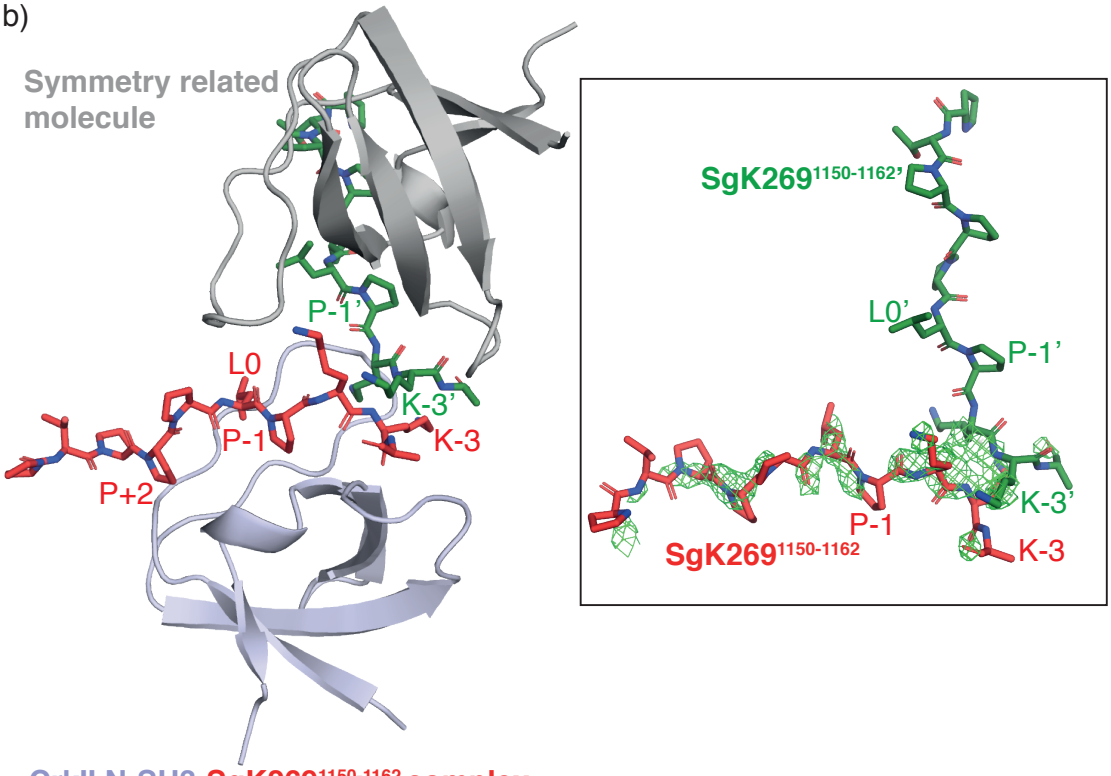
that usually bind the consensus lysine of the PRM (see section 4.2.3.7 and [110, 115, 119]). The K-3 residue of SgK269$^{1150\text{-}1162}$ is stabilised by the interaction with Glu149 of the CrkII N-SH3 domain RT-loop. To visualise the density of the peptide more clearly, we may need to design a shorter peptide or find different crystal forms where the peptide is not involved in crystal contacts.

a)



CrkII N-SH3:SgK269$^{1150\text{-}1162}$ complex

b)



CrkII N-SH3:SgK269$^{1150\text{-}1162}$ complex

Figure 4.2.6. Structure of CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$ peptide.
a) CrkII N-SH3 domain in complex with the SgK269$^{1150-1162}$ peptide. Peptide residues are numbered relative to the central leucine residue (designated position zero). Surface representation shows acidic residues in red, basic residues in blue, carbons not bonded to any oxygens or nitrogens in yellow (hydrophobic regions) and everything else in white. CrkII N-SH3 domain, light blue; SgK269$^{1150-1162}$ peptide, red.
b) CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$ and its symmetry related molecule. The SgK269$^{1150-1162}$ peptides come into close contact at K-2 and K-3 residues. Density overlap of the symmetry related peptides is shown in the inset. The Fo-Fc omit map (green mesh) contoured at 3σ around the peptide with a carve radius of 4 Å. CrkII N-SH3 domain, light blue; SgK269$^{1150-1162}$ peptide, red; symmetry related molecule of the CrkII N-SH3 domain, grey; peptide of symmetry related molecule, green.

4.2.3.5 Comparison of SgK269$^{1150-1162}$ peptide binding to the other PEAK family peptides

Our SPR interaction studies confirmed that the PRM peptides with the highest affinity for the CrkII N-SH3 domain were those containing the consensus motif PxLPxK (Table 4.2.2). Peptide SgK269$^{1117-1129}$, which demonstrated negligible binding to CrkII N-SH3 domain by SPR, lacks the consensus leucine and lysine residues. As seen in our structure of CrkII N-SH3:SgK269$^{1150-1162}$, both the consensus leucine and lysine residues are important for stabilising the peptide in the N-SH3 binding groove (Fig. 4.2.6a). The lysine residue (K-3) interacts with Glu149 and the leucine residue (L0) interacts with the hydrophobic residues Phe143, Trp169 and Tyr186 of the N-SH3 domain (Fig. 4.2.6a). In place of the leucine residue, SgK269$^{1117-1129}$ has a glutamine, which has a longer chain and is a polar residue (Fig. 4.2.6a). As the N-SH3 domain peptide binding groove is largely hydrophobic and aromatic (Fig. 4.2.6a surface hydrophobicity shown in yellow), a glutamine residue would likely be less well accommodated and cause unfavourable interactions. The lysine residue in the K-3 position is critical for CrkII N-SH3:PRM specificity and in SgK269$^{1117-1129}$ this position is also occupied by a glutamine residue (Fig. 4.2.6a). As the lysine residue interacts with the acidic residue Glu149, a polar but uncharged glutamine residue would lead to a loss of this critical interaction and instability of binding (Fig. 4.2.6a).

The SgK223$^{709-721}$ peptide bound to the CrkII N-SH3 domain, but with weaker affinity than the other peptides that follow the consensus motif for CrkII N-SH3 binding (Table 4.2.2). SgK223$^{709-721}$ peptide has a proline in place of the consensus leucine in positon L0, but maintains the consensus lysine residue in position K-3 (Fig. 4.2.6a). The proline residue, although non-polar, has no side chain, thus would be unable to fill the hydrophobic pocket as would the side-chain of a leucine residue.

As described, the main residues from SgK269$^{1150-1162}$ peptide that interact with CrkII N-SH3 domain are P-2, L0, P-1 and K-3 (Fig. 4.2.6). These residues are present in the three peptides that interact with CrkII N-SH3 with the highest affinity (PEAK3$^{54-66}$, SgK223$^{809-821}$, SgK269$^{1150-1162}$) and Abl peptide. Thus, we predict that the two peptides that we are yet to crystalise with CrkII N-SH3 domain may bind into the peptide binding site in a very similar way to the SgK269$^{1150-1162}$ peptide and Abl peptide.

4.2.3.6 Comparison of the crystal structure of CrkII N-SH3:SgK269$^{1150-1162}$ complex to the crystal structure of CrkII N-SH3 apo

The CrkII N-SH3 domain in complex with the SgK269$^{1150-1162}$ peptide largely resembles the apo structure of CrkII N-SH3 domain, with the canonical SH3 domain fold. We sought to understand if there were any differences between the apo CrkII N-SH3 structure and the CrkII N-SH3:SgK269$^{1150-1162}$ complex structure. The root mean square deviation (RMSD), based on Cα position, between the two structures is 0.27 Å^2, indicating they are structurally very similar (Fig. 4.2.7). The overlay of the two structures reveals no major conformational changes in the SH3 domain structures (Fig. 4.2.7). The hydrophobic aromatic residues (Phe141, Phe143, Trp169, Pro183, Pro185 and Tyr186) that stabilize the prolines P+2 and P-1 and leucine L0 of the peptide are in the same conformation in the apo structure and the peptide bound structure. This is most likely because these residues maintain the intramolecular structure of the N-SH3 domain, holding this groove open for peptide binding.

The residues from the RT- and nSrc-loops that constitute the main interactions with the PRM peptides align quite well, showing very similar conformations in the peptide bound and unbound structures (Fig. 4.2.7) [120]. The side chains of the acidic residues from the RT-loop (Asp147 and Glu148) are not visible in the apo structure, which could be due to flexibility of this loop (Fig. 4.2.7).

The primary difference between the crystal structure of CrkII N-SH3 apo and CrkII N-SH3 in complex with SgK269$^{1150\text{-}1162}$ lies in their packing. This crystal symmetry of CrkII N-SH3 apo precludes binding of a peptide to the peptide binding groove. This is highlighted in Figure 4.2.8a and b, showing an overlay of the CrkII N-SH3: SgK269$^{1150\text{-}1162}$ structure with the CrkII N-SH3 apo structure and its symmetry related molecule. Figure 4.2.8b demonstrates the unit cell of the CrkII N-SH3 domain and its three symmetry related molecules, again with the CrkII N-SH3:SgK269$^{1150\text{-}1162}$ structure overlaid.

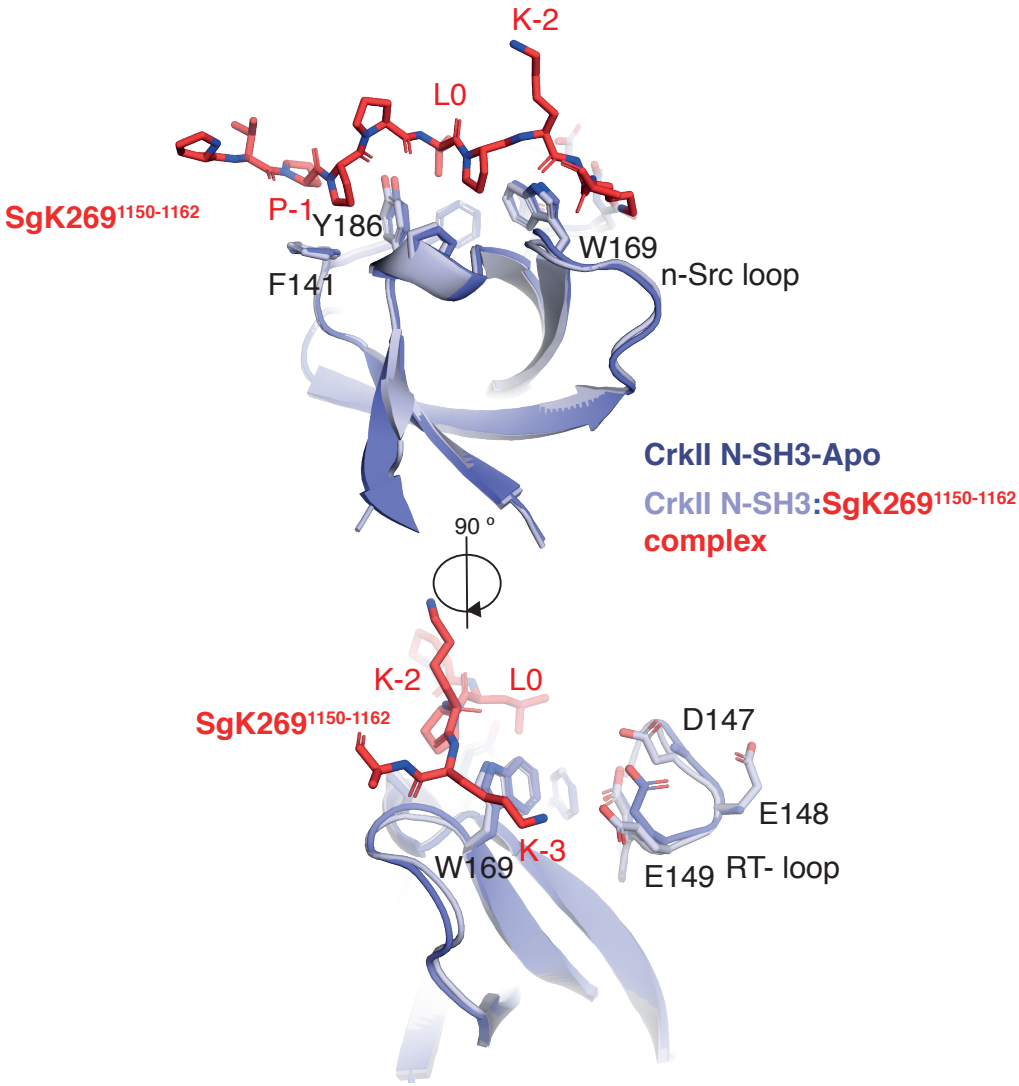


Figure 4.2.7. Comparison of CrkII N-SH3 apo domain to CrkII N-SH3 domain in complex with SgK269$^{1150\text{-}1162}$ peptide.
Overlay of CrkII N-SH3 domain in apo form and CrkII N-SH3 complexed with SgK269$^{1150\text{-}1162}$. Residues of the peptide are labelled with the leucine designated residue zero. The important residues in the peptide binding groove of the CrkII N-SH3 domain are shown and labelled. CrkII N-SH3 domain apo, dark blue; CrkII N-SH3- SgK269$^{1150\text{-}1162}$, light blue; SgK269$^{1150\text{-}1162}$ peptide, red.

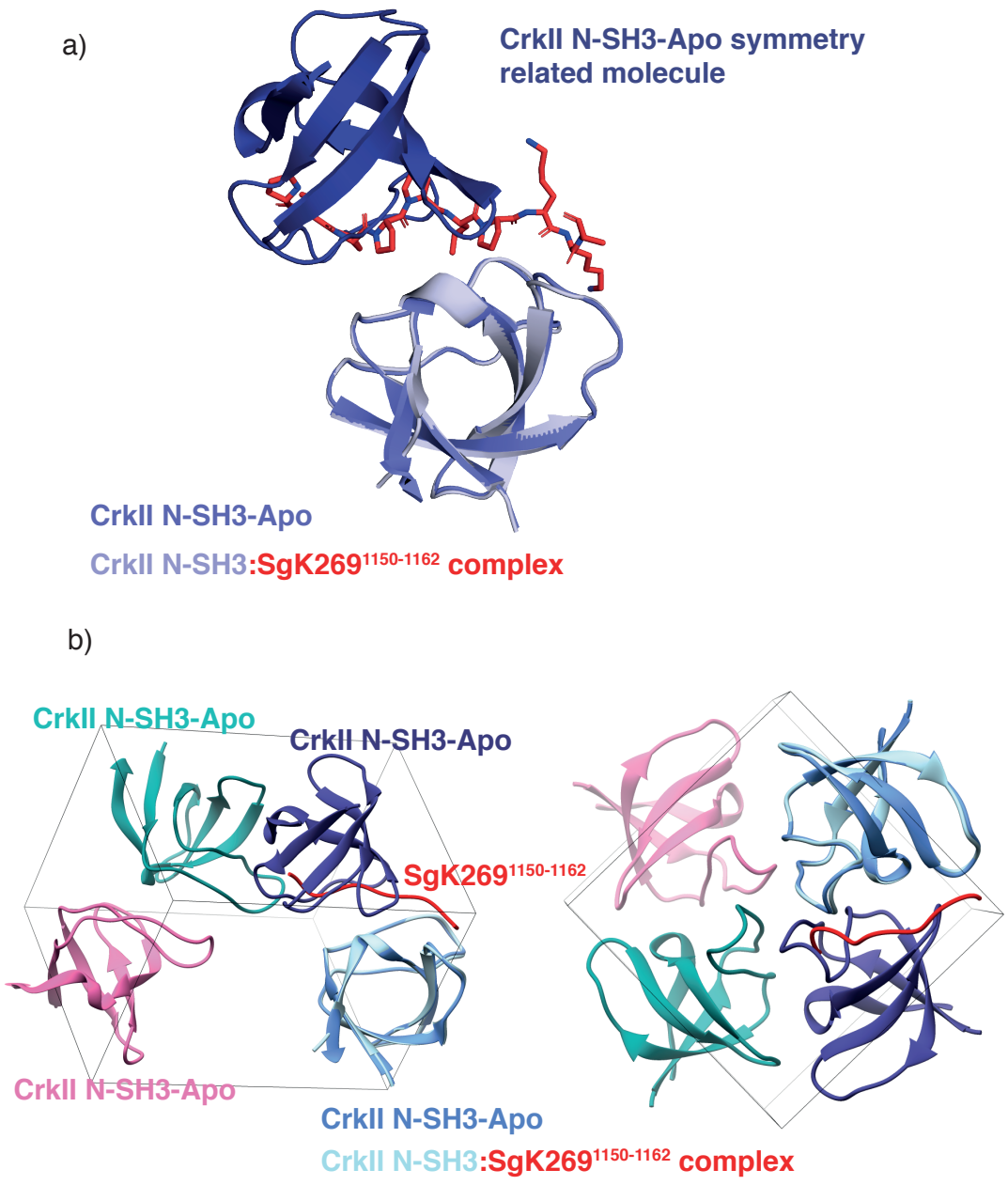


Figure 4.2.8. Crystal symmetry of CrkII N-SH3 apo.
a) Crystal structure of CrkII N-SH3 domain in apo form and its symmetry related molecule. CrkII N-SH3 domain in complex with SgK269$^{1150\text{-}1162}$ is overlaid on the CrkII N-SH3 apo structure to demonstrate that crystal packing would not allow peptide binding to CrkII N-SH3 apo. CrkII N-SH3 apo, medium blue, dark blue; CrkII N-SH3:SgK269$^{1150\text{-}1162}$, light blue; SgK269$^{1150\text{-}1162}$, red.
b) Unit cell and four symmetry related molecules of CrkII N-SH3 apo. Overlaid is the CrkII N-SH3: SgK269$^{1150\text{-}1162}$ complex, illustrating the symmetry in the unit cell, preventing peptide binding. CrkII N-SH3, medium blue, dark blue, green, pink; CrkII N-SH3: SgK269$^{1150\text{-}1162}$, blue; SgK269$^{1150\text{-}1162}$, red.

