Structural and biochemical characterisation of the regulatory principles of Twitchin Kinase, a member of the titin-like family of muscle kinases

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Rhys Mathew Williams

September 2016

Abstract

The titin-like kinases are mechanotransducers of stretch-activated signalling pathways in the muscle sarcomere, thought to undergo mechanoactivation in response to stretch in the myofibril. This work explores the mechanistic principles of twitchin kinase, a titin-like kinase from *Caenorhabditis elegans*. We show that twitchin kinase undergoes extensive autophosphorylation of its catalytic domain, identifying four modification sites in key mechanistic regions. We find that twitchin kinase is not activated by autophosphorylation as is common in protein kinases, but uses autophosphorylation to inhibit catalysis. We conclude that modification of the twitchin kinase substrate-binding region is likely to be a major contributor to inhibition, with phosphorylation at multiple sites having a cumulative effect to silence catalysis.

We present the crystal structure of the active twitchin kinase conformation, showing that conformational changes in twitchin kinase are limited to closure of the glycine-rich loop, with the αC-helix adopting an unusually open conformation while still maintaining catalytically essential active site interactions. Our structure supports the hypothesis of mechanical activation, in which significant rearrangement of the catalytic domain would not be possible under mechanical stress.

Finally, we establish the use of DNA molecular springs to study the mechanoenzymatic properties of twitchin kinase. Twitchin kinase:DNA chimeras have been constructed to apply mechanical stress while monitoring catalysis. We show that twitchin kinase is catalytically active under mechanical deformation, supporting the hypothesis of mechanoactivation and providing a reliable method for future work.

Together, these results show that twitchin kinase is likely to employ a triple mechanism for regulating catalysis, combining biochemical regulation by autophosphorylation with intrasteric inhibition by N- and C-terminal regulatory tails, indicative of the need for strict regulation of twitchin kinase activity.

Table of Contents

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Olga Mayans, for giving me the opportunity to join her lab and work on such an interesting project, for the continued support over the past four years, the discussions, the stories, and for the chance to experience living and working in both Liverpool and Konstanz. It's been a great adventure and I am extremely grateful.

I thank my second supervisor, Dr. Daniel Rigden for his helpful advice and guidance, Dr. Mark Wilkinson for his work and ideas that contributed greatly to this project, and to Dr. Patrick Eyers for teaching me new techniques and always being available to give advice. I'm grateful to Prof. Giovanni Zocchi and his group for hosting me in Los Angeles, giving me a great experience and the opportunity to learn new skills. Thank you to Dr. Julius Bogomolovas for all the help and materials provided.

I would like to thank all of the members of the Mayans group and Lab C, it's been a pleasure to work with you all. In particular, to Chris, Jen, Dani, Jens and Bruno, your support and friendship both in and out of the lab has meant so much. A special thank you goes to Barbara who has taught me, guided me, looked after me and even fed me every step of the way. I really couldn't have done it without you.

To Anshul and Sean, thank you for every drink, every laugh, every trip we took and every bizarre situation that we found ourselves in. You were always there for me and my time in Liverpool was better for having shared it with you.

Finally, thank you to my family. To my brother for making me feel like I've never left whenever I come home. To my Nan for always being so proud and encouraging. To my Mum and Dad, I don't think I could ever find the words to truly describe how much your love and support has meant to me. I am where I am because of you and to you I dedicate this thesis.

List of Figures

List of Tables

List of Abbreviations

AEC	Anion exchange chromatography
ADP	Adenosine diphosphate
AFM	Atomic force microscopy
AmC3	C6 amine (DNA modification)
AmC6	C3 amine (DNA modification)
Aplysia	*Aplysia californica*
approx.	Approximately
ATP	Adenosine triphosphate
Ball	Bällchen kinase
Bis Tris	2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol
β-ME	β-Mercapto-ethanol
BSA	Bovine serum albumin
Bub1	Budding uninhibited by benzimidazoles
CAK	Cdk-activating kinase
CAMK	Ca2+/calmodulin-dependent protein kinase
Cdk	Cyclin dependent kinase
C. elegans	*Caenorhabditis elegans*
CPNA-1	Copine domain protein, atypical 1
CSK	C-terminal Src kinase
c-Src	Cellular Src kinase
CTD	C-terminal domain(-type phosphatase)
DAPK	Death-activated protein kinase
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol

E. coli	*Escherichia coli*
EPR	Electron paramagnetic resonance spectroscopy
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
FHL	Four-and-a-half LIM domains
Fig	Figure
FnIII	Fibronectin type III
HEPES	4-(2-Hydroxyethyl)-pperazine-1-sulfonic acid
Ig	Immunoglobulin
IPTG	Isopropyl-b-D-thiogalactopyranoside
IRE1	Inositon-requiring enzyme 1
ITC	Isothermal titration calorimetry
Kin	Kinase domain (in reference to titin-like kinases)
LIM	Lin11, Isl-1 & Mec-3
MAK-1	*C. elegans* orthologue of MAPKAP kinase 2
MAPK	Mitogen-activated protein kinase
MAPKAP	Mitogen-activated protein-activated protein kinase
MARK2	Microtubule affinity regulating kinase 2
MARKK	Microtubule affinity regulating kinase-kinase
MASK	Multiple ankyrin repeats single KH domain
MD	Molecular dynamics
MEI-1	Meiotic double-stranded break formation protein
MEL-26	Maternal effect lethal-26
MLCK	Myosin light-chain kinase
MOPS	(3-(N-morpholino)propanesulfonic acid)
MuRF	Muscle specific RING finger protein
MWCO	Molecular weight cut-off

Nbr1	Neighbour of BRCA1
NL	N-terminal linker
OD_{600}	Optical density at 600 nm
OT	Optical tweezers
PAT-4	Paralysed arrest at two-fold 4 (*C. elegans* integrin-linked kinase homologue)
PAT-6	Paralysed arrest at two-fold 6 (*C. elegans* actopaxin/α-parvin homologue)
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEVK	Proline Glutamate Valine Lysine
Phos	Phosphorylation (DNA modification)
pI	Isoelectric point
PINCH	Particularly interesting new cysteine-histidine-rich protein
PK1	Protein kinase domain 1
PK2	Protein kinase domain 2
PKA	Protein kinase A (cAMP-dependent protein kinase)
PKC	Protein kinase C
PP2A	Protein phosphatase 2
Raf-1	Cellular rapidly accelerated fibrosarcoma kinase
RLC	Regulatory light chain
RSK1	Ribosomal kinase 1
SAXS	Small angle X-ray scattering
SCPL-1	Small CTD phosphatase-like 1
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SH2	Src homology 2

SH3	Src homology 3
SRF	Serum response factor
TCEP	Tris(2-carboxyethyl) phosphine hydrochloride
TEV	Tobacco etch virus
TK	Titin kinase
Tris	Tris-(hydroxymethyl)-aminomethane
TTN-1	Small titin
TwcK	Twitchin kinase
UNC	Uncoordinated
Y2H	Yeast two-hybrid
ZIPK	Zipper-interacting kinase

Chapter 1
General Introduction

Muscle building and function relies on mechanical stimulation yet little is currently known about how mechanical signals are sensed by the myofibril and translated into cellular signalling pathways to control muscle homeodynamics. It appears increasingly likely that the kinase domains of the titin-like family of giant sarcoskeletal proteins are involved in the sensing and transduction of mechanical signals, acting as mechanosensory nodes in stretch-activated signalling pathways (Lange *et al.,* 2005; Butler & Siegman, 2011). However, the molecular principles of how and when these kinases may carry out mechanotransduction remain unknown. In this work, we use the kinase domain of twitchin from *Caenorhabditis elegans* (*C. elegans*), twitchin kinase, as a representative member of the titin-like family of kinases to explore the molecular mechanisms of catalysis and mechanosensitivity.

1.1 The sarcomere and its components

The sarcomere is the basic contractile unit of muscle, many repeats of which align to form myofibrils and these, in turn, associate into muscle fibres. The sarcomere is composed of a number of protein filaments that are involved in muscle contractility, signalling and cytoskeletal assembly and is divided into regions based on early observations by light microscopy (**Fig 1.1.1a**). In vertebrate sarcromeres, the Z (zwischen)-disc appears as a dark line and delineates the boundary between adjacent sarcomeres. The I (isotropic)-band is seen as a narrow region immediately

following the Z-disc and is named for its isotropic nature in polarised light. The largest part of the saromere is formed by the A (anisotropic)-band, named for its anisotropy under polarised light. The A-band is intersected by the less dense H (hell)-zone, at the centre of which is found the dense M (mittel)-line, marking the centre-point of the sarcomere (Sjostrom and Squire, 1977).