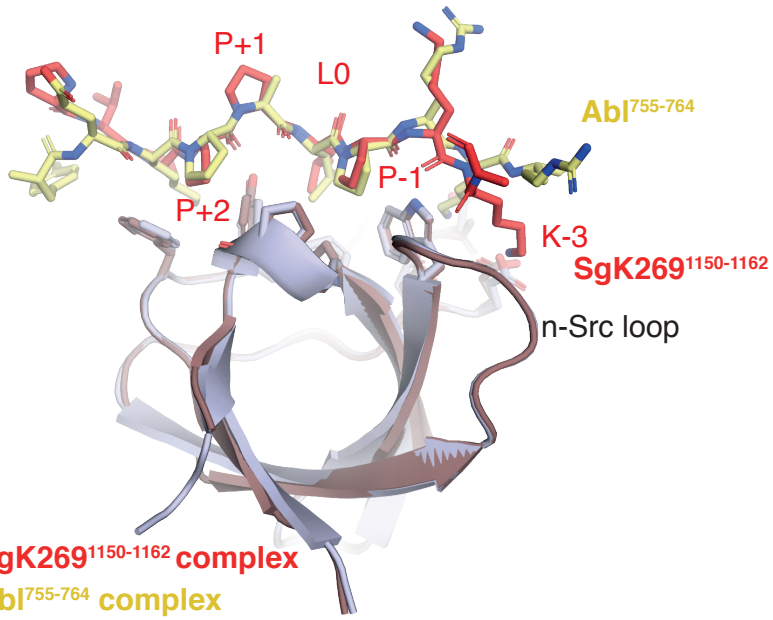
4.2.3.7 Comparison of the crystal structures of CrkII N-SH3 domain in complex with SgK269$^{1150-1162}$ peptide and CrkII N-SH3 domain in complex with Abl peptide

Recently, Bhatt *et al.* described the X-ray crystal structure of the CrkII N-SH3 domain in complex with a PRM peptide from Abl kinase (PDB: 5IH2) [110]. We compared the binding mode of the PRM peptides between our CrkII N-SH3:SgK269$^{1150-1162}$ co-crystal structure and that of the CrkII N-SH3:Abl PRM peptide co-crystal structure. Overlay of the two structures revealed that the CrkII N-SH3 domains are very similar with an RMSD of 0.18 Å^2 over Cα. Specifically, the RT- and nSrc- PRM binding loops have a similar conformation (Fig. 4.2.9a). The key hydrophobic residues of the RT- and nSrc-loops interact with the proline and leucine residues of both peptides as described for the SgK269$^{1150-1162}$ peptide previously (Fig. 4.2.6a).

The main difference in interactions begins at the K-2 lysine of SgK269$^{1150-1162}$ (an arginine in the Abl peptide). In the CrkII N-SH3:SgK269$^{1150-1162}$ structure, the K-2 lysine binds the selectivity groove of the symmetry related molecule, however in the CrkII N-SH3:Abl structure, the arginine points into the solvent (Fig. 4.2.9b). The Abl peptide and SgK269$^{1150-1162}$ peptide are in the same register, however K-3 of SgK269$^{1150-1162}$ does not bind as deeply in the selectivity groove as the Abl K-3 does, instead the K-2 of the symmetry related molecule occupies this position (Fig. 4.2.9b). This is due to the crystal packing of the CrkII N-SH3:SgK269$^{1150-1162}$ complex forming a symmetrical dimer as described in section 4.2.3.4. Considering that other structures of CrkII N-SH3 complexes (PDB: 1CKA, 5UL6 and 5L23) have demonstrated the K-3 lysine residue to bind in the same pocket as CrkII N-SH3:Abl PRM peptide, this suggests that SgK269$^{1150-1162}$ could bind in the same manner in solution [110, 115, 119] (PBD: 5L23 is unpublished, but available in the PDB). There is no evidence for the relevance of the symmetrical dimer of CrkII N-SH3:SgK269$^{1150-1162}$ in solution, especially as the concentration of CrkII in a biological system would be much lower than in the crystal. Obtaining different crystal forms of SgK269$^{1150-1162}$ peptide with the CrkII N-SH3 or crystallising the other PEAK family peptides in complex with the CrkII N-SH3

domain would help to understand the interactions of the PEAK family peptides with CrkII N-SH3 domain.

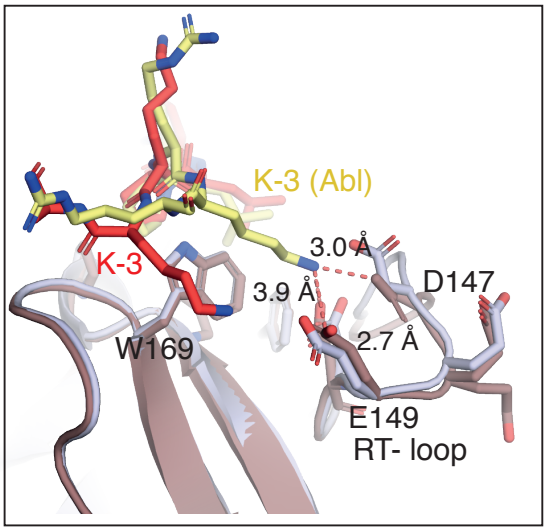
a)

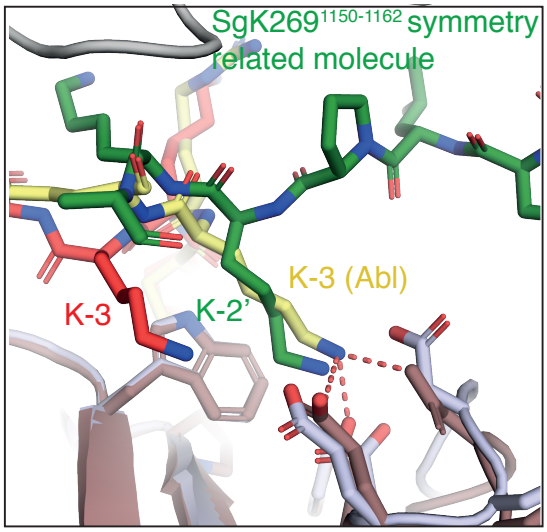


CrkII N-SH3:SgK269$^{1150-1162}$ complex
CrkII N-SH3:Abl$^{755-764}$ complex

90 °

b)





SgK269$^{1150-1162}$: PTPPPLPKKM
c-Abl PRM: YEKPALPRKR

Figure 4.2.9. Overlay of the CrkII N-SH3:Abl peptide bound structure with the CrkII N-SH3:SgK269$^{1150-1162}$ peptide bound structure.
a) Overlay of CrkII N-SH3 in complex with Abl peptide (PDB: 5IH2) and CrkII N-SH3 in complex with SgK269$^{1150-1162}$ peptide. The central leucine is designated the zero position. CrkII N-SH3-Abl, brown; Abl peptide, yellow; CrkII N-SH3: SgK269$^{1150-1162}$, light blue; SgK269$^{1150-1162}$ peptide, red.
b) Structural difference between the CrkII N-SH3:Abl complex and CrkII N-SH3:SgK269$^{1150-1162}$ complex. Lysine K-3 from Abl peptide interacts with Glu149 and Asp147 of the N-SH3 domain. Whereas, as shown in Figure 4.2.6, the SgK269$^{1150-1162}$ peptide lysine K-3 does not interact with these residues as lysine K-2 of SgK269$^{1150-1162}$ from the symmetry related molecule (shown in green) docks into this region of the peptide binding groove of the N-SH3 domain. Salt bridges shown as red dashed lines.

4.3 Discussion

In this Chapter, we investigated the interactions of the adaptor protein CrkII with the PEAK family members. Specifically, we characterised the interactions of the proline-rich regions of SgK223, SgK269 and PEAK3 with the CrkII N-SH3 domain biophysically and the interaction of CrkII N-SH3 domain with SgK269$^{1150\text{-}1162}$ PRM structurally. We confirmed that CrkII N-SH3 domain binding to PEAK family PRMs followed the consensus motif PxLPxK (as previously reported for interaction of other PRMs with the CrkII N-SH3 domains), and PRMs with a central proline residue instead of a leucine residue still allowed binding of this motif to CrkII, however at a five-fold decreased affinity (13 μM and 2.4 μM, respectively). The affinity of the PEAK family for the CrkII N-SH3 domain is approximately 1 μM (Table 4.2.2), which is similar to the previously described affinities of other PRM containing proteins with CrkII N-SH3 domain, such as Abl, c-Jun N-terminal Kinase (JNK) and C3G [110, 115, 119, 121-123].

Currently in the protein data bank there are only structures of CrkII N-SH3 domain in complex with Abl peptide, an influenza protein (NS1) peptide and C3G peptide (PDB: 5IH2, 5UL6, 1CKA) [110, 119]. However, through immunoprecipitation and ELISA experiments, groups have demonstrated additional proteins that interact with CrkII N-SH3, such as JNK and Son of Sevenless (Sos) [110, 115, 119, 121-123]. Although the initial experiments investigating the binding of C3G and Sos with CrkII characterised the CrkII N-SH3 consensus sequence, the investigations carried out by Bhatt *et al.* and Wu *et al.* provided structural insights into why minor alterations to the consensus sequence can lead to large changes in the affinity of PRM binding to the CrkII N-SH3 domain [110, 115]. They demonstrated that the largely hydrophobic PRM binding pocket of CrkII N-SH3 interacts with the proline and leucine residues of the PRM to stabilise binding and the three key acidic residues interact with the consensus lysine of the PRM, bringing interaction specificity [110]. Our structural data confirmed their findings, with the PRMs of the PEAK family sharing the consensus leucine (L0) and lysine (K-3) residues. Additionally, our structural data demonstrated that the leucine (L0) residue of the

PRM participates in critical hydrophobic interactions with CrkII N-SH3 domain. Although, possibly an artifact of crystal packing within our structure, the consensus lysine K-3 does not interact with all three acidic residues from the N-SH3 domain, rather just one of the residues, Glu149 (Fig. 4.2.6a, Fig. 4.2.9b). It is quite clear, however, that this region interacts with a lysine residue in position K-3, as the size of the groove would not accommodate an arginine residue (Fig. 4.2.9b). The side chains of the arginine would clash with the side chains of the acidic CrkII residues. Additionally, as this is the only region of the peptide binding pocket that is acidic, it would unlikely bind a hydrophobic residue.

As there are a large number of SH3 domains in the proteome, as well as PRM containing proteins, it is critical that SH3 domain-PRM interactions are regulated to prevent aberrant signalling [97]. PRM containing proteins can either interact with specific SH3 domains or behave promiscuously, interacting with multiple SH3 domains to integrate and converge signalling pathways. For example, Sos is a promiscuous SH3 domain binder, interacting with CrkII, Grb2, Src and Nck SH3 domains [123]. Sos has a lower affinity for each SH3 domain, with an arginine in place of the consensus lysine for CrkII N-SH3 domain binding [123]. Through its interactions with multiple SH3 domains, Sos integrates different receptor signalling pathways for subsequent Ras signalling [123]. Comparatively, the PRMs C3G and Abl, are specific to the N-SH3 domain of CrkII [110, 115, 123].

It is yet to be determined if the PEAK family are promiscuous or specific SH3 domain binders. As demonstrated in Chapter 4, the PEAK family interactions with CrkII N-SH3 domain have an affinity of ~1 μM, which is a tight interaction for an SH3 domain with a PRM, however not as tight as a SH2 domain interaction (K_D ~10^{-3} μM), for instance [124]. The SPR experiments demonstrated that the interaction between the PEAK family and CrkII N-SH3 domain had very fast association and dissociations (on/off rates). The affinity of interaction and association/dissociation rates between the CrkII and the PEAK family could be important for their roles in signalling. Specifically, the fast association and

dissociation could allow multiple molecules of the PEAK family to interact with CrkII, thus propagating signalling. In the case the PEAK family proteins had slower association/dissociation rates, they would remain bound for much longer, preventing interaction with different molecules and perhaps a decrease in signalling output. Interestingly, the N-terminal domain of SgK223, that we began characterising in Chapter 3, has a proline-rich motif that follows the consensus motif for CrkII N-SH3 domain, except it has an arginine in place of the K-3 lysine. We have yet to investigate if this region can interact with CrkII N-SH3 domain, however if it is similar to the Sos PRM, it could possibly be a promiscuous SH3 domain interactor. Further investigations will also need to be carried out to characterise the interaction of the PEAK family with other potential SH3 domains, such as Grb2, Src, p130Cas, Sorbs1, Csk and DOCK1. This will provide insight into whether they are promiscuous SH3 binders or specific to CrkII. Understanding the interactions of the PEAK family with SH3 domains will help to characterise the role of SgK223, SgK269 and PEAK3 in signalling pathways.

A possible limitation of this study is that all of the experiments in Chapter 4 were carried out on the N-SH3 domain of CrkII, in isolation from the other domains of CrkII, and short peptides of the PEAK family PRMs. To overcome this and verify if this interaction is representative of the protein-protein interaction occurring in cells, we could use the full-length CrkII protein and the recombinant PEAK family proteins for the SPR binding assays.

The literature has provided insights into potential interactors of SgK223, SgK269 and PEAK3, such as Csk (SgK223), Grb2 and Shc1 (SgK269), however very little structural and biophysical data have been obtained characterising the interactions of this family at focal adhesions. The interactions that have been discovered through immunoprecipitation experiments may not be direct, as there could be an intermediary protein facilitating the interactions. We are yet to understand what drives localisation of the PEAK family to focal adhesions, whether hetero-dimerisation of SgK223 and SgK269 is required, or interaction with an adaptor, such as CrkII, is required. Additionally, it has been reported that

SgK269 localises with F-actin, however no direct actin binding site has been reported on SgK269 [3]. Possibly the localisation of SgK269 with actin is mediated through an adaptor such as CrkII, linking SgK269 to paxillin or vinculin, which have been demonstrated to interact directly with actin.

The next steps to validate these interactions in cells would be to undertake co-localisation microscopy of CrkII and the PEAK family. Mutagenesis of the PEAK family PRMs to prevent their interaction with CrkII could shed light on the importance of their interactions for PEAK family localisation and signalling.

We could also investigate the importance of PEAK family interactions with CrkII on cell migration and morphology changes. Both CrkII and SgK223/SgK269 can enhance cell migration and cause an elongated cell morphology upon overexpression [1, 6, 107]. It was also demonstrated recently that dimerised PEAK3 reduced the propensity of CrkII to elongate cell morphology [52]. However, it is unknown whether the co-expression of CrkII and SgK223 and SgK269 enhances cell migration and cell elongation further or if there is a decrease in cell migration and morphology when they cannot interact.

5.0 An investigation into the role of SgK223 and SgK269 homo- and hetero-associations for their localisation

5.1 Introduction

SgK223 and SgK269, as previously described, are pseudokinase scaffolds that promote cell migration and alter cell morphology [1, 6]. SgK223 and SgK269 have been demonstrated to localise to focal adhesions [2, 5]. Here, they act as scaffolds of multiple critical signalling pathways, such as Src kinase signalling, MAPK signalling and EGF stimulated signalling [2, 5, 56]. Focal adhesions are the important contact point between the cell and the extracellular matrix (ECM) [125]. Numerous proteins are found at focal adhesions, ranging from enzymes to adaptors, and all carry out roles to promote signalling for modulating cell adhesion, migration and morphology in response to cellular and environmental cues [125, 126]. For example, the pseudokinase scaffold ILK plays an important role at focal adhesions, interacting with the adaptors PINCH1 and α-parvin to modulate cell adhesions dynamics. Microscopy studies on the ILK-PINCH1-α-parvin (IPP) complex have demonstrated that disrupting their interactions abrogates the localisation of the ILK to focal adhesions, thus impairing cell spreading and cell migration [127].

Upon phosphorylation of Tyr411 in the PEST linker of SgK223, the Src family kinase Csk can interact, through its SH2 domain, with SgK223. Using immunofluorescence imaging, SgK223 and Csk were found to localise at focal adhesions with the adaptor protein vinculin, as well as to the cytosol [5]. It was suggested that vinculin played a role in localising the SgK223-Csk complex to focal adhesions, however this was not experimentally verified [5]. Using live-cell TIRF microscopy, in contrast to the immunofluorescence studies on SgK223, SgK269 fused N-terminally to GFP was found to co-localise with the focal adhesion protein paxillin and actin, as well as to the cytosol [2]. Additionally, it was demonstrated that for SgK269 to localise to focal adhesions, dynamic phosphorylation and dephosphorylation of the residue Tyr665 of SgK269 PEST linker was required [2]. Together, these studies provided strong evidence that the

localisation of SgK223 and SgK269 to focal adhesions is based on key motifs in their PEST linker for protein-protein interactions, such as Tyr411 in SgK223 for its interaction with Csk, and Tyr665 in SgK269. However, yet to be investigated is the role homo- and hetero- association of SgK223 and SgK269 plays in their localisation. The aim of this study was therefore to express the dimerisation and oligomerisation mutants of SgK223 and SgK269 characterised in Chapter 3, and investigate their localisation in cells compared to WT SgK223 and SgK269.

5.2 Results

5.2.1 Cloning of SgK223 and SgK269 for mammalian cell expression

Full-length SgK223 and SgK269 genes were synthesised by Genscript for mammalian cells expression and cloned into a pMIH vector, kindly gifted by Dr. Kate McArthur (see Chapter 2 and Fig. 2.1 and Table 8.1 in the Appendices). As we previously synthesised constructs comprising the pseudokinase domain and the N- and C-terminal flanking helices with a BamHI restriction site prior to the N-terminal helix and a HindIII restriction site following the C-terminal helices, we incorporated in the full-length gene internal BamHI and HindIII restriction site to allow for sub-cloning of all our mutants into the full-length SgK223 and SgK269 (Fig. 5.2.1a).

In addition, SgK223 was fused N-terminally to mNeonGreen (Allele Biotechnology), the brightest green fluorescent protein available with an excitation maximum/emission maximum (ex/em) of 506/517 nm (Fig. 5.2.1b) [128]. SgK269 was fused N-terminally to a Halo tag (Promega), which only fluoresces in the presence of a Halo dye (Fig. 5.2.1b) (see Chapter 2 for experimental details). There are multiple different ex/em wavelength Halo dyes available, providing flexibility for imaging. As we planned to image SgK223 and SgK269 together, as well as a number of focal adhesion associating proteins and cell structural components, like actin, this flexibility of fluorophore choice for SgK269 was necessary.

To visualise focal adhesions in cells, we used paxillin fused to either mApple (ex/em 568/592) or mEmerald (ex/em 487/509) fluorophores (Addgene #54935, #54219, a gift from Michael Davidson). Paxillin has been widely used as a marker of focal adhesions, and there are many readily available and well-established expression plasmids, such as those we used. To stain actin, we used the dye SIR-Actin, in the far-red channel (652/674 ex/em).

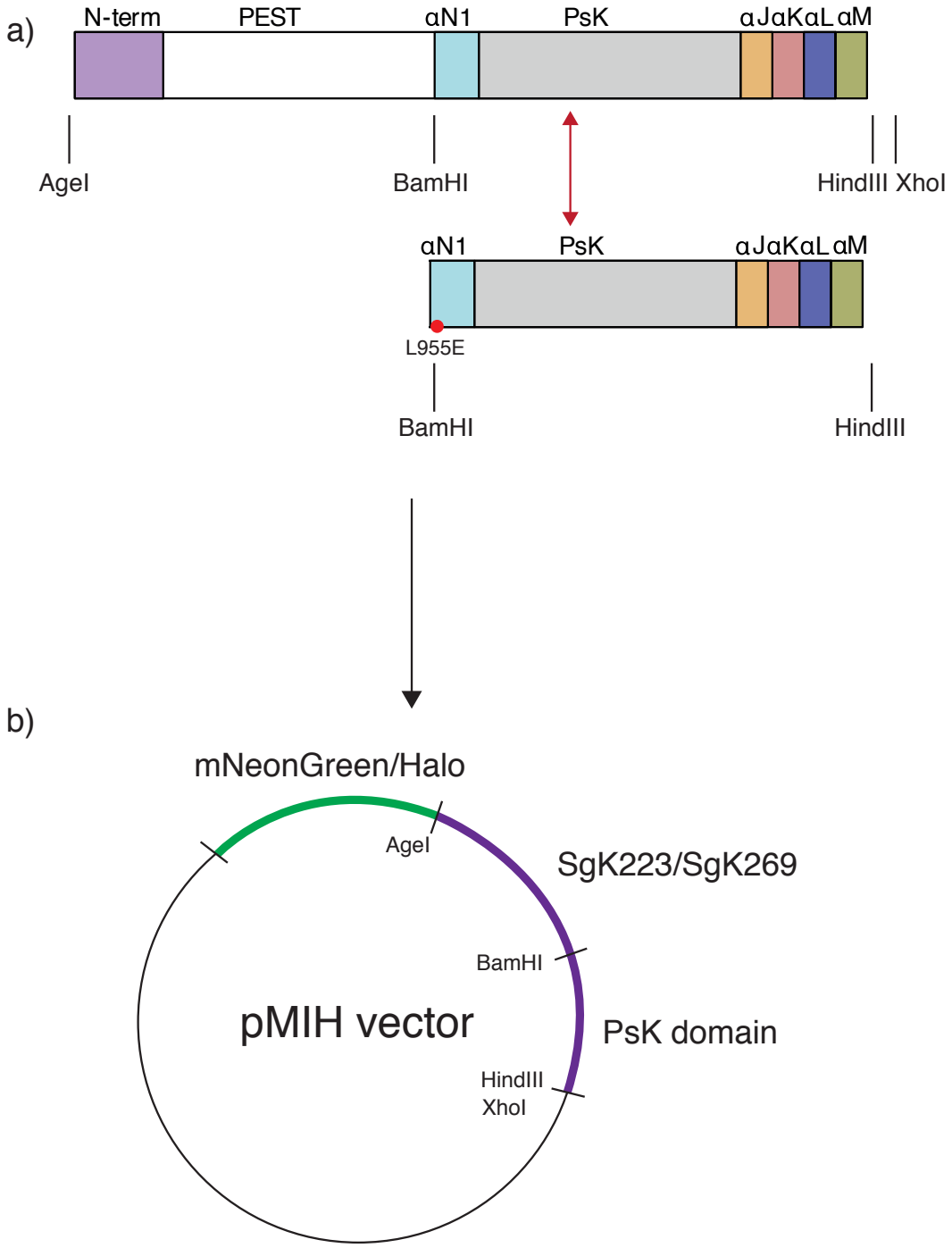


Figure 5.2.1. Cloning scheme for SgK223 and SgK269 expression in mammalian cells.
a) Linear domain layout of SgK223/SgK269. The restriction sites are mapped onto the layout to demonstrate the C-terminal domain swapping designed to incorporate the mutations in the pseudokinase and dimerisation domain.
b) pMIH vector map demonstrating SgK223/SgK269 gene and the N-terminal fusion fluorophore, either Halo tag or mNeonGreen.

Table 5.2.1. Constructs designed and cloned for imaging experiments.

mNeonGreen SgK223	Halo SgK269	Focal adhesion markers *from Addgene
SgK223 WT	SgK269 WT	mApple paxillin
SgK223 L955E, A1367D mutant	SgK269 I1290E, A1707D mutant	mEmerald paxillin
SgK223 L955E, F1271A, A1367D mutant	SgK269 I1290E, F1609A, A1707D mutant	
SgK223 L966A mutant		

5.2.2 Optimisation of the transient transfection of SgK223 in human cell lines

As we aimed to carry out live cell imaging, we needed to establish the mammalian cell expression of SgK223 and SgK269 fused to mNeonGreen or Halo tag, respectively. To create a stable cell line expressing SgK223 and SgK269, we used retroviral transfection. However, this resulted in cells downregulating SgK223 and SgK269 expression over time, as demonstrated through microscopy images of cells at different passage times. This was not ideal, as very low passage number cells would be needed for imaging, requiring fresh thawing of cell lines for each experiment and thus, several days for the cells to recover. Therefore, we carried out transient transfections, whereby cells are transfected and only express the DNA products for a short period of time, often 1-7 days. Transfection was performed 12-24 hours prior to imaging and there was no requirement for selection of cells expressing the transfected gene. To optimise the transfection protocol, cells were transfected initially with just SgK223 fused N-terminally to mNeonGreen, before carrying out experiments on both SgK223 and SgK269.

As SgK223 and SgK269 have been reported to localise to focal adhesions and play a role in promoting cell migration signalling and elongating cell morphology,

we aimed to image cells known to form prominent focal adhesions, thus we chose the metastatic breast adenocarcinoma cell line MDA-MB-231 [54, 129, 130].

We transfected the MDA-MB-231 cells with either SgK223 mNeonGreen or mApple paxillin (a focal adhesion marker protein) or both, using the lipid based transfection reagent Lipofectamine 3000 (Fig. 5.2.2a). Cells were seeded into an 8-well chamber dish and incubated for 24-30 hours before the DNA and Lipofectamine 3000 were added, and then incubated for a further 12 hours (Fig. 5.2.2a). SIR-Actin, a fluorescent actin marker, was added to the cells prior to imaging, since SgK223 and SgK269 have been shown to localise to actin and focal adhesions occurs at the end of actin filaments (Fig. 5.2.2a) [2, 60]. Imaging was performed using the Zeiss LSM 880 confocal microscope with Airyscan to visualise SgK223 localisation. Figure 5.2.2b demonstrates the MDA-MB-231 cells stained with paxillin in magenta and actin in yellow. The field of view has multiple cells present and the actin staining helps to define cell boundaries. All of the cells in Figure 5.2.2b are expressing paxillin, however in general there was a low transfection efficiency of paxillin within the overall cell population. In the MDA-MB-231 cells, paxillin localisation demonstrated the presence of focal adhesions, which are generally at the ends of the actin filaments, indicated by white arrows in the merge image (Fig. 5.2.2b). However, there was very low expression of SgK223 in these cells demonstrated through a lack of mNeonGreen fluorescence. Any cells that were expressing SgK223 were not co-expressing paxillin. Thus, we could not obtain any images demonstrating the localisation of SgK223 at focal adhesions, using paxillin as a marker of focal adhesions. Additionally, a large degree of cell death was observed upon treatment of cells with Lipofectamine 3000. Cells in the process of dying become rounded and lifted from the surface of the dish, losing their spread-out cell morphology that usually demonstrates obvious focal adhesions.

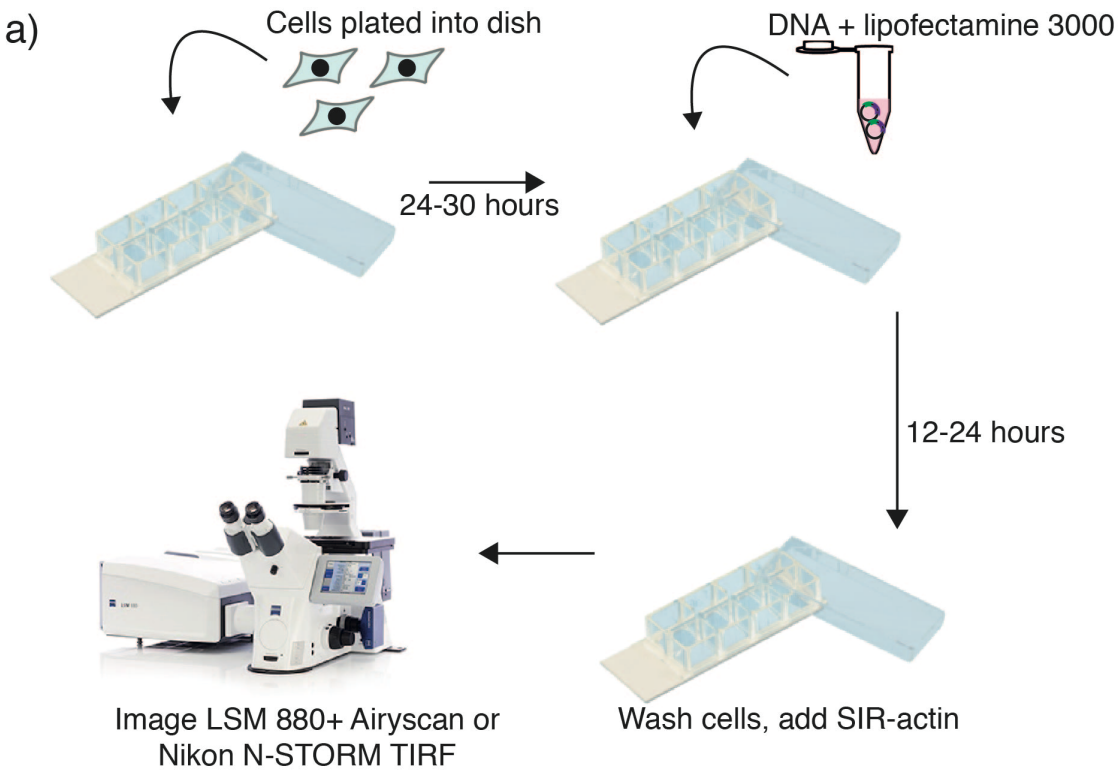
As paxillin has been shown to express well in transient transfection systems and in MDA-MB-231 cells usually have a high transfection efficiency, we aimed to improve the transfection of SgK223 and paxillin in this cell line [131].

To optimise transfection of SgK223 and paxillin, we initially trialed a number of conditions including: altering the number of cells in each well, the dish type (8-well chamber or 35-mm dish), the amount of DNA-to-Lipofectamine 3000 reagent ratio and the length of incubation of the cells with Lipofectamine 3000 and the DNA. As the optimisation steps did not improve the transfection efficiency of SgK223 and paxillin in MDA-MB-231 cells, we trialed different transfection reagents, such as Fugene and Lipofectamine 2000. Fugene caused a significant decrease in cell death compared to Lipofectamine 3000. However, upon imaging, the cells transfected with Fugene demonstrated no expression of paxillin or SgK223, indicating transfection efficiency was very low or failed. Lipofectamine 3000 is the new and superior reagent from the Lipofectamine products and is marketed to have a better transfection efficiency than Lipofectamine 2000. In our hands, we found that both reagents caused a similar degree of cell death and similar transfection efficiencies in MDA-MB-231 cells. Thus, we continued to use Lipofectamine 3000 as our transfection reagent and investigated using different cell lines to increase SgK223 and paxillin transfection efficiency.

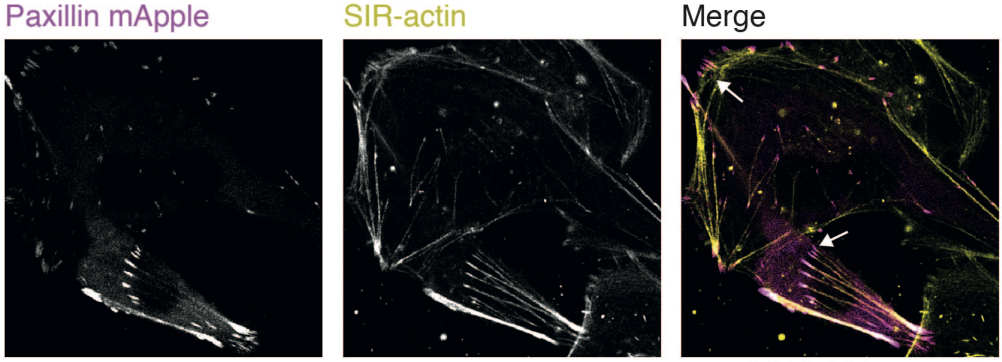
We chose DLD1 colorectal cancer cells and U2OS osteosarcoma cells and transfected them using Lipofectamine 3000, as these have been shown to form focal adhesions, [132, 133]. Cells were imaged using total internal reflection fluorescence (TIRF) microscopy. TIRF is a microscopy technique designed to increase the signal-to-noise ratio and provide high axial resolution compared to wide-field microscopy, by creating total internal reflection of the excitation light. In contrast to wide-field techniques, TIRF creates evanescent light that does not excite the whole sample, rather just the 100-200 nm distance from the coverslip, making it an ideal technique to image processes happening at or very close to the plasma membrane, such as focal adhesions [134].

The DLD1 cells did not appear to express either paxillin or SgK223. The U2OS cells showed paxillin staining to focal adhesions, however the efficiency (<10% cells transfected) was not as high as we would require for data collection (>70%

cells transfected) (Fig. 5.2.2c). Similarly, the transfection efficiency of SgK223 was very low and we found only one U2OS cell that demonstrated expression of both paxillin (magenta) and SgK223 (cyan) (Fig. 5.2.2c). This cell demonstrated that SgK223 and paxillin co-localise to focal adhesions (Fig. 5.2.2c). However, due to the poor expression of SgK223 and paxillin, we could not continue to investigate SgK223 localisation using this cell line.

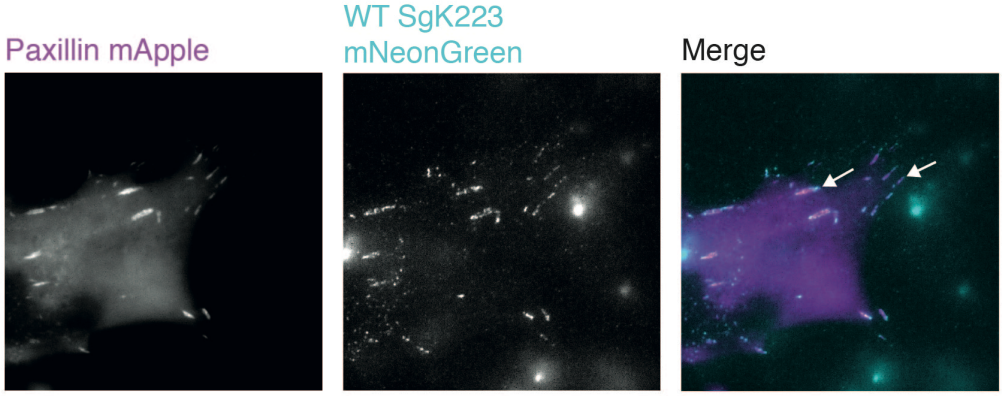


b) CONFOCAL AIRYSCAN



MDA-MB-231 cells

c) TIRF



U2OS cells

Figure 5.2.2. Optimisation of SgK223-mNeonGreen transfection in MDA-MB-231 and U2OS cell lines.
a) Experimental set up for transient transfection of all cells lines. Cells were seeded into the appropriate dish (shown here is an 8-well chamber) and following incubation they were treated with the transfection reagent (shown here is Lipofectamine 3000). After treatment, cells were washed and incubated briefly with the SIR-actin dye prior to imaging.
b) Confocal Airyscan images of MDA-MB-231 cells transfected with mApple paxillin and SgK223 mNeonGreen, and stained with SIR-actin. Cells failed to co-express paxillin and SgK223, thus this representative cell demonstrates focal adhesion formation shown by paxillin staining (magenta) and actin (yellow).
c) TIRF images of U2OS cells transfected with mApple paxillin and SgK223 mNeonGreen. Representative cell demonstrating focal adhesion formation by paxillin staining (magenta). SgK223 (cyan) co-localised with paxillin to focal adhesions. Single colour images are shown in greyscale for greater visual contrast.