The sarcomeres of invertebrate muscle are similar in structure to vertebrate sarcomeres but show several key differences **(Fig 1.1.1c)**. Invertebrate muscles are oblique, meaning that the sarcomeres are arranged at an oblique angle with respect to the longitudinal axis. In place of the vertebrate Z-disc are found analogous structures known as dense bodies, the shape of which help give rise to the oblique sarcomeric arrangement. In the nematode *C. elegans*, all M-lines and dense bodies are anchored to the basal lamina and hypodermis. In body wall muscle for example, this provides a direct connection between the muscle and the cutical to transmit the force of muscle contraction into movement (Benian & Epstein, 2011).

The sarcomere contains three main filament systems **(Fig 1.1.b)**. The thin filaments are largely composed of globular actin (G-actin) polymerised into actin fibres (F-actin) and the thick filaments are mostly composed of myosin. Actin and myosin act as the molecular motors for muscle contraction, with conformational changes induced by myosin hydrolysis of ATP allowing for the sliding of the thin and thick filaments along each other as described by the sliding filament mechanism of muscle contraction (Clark *et al.*, 2002).

The 3rd sarcomeric filament system in vertebrates is formed by titin, a giant filamentous protein of up to 4 MDa in molecular mass that spans a half sarcomere from Z-disc to M-line (~1.2 μm). Titin is composed of a large number of repeating

Immunoglobulin (Ig) and Fibronectin III (FnIII) domains as well as several unique sequence elements, including the flexible PEVK-rich region that acts as an entropic spring during extension (Linke *et al.,* 2002). Titin contributes to elasticity and passive tension in the sarcomere and acts as a scaffold for interactions with motor proteins and a variety of components involved in processes such as muscle turnover and stress response (Kontrogianni-Konstantopoulous *et al.,* 2009; Linke and Krüger, 2010; Bogomolovas *et al.,* 2014). Various titin-like proteins are found across the vertebrate and invertebrate muscle diversity, ranging from 0.7 to 4 MDa in molecular mass. Members of the titin protein family include titin and obscurin in mammals; twitchin/UNC-22, the obscurin homologue UNC-89 and the small titin TTN-1 in nematodes; twitchin in molluscs; projectin, UNC-89 and stretchin in insects (Bullard *et al.,* 2002; Kontrogianni-Konstantopoulous *et al.,* 2009; Krüger & Linke, 2011). The proteins of the titin-like family share structural similarity with vertebrate titin, although their functions vary.