5.2.3 Expression and localisation SgK223 WT and paxillin in HEK293 cells

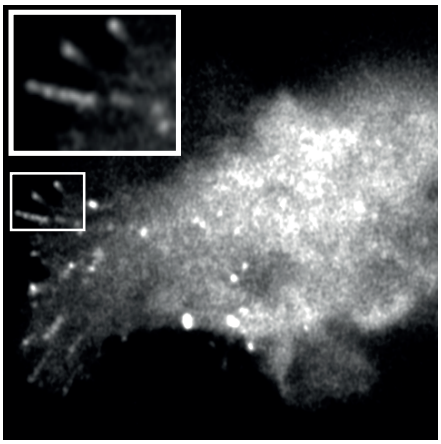
It is well established that Human Embryonic Kidney cells (HEK293) have a very high transfection efficiency and are often used for large scale protein production [135]. We transfected SgK223 mNeonGreen and mApple paxillin into HEK293 cells using the standard Lipofectamine 3000 protocol (Fig. 5.2.2a). After transfection and just prior to image acquisition, we incubated the cells with SIR-Actin dye. Cells were imaged using the Zeiss LSM 880 confocal microscope with Airyscan. Figure 5.2.3a demonstrates the localisation of paxillin (magenta) to focal adhesions. Although only a proportion of HEK293 cells (~20%) demonstrated focal adhesions, the transfection efficiency of paxillin was markedly improved. The SIR-actin stain was not visible in the HEK293 cells; thus, we did not acquire data in this channel. HEK293 cells transfected with both paxillin (magenta) and SgK223 (cyan) demonstrated high transfection efficiency of both plasmids (Fig. 5.2.3a).

SgK223 co-localised with paxillin at focal adhesions in HEK293 cells, as shown in Figure 5.2.3a. However, large background fluorescence was present in the cytosol due to the very high expression of SgK223 and paxillin. The staining in Figure 5.2.3a was representative of the cell population.

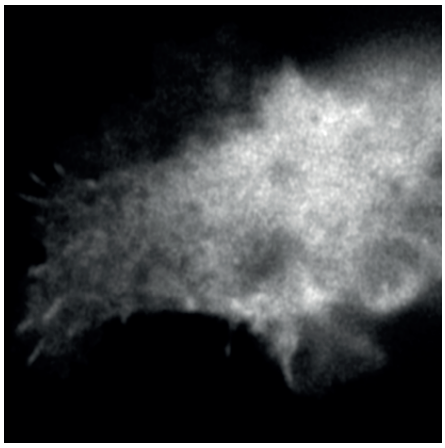
We next used TIRF microscopy to image the focal adhesions with high-level contrast to eliminate the background fluorescence (Fig. 5.2.3b). We carried out the experimental set up as shown in Figure 5.2.2a, and imaged the HEK293 cells expressing mApple paxillin and SgK223 mNeonGreen on the Nikon dSTORM microscope equipped with TIRF microscopy (Fig. 5.2.3b) As a comparison of techniques, the images of HEK293 cells expressing SgK223 acquired using TIRF microscopy compared to the images acquired using confocal Airyscan (Fig. 5.2.3). As expected, the TIRF images have significantly less background fluorescence compared to confocal microscopy.

a) CONFOCAL AIRYSCAN

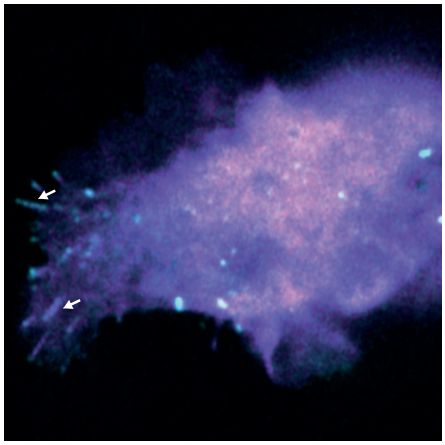
WT SgK223 mNeonGreen



Paxillin mApple



Merge



b) TIRF

WT SgK223 mNeonGreen

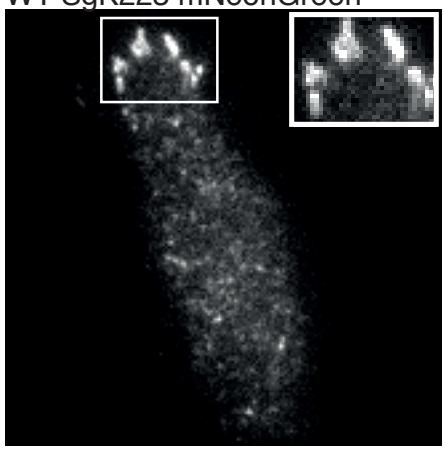


Figure 5.2.3. Localisation of SgK223 mNeonGreen in HEK293 cells.
a) Confocal Airyscan images of HEK293 cells transfected with mApple paxillin and SgK223 mNeonGreen. SgK223 mNeonGreen (cyan) co-localised with mApple paxillin (magenta) to focal adhesions, as highlighted by the magnified panel on the SgK223 mNeonGreen image and by white arrows on the merge image. A single plane from the z-stack is shown.
b) TIRF image of SgK223 mNeonGreen localisation as highlighted by the magnified panel on the SgK223 mNeonGreen image. TIRF microscopy demonstrates decreased background fluorescence compared to confocal microscopy in a). Single colour images are shown in greyscale for greater visual contrast.

5.2.4 Expression and localisation of SgK223 dimerisation and oligomerisation mutant in HEK293 cells

As we had verified the localisation of SgK223 to focal adhesions in HEK293 cells, we next investigated the impact of disrupting dimerisation and oligomerisation of SgK223 on its localisation, using the SgK223 triple mutant (L955E, A1367D, F1609A) previously characterised in Chapter 3 and shown to disrupt dimerisation and oligomerisation (Fig. 5.2.4b).

To generate the SgK223 mutant in the full-length protein, we employed the cloning strategy described in section 5.2.1 (Fig. 5.2.1). Using restriction digestion, the WT pseudokinase domain and flanking N- and C-terminal helices were excised from the SgK223 gene in the pMIH expression vector. The SgK223 triple mutant pseudokinase domain and flanking N- and C-terminal helices were then sub-cloned into the SgK223 full-length gene. The mutant SgK223 was then co-transfected with paxillin into HEK293 cells, as previously described. Control cells were transfected with SgK223 WT and paxillin as a localisation comparison.

We imaged these cells using confocal Airyscan as TIRF microscopy can only image the first 100 nm of the cell attached to the coverslip. If the mutant SgK223 localised further into the 3D volume of the cell, we would be unable to image its localisation using TIRF microscopy. Cells co-transfected with mutant SgK223 and paxillin demonstrated that the transfection efficiency and expression level of the mutant SgK223 was very similar to SgK223 WT, indicating the mutations to SgK223 are not affecting the stability of the protein in cells (Fig. 5.2.4a, b). Mutant SgK223 appeared to localise to focal adhesions and the cytosol, similarly to the SgK223 WT (Fig. 5.2.4a, b). The focal adhesions in the cells imaged expressing the mutant SgK223 appeared less prominent than those in the cells expressing SgK223 WT (Fig. 5.2.4b). However, as only ~20 cells expressing the mutant and SgK223 WT were imaged, we cannot draw conclusions about the phenotype of the cells expressing the mutant SgK223 compared to the SgK223 WT. Future experiments would involve imaging more cells over a number of different independent experiments, using TIRF microscopy. This would provide a larger

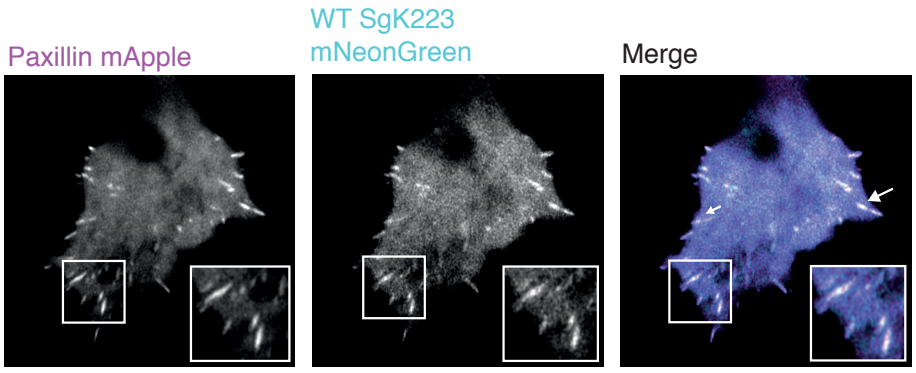
data set for analysis, as well as higher quality images with higher signal-to-noise ratio and contrast.

To evaluate if the mutant SgK223 localises to focal adhesions to the same extent as the SgK223 WT, we quantified their co-localisation with paxillin, using the image analysis software Imaris. The quantification method plots the fluorescence of SgK223 versus paxillin (Fig.5.2.4c). High co-localisation between the two proteins would result in overlap of the two channels, creating a positive linear plot (Fig. 5.2.4c). We calculated the Pearson's coefficient of the plot for each image to determine the correlation of the co-localisation of the two channels (Fig. 5.2.4c). A high Pearson's coefficient (>0.5) indicates strong correlation and good co-localisation of the two proteins. The Pearson's coefficients calculated from the plot of SgK223 WT versus paxillin and the plot of mutant SgK223 versus paxillin demonstrated similar levels of co-localisation between the two proteins (Fig. 5.2.4c). However, the caveat to this method of determining co-localisation is that diffuse areas of staining, such as the cytosol in these images, have a tendency to demonstrate a large degree of co-localisation and thus high Pearson's coefficients. Therefore, these values may not be representative of the focal adhesion localisation of the WT and mutant SgK223.

Mutations to the ILK, PINCH1, α-parvin (IPP) complex prevent its role in inducing actin bundling and lead to decreased cell area [127]. We hypothesised that disrupting dimerisation and oligomerisation of SgK223 could disrupt its role in promoting cell elongation, leading to a change in cell area [1]. We calculated cell area using ImageJ and found that cells expressing SgK223 WT and mutant SgK223 had a similar area, approximately 250 μm^2. However, these data were calculated from the images we acquired using confocal Airyscan and many of the images were of regions of the cell, rather than the entire cell, therefore we could not use these for the quantification. To improve cell area quantification, we could capture more cells in the field of view (>10) expressing WT or mutant SgK223, creating a larger data set.

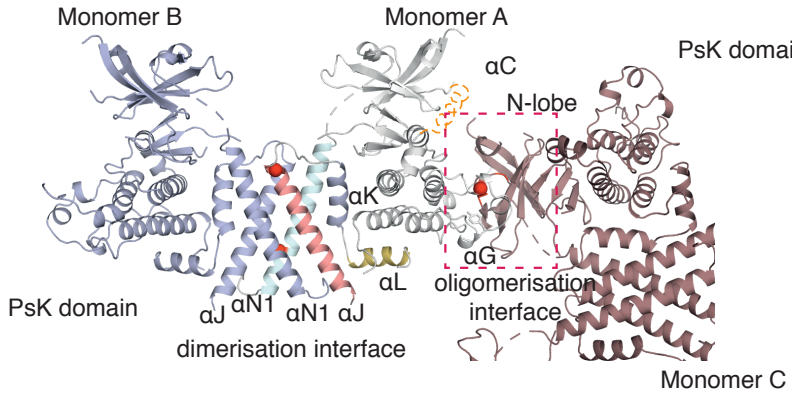
Additionally, focal adhesion size and number on a per cell basis could be quantified, using paxillin as a marker of focal adhesions. Capturing a large data set using TIRF microscopy would be ideal for this kind of quantitative comparison.

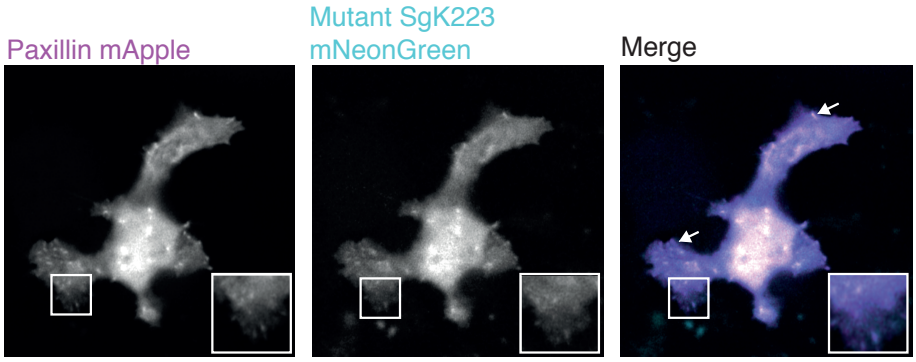
a) CONFOCAL AIRYSCAN



Hek293 cells

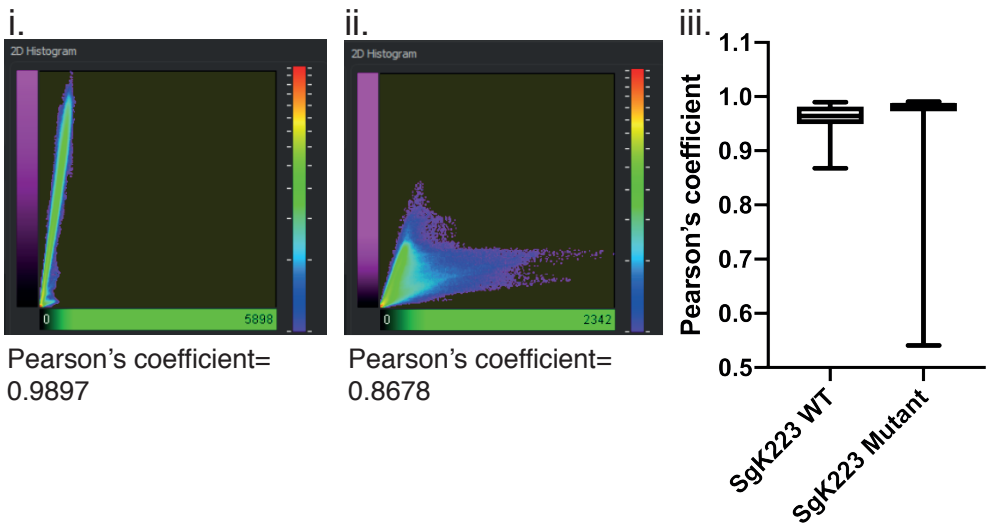
b)





Hek293 cells

c)



Pearson's coefficient= 0.9897

Pearson's coefficient= 0.8678

Figure 5.2.4. Confocal Airyscan of HEK293 cells expressing SgK223 WT or mutant SgK223.
a) Confocal Airyscan images of HEK293 cells transfected with mApple paxillin and SgK223 WT mNeonGreen. SgK223 mNeonGreen (cyan) co-localised with mApple paxillin (magenta) to focal adhesions. Focal adhesions are highlighted in the magnified panel on each image and the white arrows on the merge image. Single colour images are shown in greyscale for greater visual contrast.
b) Structure of SgK223 pseudokinase domain and flanking helices with red spheres indicating the residues that were mutated in the dimerisation/oligomerisation defective SgK223 mutant. Confocal Airyscan images of HEK293 cells transfected with mApple paxillin and mutant SgK223 mNeonGreen. Mutant SgK223 mNeonGreen (cyan) co-localised with mApple paxillin (magenta) to focal adhesions. Focal adhesions are highlighted in the magnified panel on each image and the white arrows on the merge image.
c) Panels i. and ii. plots of mNeonGreen fluorescence vs. mApple fluorescence. i. A strongly positive plot with a high Pearson's coefficient, ii. A less positive plot and lower Pearson's coefficient, iii. Pearson's coefficients of SgK223 WT compared to mutant SgK223, both demonstrating co-localisation between SgK223 and paxillin in HEK293 cells. N=1 independent experiment with 10 cells per condition. Error bars are +/- SEM.

5.3 Discussion

In this Chapter, we established the expression of SgK223 mNeonGreen in HEK293 cells and demonstrated its co-localisation with mApple paxillin at focal adhesions.

Further characterisation needs to be carried out on the cells expressing mutant SgK223 compared to SgK223 WT. We confirmed that TIRF microscopy can provide an increased signal-to-noise ratio compared to confocal microscopy for visualising the localisation of SgK223 and paxillin at focal adhesions.

Future experiments will require optimisation of SgK269 Halo expression in HEK293 cells. Once we have verified the localisation of SgK269, we will also express the mutant of SgK269 that is unable to dimerise and oligomerise to investigate the importance of SgK269 associations for its localisation in cells. We will characterise the cells expressing WT or mutant SgK223 and SgK269, using TIRF microscopy, comparing cell phenotypes such as cell area and focal adhesion size and number. Additionally, it was previously shown through TIRF time-lapse imaging that knocking down SgK269 in HT1080 fibrosarcoma cells altered the dynamics of focal adhesion assembly and disassembly [2]. We could carry out similar time-lapse experiments to determine if the mutant SgK223 and SgK269, can alter focal adhesion assembly and disassembly rates in HEK293 cells, similarly to knocking-down SgK269.