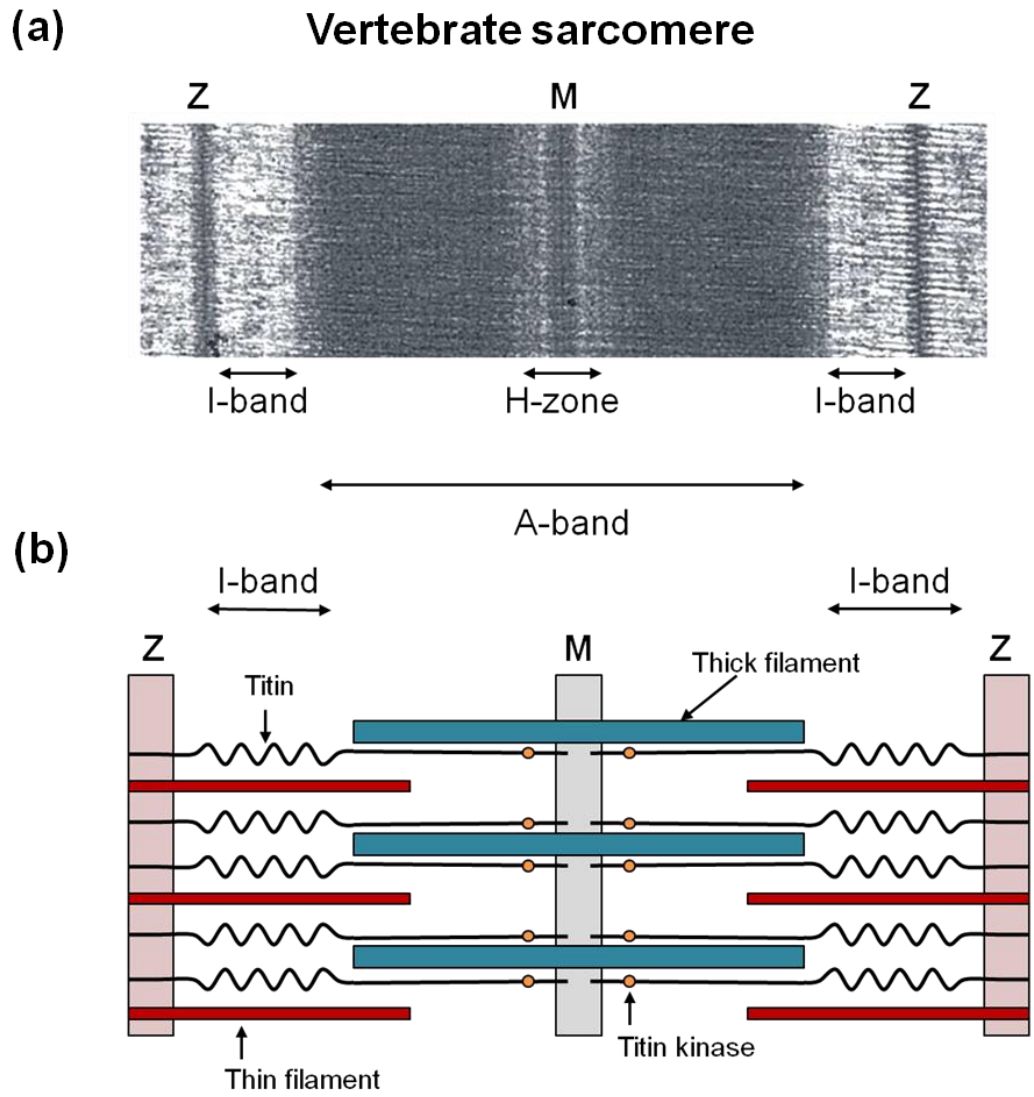
Of particular interest, titin-like proteins contain one or two protein kinase domains near their C-terminus. The kinase locus of titin-like proteins is highly conserved amongst family members and is characterised by an Ig-Ig-FnIII-linker-kinase-tail-Ig domain composition **(Fig.1.2.1)**. Muscle function, development and regulation are mediated by mechanical stimulation and the kinase regions of titin-like proteins have been identified as likely sites of mechanical signal transduction (Lange *et al.,* 2005; Butler & Siegman, 2011). However, the regulatory mechanisms of the titin-like kinases are not fully understood. The investigation of the structure and regulation of the kinase region from twitchin, a titin-like protein from the *C. elegans* sarcomere forms the basis of this thesis.

In *C. elegans*, twitchin is a titin-like protein encoded by the *unc-22* gene that localises to the A-band of the sarcomere, associating with the myosin-based thick filaments in body wall muscle **(Fig 1.1.1c)** (Moerman *et al.,* 1988) Worms with mutations in *unc-22* show a characteristic twitching phenotype (Waterston *et al.,* 1980), from which twitchin takes its name. Twitchin is composed of 31 FnIII, and 30 Ig domains, several unique sequences and a single protein kinase domain near to the C-terminus. The FnIII and Ig domains are mostly arranged into Ig-FnIII-FnIII repeats, showing a high similarity to the organisation of the A-band region of human titin. In particular, the organisation of these repeats surrounding the kinase domain, comprising 16 domains, is conserved between twitchin and human titin (Benian *et al.*, 1996).

Twitchin does not span the half sarcomere like vertebrate titin but its localisation to the A-band and domain organisation suggests that twitchin plays a similar scaffolding role to titin and that the location of the titin and twitchin kinase domains are likely to be comparable. The sarcomeres of *C. elegans* muscle are considerably larger than their vertebrate counterparts with a thick filament length of approx. 9.7 µm in adult muscle (Mackenzie & Epstein, 1980). *C. elegans* sarcomeres therefore do not contain a single protein that spans both the I and A bands like vertebrate titin and it is thought that multiple titin-like proteins may combine to provide the equivalent function of vertebrate titin (Benian & Epstein, 2011). The small titin, TTN-1, is anchored to the dense body of the *C. elegans* sarcomere (equivalent of the vertebrate Z-disc) and extends through the I-band, possibly extending to the outer edge of the A-band (Forbes *et al.,* 2010). The domain organisation of TTN-1 is similar to the I-band region of titin, containing tandem Ig domains and several sequences that are likely to be elastic elements, as well as a C-

terminal kinase domain (Benian & Epstein, 2011). The obscurin homologue, UNC-89, contains two C-terminal kinase domains and associates with the M-line of the *C. elegans* sarcomere, with *unc-89* mutants showing disorganisation of the myofilament lattice and a lack of M-lines (Waterston *et al.*, 1980). It has been hypothesised that a combination of twitchin, TTN-1 and UNC-89 could be functionally equivalent to vertebrate titin (Benian & Epstein, 2011), although this is yet to be shown experimentally.

Fig 1.1.1 The vertebrate and *C. elegans* sarcomeres (a) Electron micrograph of the vertebrate sarcomere from striated muscle. Adapted from Luther, 2009. **(b)** Schematic representation of the vertebrate sarcomere showing the three filament systems. Thick filament (blue), thin filament (red) and titin (black) are shown as lines, the C-terminal kinase domain of titin is shown as orange circles. **(c)** Schematic representation of the *C. elegans* body wall muscle sarcomere. Thick filament (blue), thin filament (red) and intermediate filaments (purple) are shown as lines. Twitchin (black) is localised to the A-band and TTN-1 (purple) is localised to the I-band. Kinase domains of titin-like proteins are shown as circles. Unc-89 (not shown) is localised to the M-line.


(a)
Vertebrate sarcomere
Z
M
Z
I-band
H-zone
I-band
A-band
(b)
I-band
I-band
Z
M
Z
Titin
Thick filament
Titin kinase
Thin filament
(c)
C. elegans sarcomere
Dense body
Twitchin
M
TTN-1
Dense body
Thin filament
Thick filament
Basal lamina
Hypodermis
Intermediate filaments
Cuticle

1.2 Titin-like kinases as platforms for muscle signalling

The kinase regions of titin-like proteins were identified as regulators of stretch-activated signalling pathways in muscle and involvement of titin-like kinases has been demonstrated in diverse cellular pathways. *C. elegans* twitchin kinase (TwcK) has been shown through yeast two-hybrid (Y2H) screening to interact with the kinase MAK-1 **(Fig 1.2.1a),** a nematode orthologue of mammalian MAPKAP kinase 2 (Mitogen-activated protein kinase-activated protein kinase 2) (Matsunaga *et al.,* 2015). TwcK can be phosphorylated by MAK-1 *in vitro* and this has been suggested as a potential mechanism of TwcK activation (Matsunaga *et al.,* 2015). The kinase domain of MAK-1 is 53% identical to the kinase domain of MAPKAP kinase 2, and is expressed in *C. elegans* body wall muscle, localised to the sarcomere around dense bodies and overlapping with the A-band localisation of twitchin (Matsunaga *et al.,* 2015). MAPKAP kinase 2 is involved in stress-activated signalling in mammalian cardiac muscle (Zu *et al.,* 1997) and has been shown to phosphorylate myosin regulatory light chain (RLC) II, activating myosin ATPase activity (Komatsu & Hosoya, 1996). Stress-induced activation of mammalian MAPKAP kinase 2 is through phosphorylation by p38 (Ben-Levy *et al.,* 1995) and these phosphorylation sites are conserved in *C. elegans* MAK-1. Along with MAK-1, three p38 MAPKs (mitogen-activated protein kinases) are also expressed in *C. elegans* body wall muscle. It is thought that through interaction with MAK-1, TwcK may be involved in a p38 signalling cascade that regulates muscle contractile activity in response to stress (Matsunaga *et al.,* 2015).

In vertebrates, titin kinase (TK) has been suggested to control muscle gene expression and protein turnover in response to mechanical demand (Lange *et al.*, 2005). TK interacts in a mechanically inducible manner with Nbr1 (Neighbour of BRCA1), a component of a ubiquitin based signalling pathway that recruits the muscle-specific E3 ubiquitin ligase, MuRF2 (muscle specific RING finger protein 2), via interaction with p62 **(Fig.1.2.1b)**. During inactivity, MuRF2 is released, leading to its nuclear translocation where it represses the regenerative response of the serum response transcription factor (SRF) (Lange *et al.*, 2005). The conserved Ig-Ig-FnIII tandem of TK has also been shown to form a scaffold for recruitment of another E3 ubiquitin ligase, MuRF1, a protein linked to muscle atrophy brought about by muscle disuse **(Fig.1.2.1b)** (Centner *et al.*, 2001; Mrosek *et al.*, 2007; Bogomolovas *et al.*, 2014). This interaction is a mediator of stress response in muscle, with MuRF1 involved in stress-induced ubiquitination in the sarcomere to regulate the trophic state of muscle via protein degradation (Centner *et al.*, 2001; Mrosek *et al.*, 2007; Bogomolovas *et al.*, 2014).