Ideally, we would like to carry out our investigations in a cell line that demonstrates obvious focal adhesions, such as the cancer cells lines AGS (gastric cancer) and HT1080 (fibrosarcoma). Additionally, developing a cell line that stably expresses paxillin, therefore only requiring the transient expression of SgK223 and SgK269, would reduce the number of plasmids transfected at once and hopefully increase the transfection efficiency.

Finally, the HEK293 cells used in these experiments still expressed endogenous SgK223, at unknown levels. Future experiments could involve knocking-out SgK223 and/or SgK269 using CRISPR/Cas-9 technology, to reduce the potential effects of endogenous protein, which could interfere with the comparison of mutant and SgK223 WT expressing cells.

These studies are critical for understanding the importance of the higher order associations of SgK223 and SgK269 for their role as scaffolds at focal adhesions.

6.0 General discussion and conclusions

This thesis began to biochemically characterise the SgK223 and SgK269 N-terminal domains. Using a number of biochemical techniques, we determined that these domains are monomeric and adopt a most likely elongated structure. However, further investigation of these domains is required, specifically determining if the cys/his motif binds ions and the relevance of this for SgK223 and SgK269 function. In order to fully understand the role of the N-terminal domains, the full-length SgK223 and SgK269 may need to be expressed and characterised. Additionally, identifying direct interacting partners of this domain will help to elucidate its function.

The investigations into SgK223 and SgK269 associations, from this study and previous published studies, demonstrated that both SgK223 and SgK269 undergo homo- and hetero- dimerisation and oligomerisation [6-8]. Dimerisation occurs through the helices that flank the pseudokinase domain N- and C-terminally, with the αN1 and αK forming the dimerisation interface [6-8]. Oligomerisation occurs through different interacting interfaces of SgK223 and SgK269, demonstrated *in vitro*. In SgK223, the N-lobe of the pseudokinase domain contributes to one interacting interface and the αG and A-loop of the pseudokinase domain contribute to another interacting interface [6]. These two interfaces in SgK223 are critical for SgK223 homo- oligomerisation and SgK223-SgK269 hetero-oligomerisation. Interestingly, in SgK269 only the αG and A-loop of the pseudokinase domain was demonstrated to contribute to SgK269 homo-oligomerisation and SgK269-SgK223 hetero-oligomerisation [6]. As SgK223 and SgK269 homo- and hetero-associations occur through different interfaces, this could be a mechanism that dictates the number of SgK223 and SgK269 molecules that can associate. Yet to be understood is the assembly of these scaffolds in the context of the entire protein including the N-terminal domain and PEST linker.

The importance of the association of SgK223 and SgK269 in cells has been studied, however not extensively. Thus far, only hetero-dimerisation of SgK223 and SgK269 has been demonstrated in cells, not hetero-oligomerisation [6]. Disrupting dimerisation of SgK223 and SgK269 prevents increased cell migration and cell elongation exhibited upon overexpression of the WT proteins and prevents the activation of the kinase Csk by SgK223 [1, 6, 8]. Recently, the dimerisation of PEAK3 has been demonstrated to be important for its interactions with CrkII. The interaction between PEAK3 and CrkII inhibits the CrkII induced membrane ruffling and leads to decreased actin filaments in cells [52]. We hypothesised that disrupting the dimerisation and oligomerisation of SgK223 and SgK269 would impair their localisation to focal adhesions. Interestingly, from our small data set obtained using confocal microscopy, we were not able to detect a difference in the localisation of SgK223 WT compared to mutant SgK223, however this requires further experiments. Further analysis of cells expressing dimerisation and oligomerisation defective mutants of SgK223 and SgK269 will help to elucidate the importance of their associations and their scaffolding role at focal adhesions.

We demonstrated that the PEAK family interact with the adaptor protein CrkII, using biophysical and structural techniques, however the biological relevance of these interactions is yet to be investigated. As demonstrated for PEAK3, dimerisation of SgK223 and SgK269 could be critical for facilitating their interactions with CrkII [52]. Carrying on from our investigations into SgK223 and SgK269 localisation, using the mutants that disrupt dimerisation and oligomerisation of SgK223 and SgK269, we could investigate if these mutants interrupt SgK223 and SgK269 interactions with CrkII, using microscopy.

Pseudokinases, although lacking catalytic function, play a critical role in numerous diseases, modulating signalling via their functions as allosteric regulators or as scaffolds. Thus, understanding how they promote signalling through their protein-protein interactions will offer clues on how they contribute to disease, and whether they can be classified as novel drug targets. The

associations of SgK223 and SgK269 may play a critical role in their ability to promote oncogenic signalling [1, 6, 136]. Through understanding the mechanisms of dimerisation and oligomerisation of the PEAK family of proteins, drugs could be developed to prevent their interactions and in doing so, the subsequent oncogenic signalling.

7.0 References

1. Liu, L., et al., *Homo- and Heterotypic Association Regulates Signaling by the SgK269/PEAK1 and SgK223 Pseudokinases.* J Biol Chem, 2016. **291**(41): p. 21571-21583.
2. Bristow, J.M., et al., *Dynamic Phosphorylation of Tyrosine 665 in Pseudopodium-enriched Atypical Kinase 1 (PEAK1) Is Essential for the Regulation of Cell Migration and Focal Adhesion Turnover.* Journal of Biological Chemistry, 2013. **288**(1): p. 123-131.
3. Wang, Y., et al., *Pseudopodium-enriched atypical kinase 1 regulates the cytoskeleton and cancer progression.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(24): p. 10920-10925.
4. Kong, R.R., et al., *Silencing NACK by siRNA inhibits tumorigenesis in non-small cell lung cancer via targeting Notch1 signaling pathway.* Oncology Reports, 2016. **35**(4): p. 2306-2314.
5. Senda, Y., N. Murata-Kamiya, and M. Hatakeyama, *C-terminal Src kinase-mediated EPIYA phosphorylation of Pragmin creates a feed-forward C-terminal Src kinase activation loop that promotes cell motility.* Cancer Science, 2016. **107**(7): p. 972-980.
6. Patel, O., et al., *Structure of SgK223 pseudokinase reveals novel mechanisms of homotypic and heterotypic association.* Nat Commun, 2017. **8**(1): p. 1157.
7. Ha, B.H. and T.J. Boggon, *The crystal structure of pseudokinase PEAK1 (Sugen Kinase 269) reveals an unusual catalytic cleft and a novel mode of kinase fold dimerization.* J Biol Chem, 2017.
8. Lecointre, C., et al., *Dimerization of the Pragmin Pseudo-Kinase Regulates Protein Tyrosine Phosphorylation.* Structure, 2018. **26**(4): p. 545-554.e4.
9. Hunter, T., *Signaling—2000 and Beyond.* Cell, 2000. **100**(1): p. 113-127.
10. Manning, G., et al., *The Protein Kinase Complement of the Human Genome.* Science, 2002. **298**(5600): p. 1912.
11. Hanks, S.K., A.M. Quinn, and T. Hunter, *The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.* Science, 1988. **241**(4861): p. 42-52.
12. Kornev, A.P., S.S. Taylor, and L.F. Ten Eyck, *A helix scaffold for the assembly of active protein kinases.* Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14377-82.
13. Kornev, A.P., et al., *Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17783-8.
14. Kung, J.E. and N. Jura, *Structural Basis for the Non-catalytic Functions of Protein Kinases.* Structure (London, England : 1993), 2016. **24**(1): p. 7-24.
15. Garnett, M.J., et al., *Wild-Type and Mutant B-RAF Activate C-RAF through Distinct Mechanisms Involving Heterodimerization.* Molecular Cell, 2005. **20**(6): p. 963-969.
16. Hu, J., et al., *Allosteric Activation of Functionally Asymmetric RAF Kinase Dimers.* Cell, 2013. **154**(5): p. 1036-1046.

17. Poulikakos, P.I., et al., *RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF.* Nature, 2010. **464**: p. 427.
18. Dey, M., et al., *Activation of Protein Kinase PKR Requires Dimerization-induced cis-Phosphorylation within the Activation Loop.* 2014. **289**(9): p. 5747-5757.
19. Langeberg, L.K. and J.D. Scott, *Signalling scaffolds and local organization of cellular behaviour.* Nature Reviews Molecular Cell Biology, 2015. **16**(4): p. 232-244.
20. Shaw, A.S., et al., *Kinases and pseudokinases: lessons from RAF.* Mol Cell Biol, 2014. **34**(9): p. 1538-46.
21. Reiterer, V., P.A. Eyers, and H. Farhan, *Day of the dead: pseudokinases and pseudophosphatases in physiology and disease.* Trends Cell Biol, 2014. **24**(9): p. 489-505.
22. Bandaranayake, R.M., et al., *Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F.* Nat Struct Mol Biol, 2012. **19**(8): p. 754-9.
23. Brennan, D.F., et al., *A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK.* Nature, 2011. **472**(7343): p. 366-9.
24. Cui, J., et al., *Structure of Fam20A reveals a pseudokinase featuring a unique disulfide pattern and inverted ATP-binding.* Elife, 2017. **6**.
25. Dai, D.L., et al., *Increased Expression of Integrin-Linked Kinase Is Correlated with Melanoma Progression and Poor Patient Survival.* Clinical Cancer Research, 2003. **9**(12): p. 4409.
26. Durzynska, I., et al., *STK40 Is a Pseudokinase that Binds the E3 Ubiquitin Ligase COP1.* Structure, 2017. **25**(2): p. 287-294.
27. Fukuda, K., et al., *The Pseudoactive Site of ILK Is Essential for Its Binding to α-Parvin and Localization to Focal Adhesions.* Molecular Cell, 2009. **36**(5): p. 819-830.
28. Min, X., et al., *Structural and Functional Characterization of the JH2 Pseudokinase Domain of JAK Family Tyrosine Kinase 2 (TYK2).* J Biol Chem, 2015. **290**(45): p. 27261-70.
29. Murphy, James M., et al., *The Pseudokinase MLKL Mediates Necroptosis via a Molecular Switch Mechanism.* Immunity, 2013. **39**(3): p. 443-453.
30. Murphy, J.M., et al., *Molecular Mechanism of CCAAT-Enhancer Binding Protein Recruitment by the TRIB1 Pseudokinase.* Structure, 2015. **23**(11): p. 2111-21.
31. Scheeff, E.D., et al., *Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site.* Structure, 2009. **17**(1): p. 128-138.
32. Toms, A.V., et al., *Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases.* Nature Structural &Amp; Molecular Biology, 2013. **20**: p. 1221.
33. Zeqiraj, E., et al., *Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation.* Science, 2009. **326**(5960): p. 1707-11.
34. Zeqiraj, E., et al., *ATP and MO25alpha regulate the conformational state of the STRADalpha pseudokinase and activation of the LKB1 tumour suppressor.* PLoS Biol, 2009. **7**(6): p. e1000126.

35. Littlefield, P., et al., *Structural analysis of the EGFR/HER3 heterodimer reveals the molecular basis for activating HER3 mutations.* Sci Signal, 2014. **7**(354): p. ra114.
36. Maubant, S., et al., *LRP5 regulates the expression of STK40, a new potential target in triple-negative breast cancers.* Oncotarget, 2018. **9**(32): p. 22586-22604.
37. Murphy, James M., et al., *Insights into the evolution of divergent nucleotide-binding mechanisms among pseudokinases revealed by crystal structures of human and mouse MLKL.* Biochemical Journal, 2014. **457**(3): p. 369.
38. Zhang, H., et al., *Structure and evolution of the Fam20 kinases.* Nature Communications, 2018. **9**(1): p. 1218.
39. Lupardus, P.J., et al., *Structure of the pseudokinase–kinase domains from protein kinase TYK2 reveals a mechanism for Janus kinase (JAK) autoinhibition.* Proceedings of the National Academy of Sciences, 2014. **111**(22): p. 8025.
40. Jura, N., et al., *Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21608-13.
41. Qiu, C., et al., *Mechanism of activation and inhibition of the HER4/ErbB4 kinase.* Structure, 2008. **16**(3): p. 460-7.
42. Xie, T., et al., *Pharmacological targeting of the pseudokinase Her3.* Nature chemical biology, 2014. **10**(12): p. 1006-1012.
43. Mishra, R., et al., *HER3 signaling and targeted therapy in cancer.* Oncology reviews, 2018. **12**(1): p. 355-355.
44. Murphy, J.M., et al., *A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties.* The Biochemical journal, 2014. **457**(2): p. 323-334.
45. Good, M.C., J.G. Zalatan, and W.A. Lim, *Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information.* Science, 2011. **332**(6030): p. 680-686.
46. Mukherjee, K., et al., *Evolution of CASK into a Mg²⁺-Sensitive Kinase.* Science Signaling, 2010. **3**(119): p. ra33.
47. Mukherjee, K., et al., *CASK Functions as a Mg²⁺-Independent Neurexin Kinase.* Cell, 2008. **133**(2): p. 328-339.
48. Atasoy, D., et al., *Deletion of CASK in mice is lethal and impairs synaptic function.* Proceedings of the National Academy of Sciences, 2007. **104**(7): p. 2525.
49. Legate, K.R., et al., *ILK, PINCH and parvin: the tIPP of integrin signalling.* Nat Rev Mol Cell Biol, 2006. **7**(1): p. 20-31.
50. Morrison, D.K., *KSR: a MAPK scaffold of the Ras pathway?* Journal of Cell Science, 2001. **114**(9): p. 1609.
51. Dhawan, N.S., A.P. Scopton, and A.C. Dar, *Small molecule stabilization of the KSR inactive state antagonizes oncogenic Ras signalling.* Nature, 2016. **537**(7618): p. 112-116.
52. Lopez, M.L., et al., *PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization.* Proceedings of the National Academy of Sciences, 2019: p. 201906360.
53. Agajanian, M., et al., *PEAK1 Acts as a Molecular Switch to Regulate Context-Dependent TGF beta Responses in Breast Cancer.* Plos One, 2015. **10**(8).

54. Croucher, D.R., et al., *Involvement of Lyn and the Atypical Kinase SgK269/PEAK1 in a Basal Breast Cancer Signaling Pathway.* Cancer Research, 2013. **73**(6): p. 1969-1980.
55. Kelber, J.A., et al., *KRas Induces a Src/PEAK1/ErbB2 Kinase Amplification Loop That Drives Metastatic Growth and Therapy Resistance in Pancreatic Cancer.* Cancer Research, 2012. **72**(10): p. 2554-2564.
56. Zheng, Y., et al., *Temporal regulation of EGF signalling networks by the scaffold protein Shc1.* Nature, 2013. **499**(7457): p. 166-71.
57. Leroy, C., et al., *Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates Src invasive activity in advanced colon carcinoma cells.* Cancer Research, 2009. **69**.
58. Safari, F., et al., *Mammalian Pragmin regulates Src family kinases via the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif that is exploited by bacterial effectors.* Proc Natl Acad Sci U S A, 2011. **108**(36): p. 14938-43.
59. Tactacan, C.M., et al., *The pseudokinase SgK223 promotes invasion of pancreatic ductal epithelial cells through JAK1/Stat3 signaling.* Mol Cancer, 2015. **14**: p. 139.
60. Tanaka, H., H. Katoh, and M. Negishi, *Pragmin, a novel effector of Rnd2 GTPase, stimulates RhoA activity.* J Biol Chem, 2006. **281**(15): p. 10355-64.
61. Leroy, C., et al., *Quantitative Phosphoproteomics Reveals a Cluster of Tyrosine Kinases That Mediates Src Invasive Activity in Advanced Colon Carcinoma Cells.* Cancer Research, 2009. **69**(6): p. 2279-2286.
62. Tactacan, C.M., et al., *The pseudokinase SgK223 promotes invasion of pancreatic ductal epithelial cells through JAK1/Stat3 signaling.* Molecular Cancer, 2015. **14**.
63. Weaver, K.L., et al., *NACK is an integral component of the Notch transcriptional activation complex and is critical for development and tumorigenesis.* Cancer Res, 2014. **74**(17): p. 4741-51.
64. Titz, B., et al., *The proximal signaling network of the BCR-ABL1 oncogene shows a modular organization.* Oncogene, 2010. **29**(44): p. 5895-5910.
65. Murphy, J.M., et al., *A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties.* Biochemical Journal, 2014. **457**: p. 323-334.
66. Andrade, M., et al., *Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network*. Vol. 6. 1993. 383-90.
67. Lobley, A., L. Whitmore, and B.A. Wallace, *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.* Bioinformatics, 2002. **18**(1): p. 211-212.
68. Fazio, V.J., T.S. Peat, and J. Newman, *A drunken search in crystallization space.* Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 10): p. 1303-11.
69. Bricogne, G., et al., *BUSTER version X.Y.Z.* Cambridge, United Kingdom: Global Phasing Ltd, 2016.
70. Kabsch, W., *XDS.* Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.
71. McCoy, A.J., *Solving structures of protein complexes by molecular replacement with Phaser.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 32-41.

72. Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.
73. Kantardjieff, K.A. and B. Rupp, *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.* Protein Sci, 2003. **12**(9): p. 1865-71.
74. 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.
75. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
76. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
77. Cassandri, M., et al., *Zinc-finger proteins in health and disease.* Cell Death Discov, 2017. **3**: p. 17071.
78. Cole, J.L., et al., *Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium.* Methods in cell biology, 2008. **84**: p. 143-179.
79. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.* Biophys J, 2000. **78**(3): p. 1606-19.
80. Smith, C., *Estimation of Sedimentation Coefficients and Frictional Ratios of Globular Proteins*, in *Biochemical Education*. 1988. p. 104-106.
81. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nature Protocols, 2006. **1**(6): p. 2876-2890.
82. Slabinski, L., et al., *XtalPred: a web server for prediction of protein crystallizability.* Bioinformatics, 2007. **23**(24): p. 3403-5.
83. Kostic, M., et al., *Solution Structure of the Hdm2 C2H2C4 RING, a Domain Critical for Ubiquitination of p53.* Journal of Molecular Biology, 2006. **363**(2): p. 433-450.
84. Lucet, I.S. and J.M. Murphy, *Characterization of Ligand Binding to Pseudokinases Using a Thermal Shift Assay*, in *Kinase Signaling Networks*, A.-C. Tan and P.H. Huang, Editors. 2017, Springer New York: New York, NY. p. 91-104.
85. Pilz, I., O. Glatter, and O. Kratky, *Small-angle X-ray scattering.* Methods Enzymol, 1979. **61**: p. 148-249.
86. Wieduwilt, M.J. and M.M. Moasser, *The epidermal growth factor receptor family: biology driving targeted therapeutics.* Cellular and molecular life sciences : CMLS, 2008. **65**(10): p. 1566-1584.
87. Bagley, C.J., et al., *The Structural and Functional Basis of Cytokine Receptor Activation: Lessons From the Common β Subunit of the Granulocyte-Macrophage Colony-Stimulating Factor, Interleukin-3 (IL-3), and IL-5 Receptors.* Blood, 1997. **89**(5): p. 1471.
88. Sriram, G. and R.B. Birge, *Emerging roles for crk in human cancer.* Genes Cancer, 2010. **1**(11): p. 1132-9.
89. Miller, C.T., et al., *Increased C-CRK proto-oncogene expression is associated with an aggressive phenotype in lung adenocarcinomas.* Oncogene, 2003. **22**(39): p. 7950-7957.

90. Rodrigues, S.P., et al., *CrkI and CrkII Function as Key Signaling Integrators for Migration and Invasion of Cancer Cells.* Molecular Cancer Research, 2005. **3**(4): p. 183.
91. Wang, L., et al., *Signaling adaptor protein Crk is indispensable for malignant feature of glioblastoma cell line KMG4.* Biochemical and Biophysical Research Communications, 2007. **362**(4): p. 976-981.
92. Flynn, D.C., *Adaptor proteins.* Oncogene, 2001. **20**(44): p. 6270-6272.
93. Takino, T., et al., *Tyrosine phosphorylation of the CrkII adaptor protein modulates cell migration.* J Cell Sci, 2003. **116**(Pt 15): p. 3145-55.
94. Pawson, T. and J. Schlessingert, *SH2 and SH3 domains.* Current Biology, 1993. **3**(7): p. 434-442.
95. Feller, S.M., *Crk family adaptors–signalling complex formation and biological roles.* Oncogene, 2001. **20**: p. 6348.
96. Mayer, B.J., *SH3 domains: complexity in moderation.* Journal of Cell Science, 2001. **114**(7): p. 1253.
97. Teyra, J., et al., *Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities.* Structure, 2017. **25**(10): p. 1598-1610 e3.
98. Saksela, K. and P. Permi, *SH3 domain ligand binding: What's the consensus and where's the specificity?* FEBS Lett, 2012. **586**(17): p. 2609-14.
99. Matsuda, M., et al., *Two species of human CRK cDNA encode proteins with distinct biological activities.* Molecular and Cellular Biology, 1992. **12**(8): p. 3482.
100. Akakura, S., et al., *C-terminal SH3 domain of CrkII regulates the assembly and function of the DOCK180/ELMO Rac-GEF.* J Cell Physiol, 2005. **204**(1): p. 344-51.
101. Sun, H., et al., *Host adaptor proteins Gab1 and CrkII promote InlB-dependent entry of Listeria monocytogenes.* Cell Microbiol, 2005. **7**(3): p. 443-57.
102. Muralidharan, V., et al., *Solution structure and folding characteristics of the C-terminal SH3 domain of c-Crk-II.* Biochemistry, 2006. **45**(29): p. 8874-84.
103. Kobashigawa, Y., et al., *Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK.* Nat Struct Mol Biol, 2007. **14**(6): p. 503-10.
104. Petit, V., et al., *Phosphorylation of Tyrosine Residues 31 and 118 on Paxillin Regulates Cell Migration through an Association with Crk in Nbt-II Cells.* The Journal of Cell Biology, 2000. **148**(5): p. 957.
105. Hasegawa, H., et al., *DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane.* Molecular and Cellular Biology, 1996. **16**(4): p. 1770.
106. Kiyokawa, E., et al., *Evidence That DOCK180 Up-regulates Signals from the CrkII-p130Cas Complex.* 1998. **273**(38): p. 24479-24484.
107. Nakashima, N., et al., *The functional role of CrkII in actin cytoskeleton organization and mitogenesis.* J Biol Chem, 1999. **274**(5): p. 3001-8.
108. Abassi, Y.A. and K. Vuori, *Tyrosine 221 in Crk regulates adhesion-dependent membrane localization of Crk and Rac and activation of Rac signaling.* The EMBO Journal, 2002. **21**(17): p. 4571.

109. Rosen, M.K., et al., *Direct demonstration of an intramolecular SH2-phosphotyrosine interaction in the Crk protein.* Nature, 1995. **374**(6521): p. 477-9.
110. Bhatt, V.S., et al., *Binding Mechanism of the N-Terminal SH3 Domain of CrkII and Proline-Rich Motifs in cAbl.* Biophys J, 2016. **110**(12): p. 2630-2641.
111. Markey, F., *Principles of Surface Plasmon Resonance*, in *Real-Time Analysis of Biomolecular Interactions: Applications of BIACORE*, K. Nagata and H. Handa, Editors. 2000, Springer Japan: Tokyo. p. 13-22.
112. van der Merwe, P., *Surface plasmon resonance.* Protein-Ligand Interactions: Hydrodynamics and Calorimetry, 2001: p. 137-170.
113. Nikolovska-Coleska, Z., *Studying Protein-Protein Interactions Using Surface Plasmon Resonance*, in *Protein-Protein Interactions: Methods and Applications*, C.L. Meyerkord and H. Fu, Editors. 2015, Springer New York: New York, NY. p. 109-138.
114. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
115. Wu, X., et al., *Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk.* Structure, 1995. **3**(2): p. 215-226.
116. Musacchio, A., et al., *Crystal structure of a Src-homology 3 (SH3) domain.* Nature, 1992. **359**(6398): p. 851-5.
117. Noble, M.E., et al., *Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin.* EMBO J, 1993. **12**(7): p. 2617-24.
118. Mayer, B.J. and M.J. Eck, *SH3 domains. Minding your p's and q's.* Curr Biol, 1995. **5**(4): p. 364-7.
119. Shen, Q., et al., *The Molecular Mechanisms Underlying the Hijack of Host Proteins by the 1918 Spanish Influenza Virus.* ACS Chem Biol, 2017. **12**(5): p. 1199-1203.
120. Xue, Y., et al., *Role of Electrostatic Interactions in Binding of Peptides and Intrinsically Disordered Proteins to Their Folded Targets. 1. NMR and MD Characterization of the Complex between the c-Crk N-SH3 Domain and the Peptide Sos.* Biochemistry, 2014. **53**(41): p. 6473-6495.
121. Girardin, S.E. and M. Yaniv, *A direct interaction between JNK1 and CrkII is critical for Rac1-induced JNK activation.* The EMBO Journal, 2001. **20**(13): p. 3437.
122. Knudsen, B.S., S.M. Feller, and H. Hanafusa, *Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology 3 domain of Crk.* 1994. **269**(52): p. 32781-32787.
123. Knudsen, B.S., et al., *Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins.* The EMBO journal, 1995. **14**(10): p. 2191-2198.
124. Schlessinger, J., *SH2/SH3 signaling proteins.* Current Opinion in Genetics & Development, 1994. **4**(1): p. 25-30.
125. Wehrle-Haller, B., *Structure and function of focal adhesions.* Curr Opin Cell Biol, 2012. **24**(1): p. 116-24.

126. Wu, C., *Focal adhesion: a focal point in current cell biology and molecular medicine.* Cell Adh Migr, 2007. **1**(1): p. 13-8.
127. Vaynberg, J., et al., *Non-catalytic signaling by pseudokinase ILK for regulating cell adhesion.* Nat Commun, 2018. **9**(1): p. 4465.
128. Shaner, N.C., et al., *A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum.* Nat Methods, 2013. **10**(5): p. 407-9.
129. Schlienger, S., R.A.M. Ramirez, and A. Claing, *ARF1 regulates adhesion of MDA-MB-231 invasive breast cancer cells through formation of focal adhesions.* Cellular Signalling, 2015. **27**(3): p. 403-415.
130. Chiu, C.-L., et al., *Nanoimaging of Focal Adhesion Dynamics in 3D.* PLOS ONE, 2014. **9**(6): p. e99896.
131. Wang, T., et al., *Systematic Screening of Commonly Used Commercial Transfection Reagents towards Efficient Transfection of Single-Stranded Oligonucleotides.* Molecules (Basel, Switzerland), 2018. **23**(10): p. 2564.
132. Stricker, J., et al., *Myosin II-mediated focal adhesion maturation is tension insensitive.* PloS one, 2013. **8**(7): p. e70652-e70652.
133. Fukumoto, M., et al., *α-Actinin-4 Enhances Colorectal Cancer Cell Invasion by Suppressing Focal Adhesion Maturation.* PLOS ONE, 2015. **10**(4): p. e0120616.
134. Johnson, D.S., J.K. Jaiswal, and S. Simon, *Total internal reflection fluorescence (TIRF) microscopy illuminator for improved imaging of cell surface events.* Curr Protoc Cytom, 2012. **Chapter 12**: p. Unit 12 29.
135. Arena, T.A., P.D. Harms, and A.W. Wong, *High Throughput Transfection of HEK293 Cells for Transient Protein Production.* Methods Mol Biol, 2018. **1850**: p. 179-187.
136. Ademuyiwa, A.O., et al., *Determinants of morbidity and mortality following emergency abdominal surgery in children in low-income and middle-income countries.* Bmj Global Health, 2016. **1**(4).

8.0 Appendices

Table 8.1. Sequences of the genes cloned in this study.

Gene	Vector	Sequence	Restriction sites
SgK223-C-term WT 932-1406	pCOLD	GGATCCGGCTCAACCCAACTGCAACTGCACGGTCT GCTGTCGAACATCTCGTCAAAAGAAGGCACCTATG CGAAACTGGGCGGTCTGTACACGCAGAGCCTGGCA CGCCTGGTTGCTAAATGCGAAGACCTGTTTATGGG CGGTCAGAAAAAAGAACTGCACTTCAACGAAAACA ACTGGAGCCTGTTCAAACTGACCTGTAATAAACCG TGCTGTGATTCTGGCGACGCAATTTATTACTGCGC TACCTGTAGTGAAGATCCGGGTTCCACGTATGCGG TCAAAATCTGCAAAGCCCCGGAACCGAAAACCGTG AGTTACTGTAGCCCGTCTGTGCCGGTTCACTTTAA CATTCAGCAAGACTGCGGCCATTTCGTCGCCAGCG TGCCGAGCTCTATGCTGAGTTCCCCGGATGCGCCG AAAGACCCGGTGCCGGCACTGCCGACCCACCCGCC GGCACAGGAACAAGATTGTGTGGTTGTCATCACCC GCGAAGTGCCGCACCAGACGGCATCTGATTTTGTT CGTGACTCAGCGGCCTCGCATCAAGCGGAACCGGA AGCCTATGAACGTCGCGTCTGCTTCCTGCTGCTGC AACTGTGTAACGGCCTGGAACATCTGAAAGAACAC GGTATTATCCATCGTGATCTGTGCCTGGAAAATCT GCTGCTGGTGCATTGTACCCTGCAAGCCGGCCCGG GTCCGGCACCGGCTCCGGCACCGGCACCGGCAGCT GCGGCCCCGCCGTGCTCATCGGCAGCTCCGCCGGC TGGCGGTACGCTGTCACCGGCGGCGGGTCCGGCAT CGCCGGAAGGTCCGCGCGAAAAACAGCTGCCGCGT CTGATTATCAGTAACTTTCTGAAAGCCAAACAAAA ACCGGGCGGCACCCCGAACCTGCAACAGAAAAAAT CTCAGGCACGCCTGGCTCCGGAAATTGTGAGTGCG TCCCAGTATCGTAAATTTGATGAATTTCAAACGGG CATTCTGATCTACGAACTGCTGCACCAGCCGAACC CGTTCGAAGTTCGCGCACAGCTGCGCGAACGTGAT TATCGTCAAGAAGACCTGCCGCCGCTGCCGGCACT GAGCCTGTACTCTCCGGGTCTGCAACAACTGGCAC ATCTGCTGCTGGAAGCTGATCCGATTAAACGCATT CGTATCGGCGAAGCGAAACGCGTTCTGCAATGCCT GCTGTGGGGTCCGCGTCGCGAACTGGTCCAGCAAC CGGGCACCAGCGAAGAAGCGCTGTGTGGTACGCTG CATAATTGGATCGACATGAAACGTGCCCTGATGAT GATGAAATTCGCGGAAAAAGCCGTTGATCGTCGCC GTGGTGTCGAACTGGAAGACTGGCTGTGCTGTCAG TACCTGGCGTCGGCGGAACCGGGTGCCCTGCTGCA ATCCCTGAAACTGCTGCAACTGCTGTGAAAGCTT	BamHI, HindIII
SgK269-C-term WT 1267-1746	pCOLD	GGATCCGGTATTCAAAAGCCGCAACGCCAAGCGCT GTATCGTGGTCTGGAGAACCGTGAGGAAGTGGTGG GCAAAATTCGTAGCCTGCACACCGACGCGCTGAAG AAACTGGCGGTTAAATGCGAGGACCTGTTCATGGC GGGTCAGAAGGATCAACTGCGTTTCGGCGTGGACA GCTGGAGCGATTTTCGTCTGACCAGCGACAAACCG TGCTGCGAGGCGGGTGATGCGGTTTACTATACCGC GAGCTACGCGAAAGACCCGCTGAACAACTATGCGG TGAAAATCTGCAAGAGCAAAGCGAAGGAAAGCCAG CAATACTATCACAGCCTGGCGGTTCGTCAGAGCCT GGCGGTTCACTTCAACATTCAGCAAGATTGCGGCC ACTTTCTGGCGGAAGTTCCGAACCGTCTGCTGCCG TGGGAGGACCCGGATGATCCGGAGAAAGATGAAGA CGATATGGAGGAAACCGAGGAAGACGCGAAGGGTG	BamHI, HindIII

		AAACCGATGGCAAAAACCCGAAGCCGTGCAGCGAA GCGGCGAGCAGCCAGAAGGAGAACCAAGGTGTTAT GAGCAAGAAACAGCGTAGCCACGTG**GTT**GTGATCA CCCGTGAAGTGCCGTGCCTGACCGTTGCGGACTTT GTGCGTGATAGCCTGGCGCAACACGGTAAAAGCCC GGACCTGTACGAACGTCAGGTGTGCCTGCTGCTGC TGCAACTGTGCAGCGGCCTGGAGCACCTGAAGCCG TACCACGTTACCCACTGCGATCTGCGTCTGGAAAA CCTGCTGCTGGTTCACTATCAACCGGGTGGCACCG CGCAAGGTTTTGGTCCGGCGGAACCGAGCCCGACC AGCAGCTATCCGACCCGTCTGATCGTTAGCAACTT TAGCCAGGCGAAACAAAAGAGCCACCTGGTGGACC CGGAAATTCTGCGTGATCAGAGCCGTCTGGCGCCG GAGATCATTACCGCGACCCAGTACAAGAAATGCGA CGAGTTCCAAACCGGTATCCTGATTTATGAGATGC TGCACCTGCCGAACCCG**TTT**GACGAGAACCCGGAA CTGAAAGAGCGTGAATACACCCGTGCGGATCTGCC GCGTATCCCGTTCCGTAGCCCGTACAGCCGTGGTC TGCAACAACTGGCGAGCTGCCTGCTGAACCCGAAC CCGAGCGAGCGTATCCTGATTAGCGACGCGAAAGG TATCCTGCAATGCCTGCTGTGGGGTCCGCGTGAGG ACCTGTTCCAAACCTTTACCGCGTGCCCGAGCCTG GTTCAGCGTAACACCCTGCTGCAAAACTGGCTGGA CATCAAGCGTACCCTGCTGATGATTAAA**TTTGCG**G AGAAGAGCCTGGACCGTGAAGGTGGCATTAGCCTG GAGGAT**TGG**CTGTGCGCGCAGTACCTGGCGTTTGC GACCACCGATAGCCTGAGCTGCATTGTGAAGATTC TGCAACACCGTTGA**AAGCTT**	
CrkII N-SH3, 134-191	pCOLD	**GGATCC**GCGGAGTACGTGCGTGCGCTGTTCGACTT TAACGGTAACGACGAGGAAGATCTGCCGTTCAAGA AAGGCGACATCCTGCGTATTCGTGATAAGCCGGAG GAACAGTGGTGGAACGCGGAGGATAGCGAAGGCAA GCGTGGCATGATCCCGGTGCCGTACGTTGAAAAAT ATCGTTAA**AAGCTT**	BamHI, HindIII
CrkII C-SH3, 218-299	pCOLD	**GGATCC**GGTCCGTACGCGCAGCCGAGCGTGAACAC CCCGCTGCCGAACCTGCAGAACGGTCCGATCTACG CGCGTGTGATTCAAAAGCGTGTTCCGAACGCGTAT GACAAAACCGCGCTGGCGCTGGAAGTGGGTGAACT GGTGAAGGTTACCAAAATCAACGTTAGCGGCCAAT GGGAGGGTGGCTGCAACGGCAAGCGTGGCCACTTC CCGTTTACCCACGTTCGTCTGCTGGATCAGCAAAA CCCGTAA**AAGCTT**	BamHI, HindIII
SgK223-WT	pMIH	**ACCGG**TATGCACCAGACCCTGTGCCTGAACCCCGA GAGCCTGAAGATGAGCGCCTGTAGCGATTTTGTGG AACATATTTGGAAGCCTGGGAGCTGTAAGAACTGC TTTTGTCTGAGGTCTGACCACCAGCTGGTGGCAGG ACCACCTCAGCCTCGCGCCGGCAGCCTGCCACCCC CTCCAAGGCTGCCACCTCGCCCAGAGAACTGCCGG CTGGAGGACGAGGGCGTGAATAGCTCCCCTTACTC CAAGCCAACCATCGCCGTGAAGCCCACAATGATGT CTAGCGAGGCCTCTGACGTGTGGACAGAGGCCAAC CTGTCCGCCGAGGTGTCTCAGGTCATCTGGCGGCG GGCCCCCGGCAAGCTGCCACTGCCTAAGCAGGAGG ACGCCCCTGTGGTGTACCTGGGCTCCTTTCGCGGA GTGCAGAAGCCTGCCGGCCCATCTACCAGCCCTGA TGGCAATAGCCGGTGTCCACCCGCCTATACAATGG TGGGCCTGCACAACCTGGAGCCAAGGGGCGAGCGC AATATCGCCTTTCACCCCGTGTCCTTCCCTGAGGA	AgeI, XhoI (internal BamHI, HindIII)