The kinase regions of obscurin (UNC-89 in invertebrates) also play a role in regulation of protein turnover in muscle through ubiquitin degradation pathways as well as sarcomeric organisation. Obscurin/UNC-89 have two kinase domains at their C-terminus, PK1 and PK2 (protein kinase domain 1 and 2), that are linked by a region of low sequence complexity (Small *et al.*, 2004) that is thought to act as an elastic component (Mayans *et al.*, 2013). In *C. elegans*, the Ig-FnIII domains N-terminal to PK2 interact with MEL-26 (maternal effect lethal-26), which recruits cullin-3, a scaffold for assembly of factors from the ubiquitin degradation pathway **(Fig 1.2.1c)** (Wilson *et al.*, 2012). The cullin-3−MEL-26 complex targets MEI-1 (Meiotic double-stranded break formation protein) during *C. elegans* embryonic

development, a protein responsible for thick filament assembly (Pintard *et al.*, 2005). *Unc-89* mutations lead to disorganisation of thick filaments, as does loss- or gain-of-function in *mel-26* or *mei-1* (Wilson *et al.,* 2012), suggesting that the role of the UNC-89−MEL-26 interaction is to inhibit cullin-3−MEL-26 mediated degradation of MEI-1, facilitating thick filament organisation (Mayans *et al.,* 2013). The link between obscurin and cullin-3 recruitment has also been shown in mice (Lange *et al.,* 2012), indicating a related role for UNC-89/obscurin in ubiquitin mediated protein degradation pathways in both vertebrates and invertebrates.

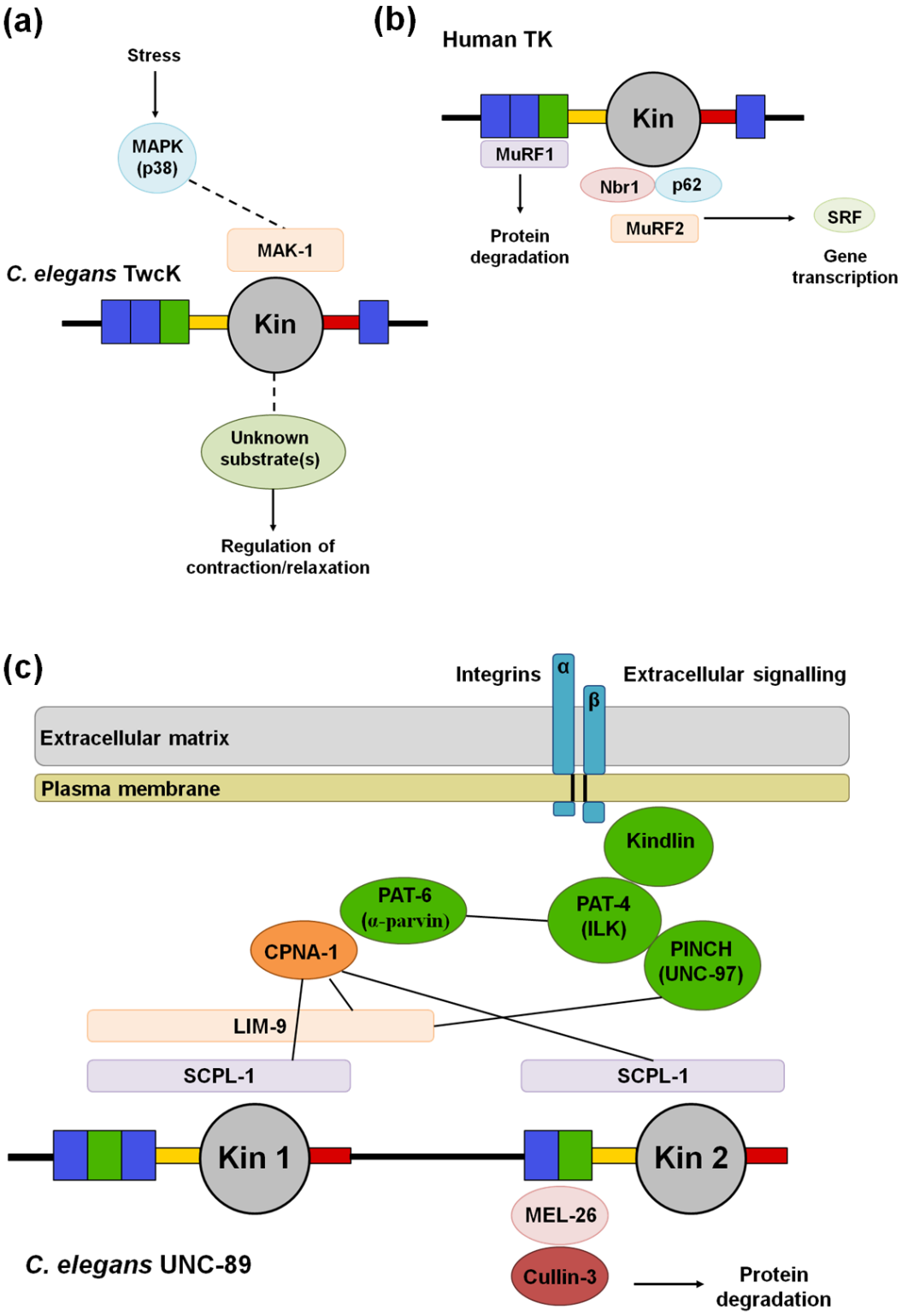
In *C. elegans,* PK1 and PK2 domains of UNC-89 interact with a CTD (C-terminal domain)-type phosphatase, SCPL-1 (small CTD phosphatase-like 1) **(Fig 1.2.1c)**, mutation of which leads to defective egg-laying muscle (Qatoda *et al.,* 2008). Both PK1 and the interkinase linker also interact with LIM-9 (Xiong *et al.,* 2009), the nematode homologue of FHL (four-and-a-half LIM domains) which associates with the elastic I-band region of titin (Krüger & Linke, 2011). In *C. elegans*, LIM-9 interacts with UNC-97 (PINCH) in the M-line (Xiong *et al.,* 2009). UNC-89, SCPL-1 and LIM9 bind CPNA-1 (copine domain protein, atypical 1), a protein that localises to integrin-adhesion sites in the M-lines and dense bodies of *C. elegans* body wall muscle. Here, through interactions with PAT-6 (Paralysed arrest at two-fold), a *C. elegans* actopaxin/α-parvin homologue, (Warner *et al.,* 2013), which in turn interacts with PAT-4, an integrin-linked kinase homologue (Lin *et al.,* 2003), CPNA-1 provides a functional link between the UNC-89 kinase domains and integrin signalling pathways of the cell surface and cytoskeleton.

In the indirect flight muscle of *Drosophila melanogaster*, the kinase domains of UNC-89 recruit proteins required for correct development and organisation of the sarcomere (Katzemich *et al.,* 2015). PK1 binds the catalytically active Ball

(Bällchen) kinase and both PK1 and PK2 bind the large ankyrin repeat containing protein, MASK (Multiple ankyrin repeats single KH domain) (Katzemich *et al.*, 2015). Reduced expression of Ball and MASK while UNC-89 was still present caused disorganisation of the sarcomere, including missing M-lines, suggesting UNC-89 acts as a platform for Ball and MASK binding (Katzemich *et al.*, 2015).

.

Fig.1.2.1 Signalling interactions of titin-like kinases. Ig domains (purple) and FnIII (green) are shown as rectangles. N-terminal (yellow) and C-terminal (red) tails are shown as lines. Kin, kinase domain. **(a)** Proposed involvement of *C. elegans* TwcK in p38-mediated stress-induced signalling. **(b)** Protein interactions mediated by human TK **(c)** Protein interactions mediated by the kinase regions of *C. elegans* UNC-89. ILK, integrin-linked kinase.