		GAAGGCCGTGCACAAGGAGAAGCCCTCTTTCCCTT ACCAGGACCGGCCCAGCACCCAGGAGTCCTTTAGA CAGAAGCTGGCAGCCTTCGCCGGCACCACATCCGG ATGCCACCAGGGCCCAGGCCCACTGAGGGAGTCCC TGCCATCTGAGGACGATAGCGACCAGAGATGCTCC CCATCTGGCGATTCTGAGGGAGGAGAGTATTGTAG CATCCTGGACTGCTGTCCAGGCAGCCCCGTGGCCA AGGCCGCCAGCCAGACCGCCGGCAGCCGGGGCCGG CACGGCGGACGCGATTGCAGCCCTACATGTTGGGA GCAGGGCAAGTGTTCCGGACCAGCCGAGCAGGAGA AGAGGGGACCTTCCTTTCCAAAGGAGTGCTGTTCT CAGGGACCAACCGCACACCCATCCTGCCTGGGACC TAAGAAGCTGAGCCTGACATCCGAGGCCGCCATCT CCTCTGATGGCCTGAGCTGTGGCAGCGGCTCCGGC TCTGGGGCCAGCAGCCCCTTCGTGCCTCACCTGGA GTCTGACTACTGCTCCCTGATGAAGGAGCCTGCCC CAGAGAAGCAGCAGGACCCCGGCTGCCCCGGCGTG ACCCCCTCCCGGTGTCTGGGCCTGACAGGCGAGCC ACAGCCTCCAGCACACCCTAGAGAGGCAACCCAGC CAGAGCCTATCTATGCCGAGAGCACAAAGAGGAAG AAGGCCGCCCCAGTGCCCAGCAAGTCCCAGGCCAA GATCGAGCACGCTGCCGCCGCCCAGGGACAGGGCC AGGTGTGCACCGGCAACGCCTGGGCCCAGAAGGCA GCATCTGGATGGGGAAGGGACAGCCCTGATCCAAC ACCACAGGTGTCCGCCACCATCACAGTGATGGCCG CCCACCCAGAGGAGGACCACAGAACCATCTACCTG TCTAGCCCAGATAGCGCCGTGGGAGTGCAGTGGCC TAGGGGCCCAGTGTCTCAGAATAGCGAAGTGGGAG AGGAGGAGACAAGCGCCGGCCAGGGCCTGTCCTCT CGCGAGAGCCACGCACACTCTGCCAGCGAGTCCAA GCCTAAGGAGCGGCCAGCCATCCCCCCTAAGCTGA GCAAGAGCTCCCCCGTGGGCTCCCCCGTGTCTCCT AGCGCCGGCGGACCACCCGTGTCCCCTCTGGCCGA CCTGTCTGATGGCTCTAGCGGAGGCTCCTCTATCG GACCTCAGCCTCCATCCCAGGGACCAGCCGACCCT GCCCCATCTTGCAGGACAAACGGCGTGGCCATCTC CGATCCATCTCGCTGTCCACAGCCTGCCGCCAGCT CCGCCAGCGAGCAGAGACGGCCCAGGTTTCAGGCA GGCACCTGGTCCCGGCAGTGTAGAATCGAGGAGGA GGAGGAGGTGGAGCAGGAGCTGCTGTCCCACTCTT GGGGCCGCGAGACAAAGAACGGCCCAACAGACCAC AGCAATTCCACCACATGGCACCGGCTGCACCCAAC CGATGGCTCTAGCGGCCAGAACTCCAAAGTGGGCA CAGGCATGTCTAAGAGCGCCTCCTTCGCCTTTGAG TTCCCTAAGGACAGAAGCGGCATCGAGACATTCAG CCCACCTCCACCACCTCCAAAGAGCAGGCACCTGC TGAAGATGAATAAGTCCTCTAGCGATCTGGAGAAG GTGAGCCAGGGCTCCGCCGAGTCTCTGAGCCCCTC CTTTAGAGGCGTGCACGTGTCCTTCACCACAGGCT CTACCGACAGCCTGGCCTCTGATAGCAGGACATGC AGCGACGGAGGACCCTCCTCTGAGCTGGCACACTC CCCTACCAATTCTGGCAAGAAGCTGTTTGCCCCAG TGCCCTTCCCTTCCGGCTCTACAGAGGACGTGAGC CCATCCGGACCACAGCAGCCACCTCCACTGCCTCA GAAGAAGATCGTGAGCCGGGCCGCCAGCTCCCCAG ATGGCTTCTTTTGGACCCAGGGCTCTCCAAAGCCC GGCACAGCAAGCCCCAAGCTGAACCTGTCTCACAG CGAGACAAATGTGCACGACGAGAGCCACTTCTCCT ATTCTCTGAGCCCAGGCAACAGACACCACCCCGTG TTTTCTAGCTCCGATCCCCTGGAGAAGGCCTTCAA GGGCTCCGGACACTGGCTGCCTGCCGCCGGCCTGG	

		CCGGCAATAGGGGCGGCTGCGGCAGCCCCGGCCTC CAGTGTAAAGGGGCTCCTAGCGCCTCCTCATCTCA GCTGTCTGTCTCTAGTCAGGCCTCTACC**GGATCC**G GCTCAACCCAACTGCAACTGCACGGTCTGCTGTCG AACATCTCGTCAAAAGAAGGCACCTATGCGAAACT GGGCGGTCTGTACACGCAGAGCCTGGCACGCCTGG TTGCTAAATGCGAAGACCTGTTTATGGGCGGTCAG AAAAAAGAACTGCACTTCAACGAAAACAACTGGAG CCTGTTCAAACTGACCTGTAATAAACCGTGCTGTG ATTCTGGCGACGCAATTTATTACTGCGCTACCTGT AGTGAAGATCCGGGTTCCACGTATGCGGTCAAAAT CTGCAAAGCCCCGGAACCGAAAACCGTGAGTTACT GTAGCCCGTCTGTGCCGGTTCACTTTAACATTCAG CAAGACTGCGGCCATTTCGTCGCCAGCGTGCCGAG CTCTATGCTGAGTTCCCCGGATGCGCCGAAAGACC CGGTGCCGGCACTGCCGACCCACCCGCCGGCACAG GAACAAGATTGTGTGGTTGTCATCACCCGCGAAGT GCCGCACCAGACGGCATCTGATTTTGTTCGTGACT CAGCGGCCTCGCATCAAGCGGAACCGGAAGCCTAT GAACGTCGCGTCTGCTTCCTGCTGCTGCAACTGTG TAACGGCCTGGAACATCTGAAAGAACACGGTATTA TCCATCGTGATCTGTGCCTGGAAAATCTGCTGCTG GTGCATTGTACCCTGCAAGCCGGCCCGGGTCCGGC ACCGGCTCCGGCACCGGCACCGGCAGCTGCGGCCC CGCCGTGCTCATCGGCAGCTCCGCCGGCTGGCGGT ACGCTGTCACCGGCGGCGGGTCCGGCATCGCCGGA AGGTCCGCGCGAAAAACAGCTGCCGCGTCTGATTA TCAGTAACTTTCTGAAAGCCAAACAAAAACCGGGC GGCACCCCGAACCTGCAACAGAAAAAATCTCAGGC ACGCCTGGCTCCGGAAATTGTGAGTGCGTCCCAGT ATCGTAAATTTGATGAATTTCAAACGGGCATTCTG ATCTACGAACTGCTGCACCAGCCGAACCCGTTCGA AGTTCGCGCACAGCTGCGCGAACGTGATTATCGTC AAGAAGACCTGCCGCCGCTGCCGGCACTGAGCCTG TACTCTCCGGGTCTGCAACAACTGGCACATCTGCT GCTGGAAGCTGATCCGATTAAACGCATTCGTATCG GCGAAGCGAAACGCGTTCTGCAATGCCTGCTGTGG GGTCCGCGTCGCGAACTGGTCCAGCAACCGGGCAC CAGCGAAGAAGCGCTGTGTGGTACGCTGCATAATT GGATCGACATGAAACGTGCCCTGATGATGATGAAA TTCGCGGAAAAAGCCGTTGATCGTCGCCGTGGTGT CGAACTGGAAGACTGGCTGTGCTGTCAGTACCTGG CGTCGGCGGAACCGGGTGCCCTGCTGCAATCCCTG AAACTGCTGCAACTGCTGTGA**GTCGACCTCGAG**	
SgK269-WT	pMIH	**ACCGGTGAATTC**TGTCCGCCTGTAATACTTTCACC GAGCACGTCTGGAAGCCTGGAGAGTGTAAGAACTG TTTTAAGCCTAAGAGCCTGCATCAGCTGCCCCCTG ATCCTGAGAAGGCCCCAATCACACACGGCAATGTG AAGACCAATGCCAACCACAGCAACAATCACAGAAT CAGGAATACCGGCAACTTCAGGCCACCCGTGGCCA AGAAGCCAACAATCGCCGTGAAGCCCACCATGATC GTGGCAGACGGACAGAGCATCTGCGGAGAGCTGTC CATCCAGGAGCACTGTGAGAACAAGCCAGTGATCA TCGGATGGAATAGGAACAGGGCCGCCCTGAGCCAG AAGCCCCTGAACAATAACAATGAGGACGATGAGGG CATCTCCCACGTGCCCAAGCCTTACGGCAACAATG ATTCTGCCAAGAAGATGAGCGACAACAATAACGGC CTGACCGAGGTGCTGAAGGAGATCGCCGGCCTGGA TACAGCCCCCCAGATCAGGGGCAATGAGACAAACA	AgeI, XhoI (internal BamHI, HindIII)

		GCCGGGAGACATTCCTGGGCAGAATCAACGACTGC TACAAGCGGAGCCTGGAGAGAAAGCTGCCTCCAAG CTGTATGATCGGCGGCATCAAGGAGACACAGGGCA AGCACGTGATCCTGTCTGGCAGCACAGAAGTGATC TCCAATGAGGGCGGCCGCTTCTGCTATCCCGAGTT TAGCTCCGGCGAGGAGTCCGAGGAGGATGTGCTGT TTTCTAACATGGAGGAGGAGCACGAGAGCTGGGAC GAGTCCGATGAGGAGCTGCTGGCTATGGAGATCAG AATGAGGGGCCAGCCAAGGTTCGCCAATTTTCGCG CCAACACCCTGAGCCCCGTGAGGTTCTTTGTGGAC AAGAAGTGGAATACAATCCCCCTGAGGAACAAGAG CCTCCAGCGCATCTGCGCCGTGGATTACGACGATT CTTATGACGAGATCCTGAATGGCTACGAGGAGAAC AGCGTGGTGTCCTATGGCCAGGGCTCTATCCAGAG CATGGTGTCTAGCGATTCCACATCTCCCGACTCCT CTCTGACCGAGGAGAGCCGCTCCGAGACAGCCAGC TCCCTGAGCCAGAAGATTTGCAATGGAGGCCTGAG CCCAGGCAACCCCGGCGACTCCAAGGATATGAAGG AGATCGAGCCCAACTACGAGTCCCCTTCTAGCAAT AACCAGGACAAGGATTCCTCTCAGGCCTCTAAGAG CTCCATCAAGGTGCCCGAGACACACAAGGCCGTGC TGGCCCTGAGGCTGGAGGAGAAGGATGGCAAGATC GCCGTGCAGACCGAGAAGGAGGAGTCCAAGGCCTC TACAGACGTGGCAGGACAGGCAGTGACAATCAATC TGGTGCCAACCGAGGAGCAGGCAAAGCCATACAGG GTGGTGAACCTGGAGCAGCCCCTGTGCAAGCCTTA TACCGTGGTGGACGTGAGCGCCGCTATGGCCTCTG AGCACCTGGAGGGCCCAGTGAATAGCCCCAAGACA AAGTCTAGCTCCTCTACCCCTAACTCCCCAGTGAC AAGCTCCTCTCTGACCCCTGGCCAGATCAGCGCCC ACTTCCAGAAGAGCTCCGCCATCAGGTATCAGGAA GTGTGGACCTCTAGCACATCTCCACGCCAGAAGAT CCCCAAGGTGGAGCTGATCACCAGCGGCACAGGCC CAAATGTGCCCCCTCGGAAGAACTGTCACAAGTCT GCCCCCACCAGCCCTACCGCCACAAATATCTCCTC TAAGACAATCCCTGTGAAGTCCCCAAATCTGTCTG AGATCAAGTTCAACAGCTACAATAACGCCGGCATG CCACCCTTTCCTATCATCATCCACGATGAGCCAAC CTATGCCAGAAGCTCCAAGAATGCCATCAAGGTGC CTATCGTGATCAATCCAAACGCCTACGACAACCTG GCCATCTATAAGTCCTTTCTGGGCACAAGCGGCGA GCTGTCCGTGAAGGAGAAGACCACAAGCGTGATCT CCCACACATACGAGGAGATCGAGACAGAGTCCAAG GTGCCAGATAATACCACATCTAAGACCACAGACTG CCTCCAGACCAAGGGCTTCTCTAACAGCACAGAGC ACAAGAGGGGCAGCGTGGCCCAGAAGGTGCAGGAG TTTAATAACTGTCTGAACCGGGGCCAGTCTAGCCC TCAGAGATCCTATTCCTCTAGCCACTCCTCTCCAG CAAAGATCCAGAGGGCAACCCAGGAGCCCGTGGCC AAGATCGAGGGCACACAGGAGAGCCAGATGGTGGG CAGCTCCTCTACAAGAGAGAAGGCCTCTACCGTGC TGAGCCAGATCGTGGCCTCTATCCAGCCTCCACAG AGCCCACCTGAGACACCACAGTCCGGACCTAAGGC CTGTTCTGTGGAGGAGCTGTACGCCATCCCACCCG ACGCCGATGTGGCCAAGTCCACACCCAAGTCTACC CCAGTGCGGCCCAAGTCCCTGTTCACCTCTCAGCC TAGCGGAGAGGCAGAGGCCCCTCAGACCACAGATA GCCCAACCACAAAGGTGCAGAAGGACCCATCCATC AAGCCCGTGACTCCATCCCCCTCTAAGCTGGTGAC ATCCCCCCAGTCTGAGCCTCCAGCCCCTTTCCCCC CTCCAAGAAGCACCAGCTCCCCCTATCACGCCGGC	

		AATCTGCTCCAGCGGCACTTTACCAACTGGACAAA GCCTACCAGCCCAACAAGATCCACCGAGGCAGAGT CCGTGCTGCACAGCGAGGGCTCCCGGAGAGCCGCC GACGCCAAGCCTAAGAGGTGGATCTCTTTTAAGAG CTTCTTTCGGCGGAGAAAGACCGATGAGGAGGACG ATAAGGAGAAGGAGCGCGAGAAGGGCAAGCTGGTG GGCCTGGACGGCACCGTGATCCACATGCTGCCACC TCCACCCGTGCAGAGGCACCACTGGTTCACAGAGG CCAAGGGCGAGTCTAGCGAGAAGCCTGCCATCGTG TTTATGTACAGATGCGATCCAGCCCAGGGCCAGCT GTCTGTGGATCAGAGCAAGGCAAGGACCGACCAGG CCGCCGTGATGGAGAAGGGCAGAGCCGAGAACGCC CTGCTCCAGGACAGCGAGAAGAAGCGGAGCCACAG CAGCCCTTCTCAGATCCCAAAGAAGATCCTGTCTC ACATGACACACGAGGTGACCGAGGATTTCAGCCCC CGCGACCCTAGGACCGTGGTGGGCAAGCAGGATGG CAGGGGCTGCACAAGCGTGACCACAGCCCTGTCCC TGCCTGAGCTGGAGAGGGAGGACGGCAAGGAGGAC ATCAGCGATCCAATGGACCCTAATCCATGTTCCGC CACCTACTCTAACCTGGGACAGTCCAGGGCCGCAA TGATCCCTCCAAAGCAGCCAAGGCAGCCAAAGGGA GCCGTGGACGATGCAATCGCCTTTGGCGGCAAGAC CGATCAGGAGGCCCCCAATGCAAGCCAGCCTACAC CACCTCCACTGCCCAAGAAGATGATCATCCGGGCC AACACCGAGCCTATCTCTAAGGACCTCCAGAAGAG CATGGAGAGCAGCCTGTGCGTGATGGCCAATCCCA CCTACGACATCGATCCTAACTGGGACGCCTCTAGC GCCGGCTCCTCTATCAGCTACGAGCTGAAGGGCCT GGATATCGAGTCCTATGACAGCCTGGAGCGGCCCC TGAGGAAGGAGCGCCCCGTGCCTAGCGCCGCCAAC TCCATCAGCTCCCTGACCACACTGTCCATCAAGGA CAGATTTAGCAACTCAATGGAGAGCCTGAGTAGTA GGAGGGGCCCAAGCTGTAGACAGGGGAGA**GGATCC** GGTATTCAAAAGCCGCAACGCCAAGCGCTGTATCG TGGTCTGGAGAACCGTGAGGAAGTGGTGGGCAAAA TTCGTAGCCTGCACACCGACGCGCTGAAGAAACTG GCGGTTAAATGCGAGGACCTGTTCATGGCGGGTCA GAAGGATCAACTGCGTTTCGGCGTGGACAGCTGGA GCGATTTTCGTCTGACCAGCGACAAACCGTGCTGC GAGGCGGGTGATGCGGTTTACTATACCGCGAGCTA CGCGAAAGACCCGCTGAACAACTATGCGGTGAAAA TCTGCAAGAGCAAAGCGAAGGAAAGCCAGCAATAC TATCACAGCCTGGCGGTTCGTCAGAGCCTGGCGGT TCACTTCAACATTCAGCAAGATTGCGGCCACTTTC TGGCGGAAGTTCCGAACCGTCTGCTGCCGTGGGAG GACCCGGATGATCCGGAGAAAGATGAAGACGATAT GGAGGAAACCGAGGAAGACGCGAAGGGTGAAACCG ATGGCAAAAACCCGAAGCCGTGCAGCGAAGCGGCG AGCAGCCAGAAGGAGAACCAAGGTGTTATGAGCAA GAAACAGCGTAGCCACGTGGTTGTGATCACCCGTG AAGTGCCGTGCCTGACCGTTGCGGACTTTGTGCGT GATAGCCTGGCGCAACACGGTAAAAGCCCGGACCT GTACGAACGTCAGGTGTGCCTGCTGCTGCTGCAAC TGTGCAGCGGCCTGGAGCACCTGAAGCCGTACCAC GTTACCCACTGCGATCTGCGTCTGGAAAACCTGCT GCTGGTTCACTATCAACCGGGTGGCACCGCGCAAG GTTTTGGTCCGGCGGAACCGAGCCCGACCAGCAGC TATCCGACCCGTCTGATCGTTAGCAACTTTAGCCA GGCGAAACAAAAGAGCCACCTGGTGGACCCGGAAA TTCTGCGTGATCAGAGCCGTCTGGCGCCGGAGATC ATTACCGCGACCCAGTACAAGAAATGCGACGAGTT	

		CCAAACAGGTATCCTGATTTATGAGATGCTGCACC TGCCGAACCCGTTTGACGAGAACCCGGAACTGAAA GAGCGTGAATACACCCGTGCGGATCTGCCGCGTAT CCCGTTCCGTAGCCCGTACAGCCGTGGTCTGCAAC AACTGGCGAGCTGCCTGCTGAACCCGAACCCGAGC GAGCGTATCCTGATTAGCGACGCGAAAGGTATCCT GCAATGCCTGCTGTGGGGTCCGCGTGAGGACCTGT TCCAAACCTTTACCGCGTGCCCGAGCCTGGTTCAG CGTAACACCCTGCTGCAAAACTGGCTGGACATCAA GCGTACCCTGCTGATGATTAAATTTGCGGAGAAGA GCCTGGACCGTGAAGGTGGCATTAGCCTGGAGGAT TGGCTGTGCGCGCAGTACCTGGCGTTTGCGACCAC CGATAGCCTGAGCTGCATTGTGAAGATTCTGCAAC ACCGTTGA**GTCGACCTCGAG**	

Note: Sequences representative of WT constructs. Mutant constructs have a single codon change for the mutation and are coloured as follows, SgK223: L955E, F992A, C996R, P999G, V1089R, F1271A, A1367D, SgK269: I1290E/A, L1301A, F1327A, S1331R, P1334G, V1460R, F1609A, F1706A, A1707D, W1722A. Restriction sites are in blue.

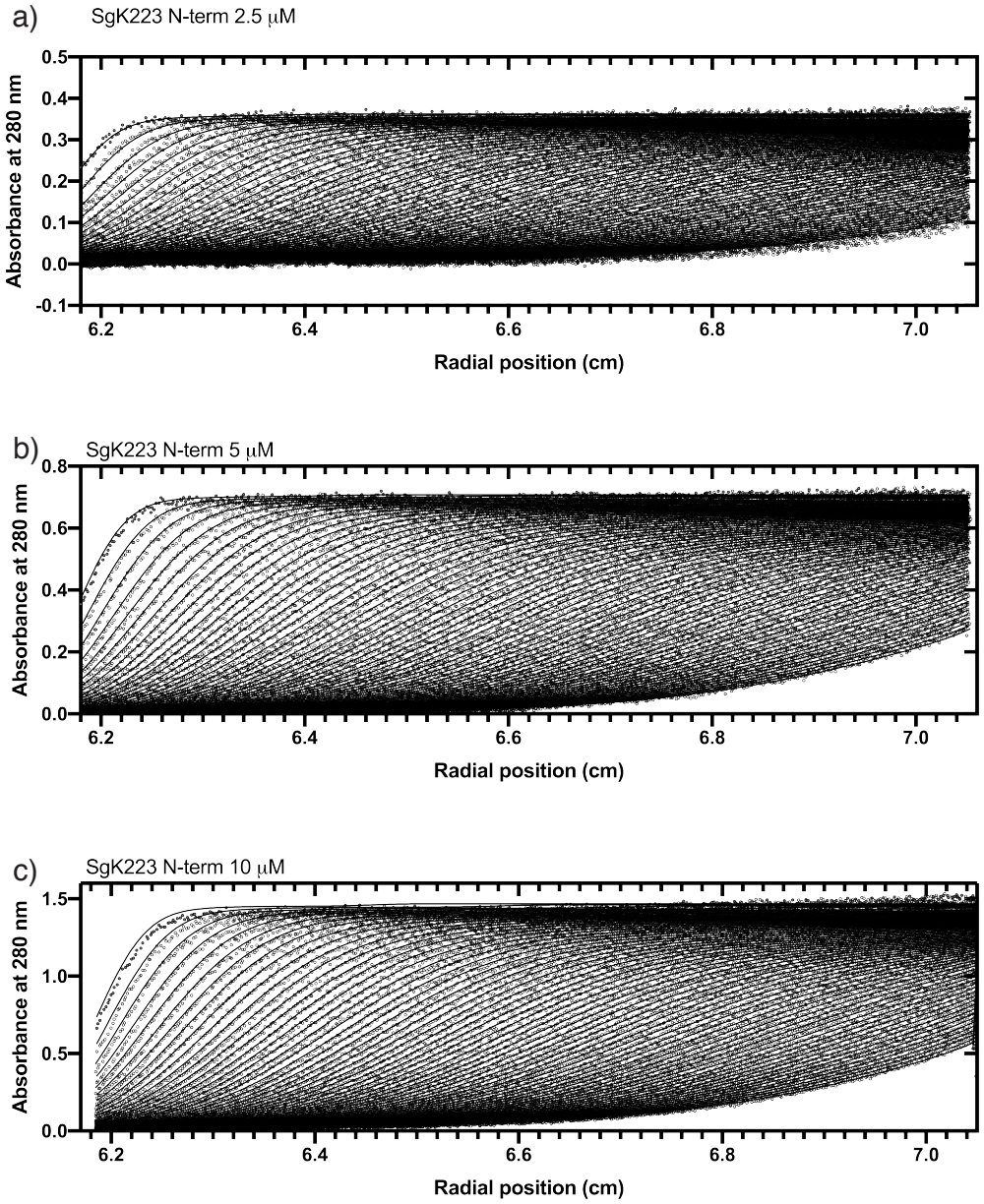


Figure 8.1. SgK223 N-terminal domain AUC data fitting.
a) SgK223 N-terminal domain AUC data fitting of AUC run at 2.5 μM sample concentration.
b) SgK223 N-terminal domain AUC data fitting of AUC run at 5 μM sample concentration.
c) SgK223 N-terminal domain AUC data fitting of AUC run at 10 μM sample concentration.

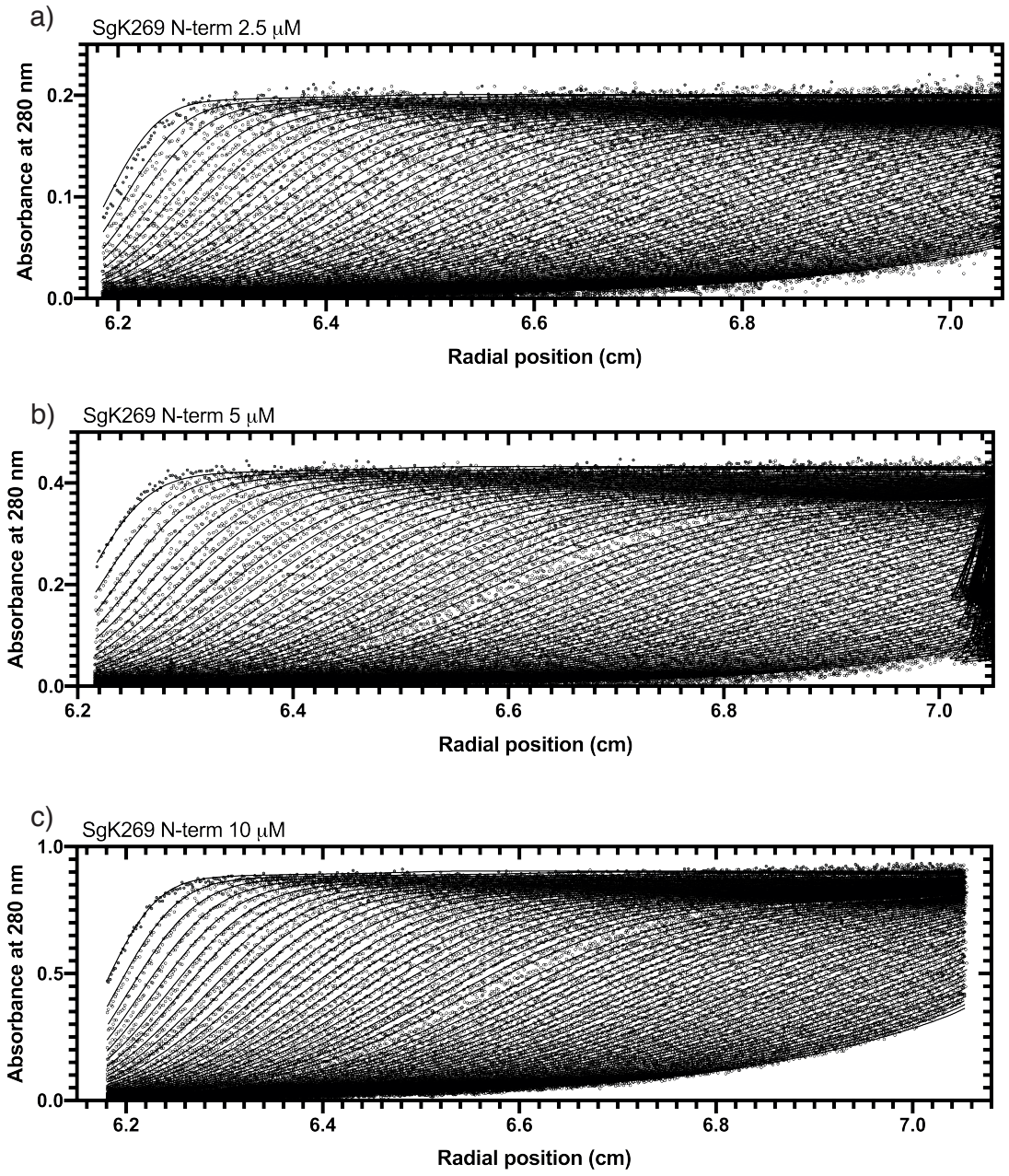


Figure 8.2. SgK269 N-terminal domain AUC data fitting.
a) SgK269 N-terminal domain AUC data fitting of AUC run at 2.5 μM sample concentration.
b) SgK269 N-terminal domain AUC data fitting of AUC run at 5 μM sample concentration.
c) SgK269 N-terminal domain AUC data fitting of AUC run at 10 μM sample concentration.

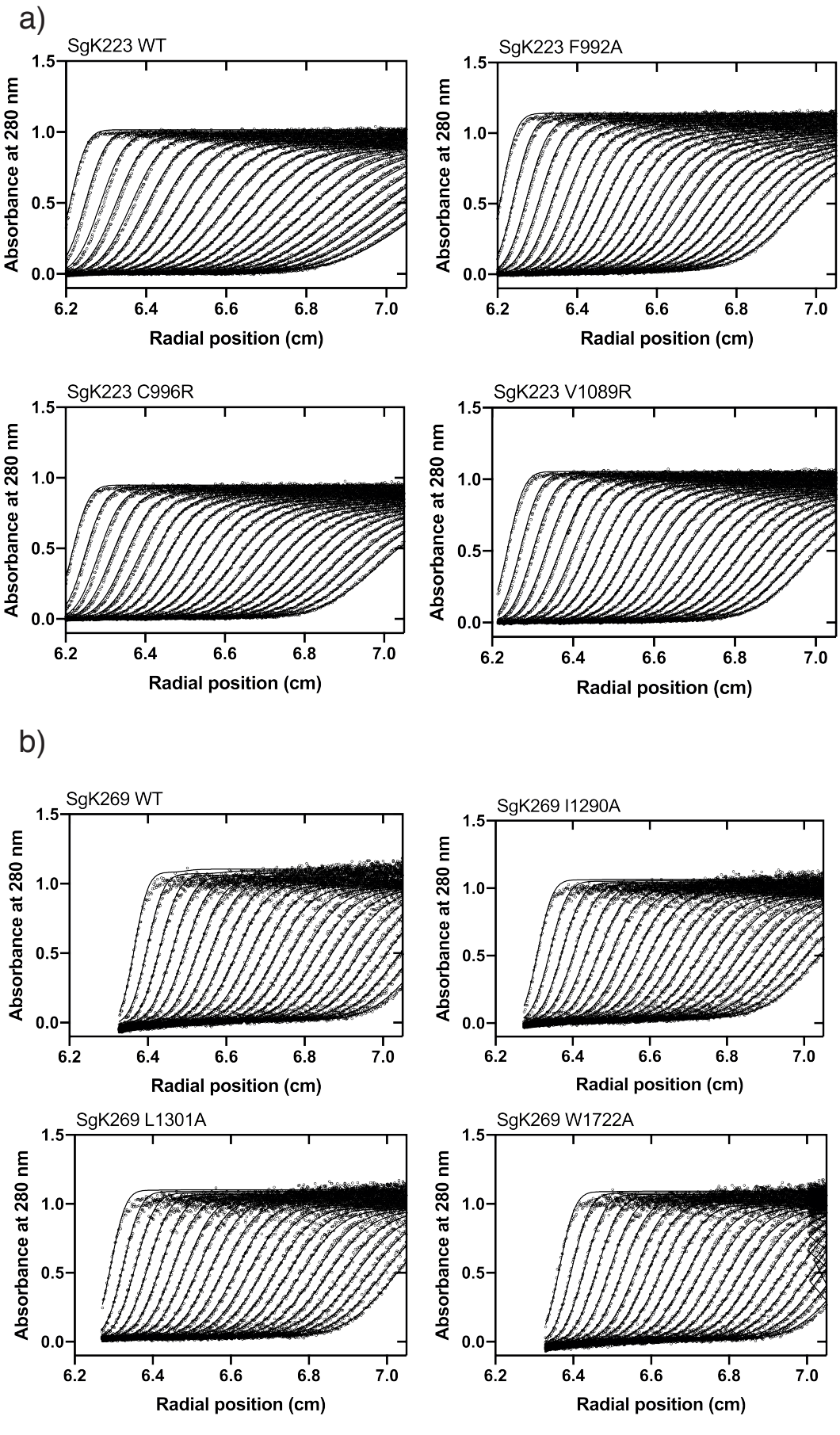

a)
SgK223 WT
SgK223 F992A
SgK223 C996R
SgK223 V1089R
b)
SgK269 WT
SgK269 I1290A
SgK269 L1301A
SgK269 W1722A
Absorbance at 280 nm
Radial position (cm)
1.5
1.0
0.5
0.0
6.2
6.4
6.6
6.8
7.0

Figure 8.3. SgK223 and SgK269 WT and mutants AUC data fitting.
a) SgK223 WT and pseudokinase domain N-lobe mutants AUC data fitting of AUC run at 20 μM sample concentration.
b) SgK269 WT and dimerisation domain mutants AUC data fitting of AUC run at 3 μM sample concentration.