(a)
Stress
MAPK (p38)
MAK-1
C. elegans TwcK
Kin
Unknown substrate(s)
Regulation of contraction/relaxation
(b)
Human TK
Kin
MuRF1
Nbr1
p62
MuRF2
SRF
Protein degradation
Gene transcription
(c)
Integrins
α
β
Extracellular signalling
Extracellular matrix
Plasma membrane
Kindlin
PAT-6 (α-parvin)
PAT-4 (ILK)
CPNA-1
PINCH (UNC-97)
LIM-9
SCPL-1
SCPL-1
Kin 1
Kin 2
MEL-26
Cullin-3
Protein degradation
C. elegans UNC-89

1.3 Titin-like kinase catalysis

The catalytic properties of titin-like kinases are poorly understood and only one physiological substrate is currently known. TwcK from the mollusc *Aplysia californica* has been shown to phosphorylate myosin regulatory light chains (RLCs) (Heierhorst *et al.,* 1995). TwcK phosphorylation of mysoin RLC occurs at Thr15 (Heierhorst *et al.,* 1995), a phosphorylation site analogous to the substrate of the myosin light chain kinases (MLCKs) that regulate muscle contraction in vertebrates (Michnoff *et al.,* 1986; Kamm & Stull, 2011). Twitchin has been implicated in the 'catch' state of molluscan smooth muscle, a state that allows for maintenance of prolonged tension that requires little energy input and TwcK catalysis is likely to be activated in a force-dependent manner during this catch state (Butler & Siegman, 2011). Crossbridging of F-actin and myosin is correlated with the phosphorylation state of twitchin, suggesting that twitchin may be involved in connecting myosin and actin filaments in a phosphorylation-dependent manner (Butler & Siegman, 2011). Increased mechanical stress in the catch state correlated with increased activity of a force-activated kinase capable of phosphorylating a myosin RLC derived peptide substrate (Butler & Siegman, 2011). This suggests a role for TwcK activity in regulation of muscle contraction/relaxation, although it has not been confirmed definitively that this force-activated kinase is indeed TwcK.

The titin-like kinase family contains a number of atypical kinases (Mayans *et al.,* 2013). Human TK has been recently classified as a pseudokinase, lacking catalytic activity and acting as an inactive signalling scaffold (Bogomolovas *et al.,* 2014) and the obscurin/UNC-89 kinases show catalytic divergence between vertebrates and invertebrates. Vertebrate obscurin kinases possess canonical kinase

active sites and are able to undergo autophosphorylation (Hu & Kontrogianni-Konstantopoulos, 2013), while the invertebrate UNC-89 kinase domains show atypical active sites that are likely to render many of them inactive (Mayans *et al.*, 2013).

1.4 Structure and mechanoactivation of titin-like kinases

To date, the best-characterised titin-like kinases are human TK and *C. elegans* TwcK. Three crystal structures of titin-like kinases have been solved; Human TK (Mayans *et al.*, 1998); *C. elegans* TwcK (Hu *et al.*, 1994; Kobe *et al.*, 1996) and TwcK from the mollusc *Aplysia californica* (Kobe *et al.*, 1996), all of which show kinases in an autoinhibited state. Initial crystal structures of TK and TwcK revealed protein kinase domains auto-inhibited by a C-terminal extension, dubbed the C-terminal regulatory domain (CRD). The CRD of TK and TwcK occupy the active-site cleft of the kinase domain, blocking the ATP-binding pocket **(Fig 1.4.1)**. The more recent crystal structure of the TwcK region (TwcKR) revealed TwcK in its molecular context (FnIII-linker-kinase-tail-Ig) (von Castelmur *et al.*, 2012). In addition to the CRD, the N-terminal linker (NL) wraps around the N-terminal lobe of the kinase domain, forming multiple contacts with the kinase domain and occupying the kinase hinge region as well as interacting with the CRD at the ATP binding site **(Fig 1.4.1)** (von Castelmur *et al.*, 2012).

Maximal phosphotransfer activity is seen in the absence of both regulatory extensions, with catalytic data showing that the TwcK NL and CRD are roughly equal contributors to autoinhibition, each reducing levels of catalysis by

approximately half. In the presence of both the NL and CRD, TwcK is fully autoinhibited (von Castelmur *et al.*, 2012). The inhibition by the NL is likely due to a restriction of hinge motions that are required for catalysis, while the CRD blocks the active site. The observation that high levels of catalysis are observed in the presence of the CRD was a surprising one, although the CRD segment responsible for blocking of the active site folds in to a rare 3_{10} helix, an energetically unfavourable conformation that may be out-competed by sufficient substrate concentration in the absence of the NL (von Castelmur *et al.*, 2012).

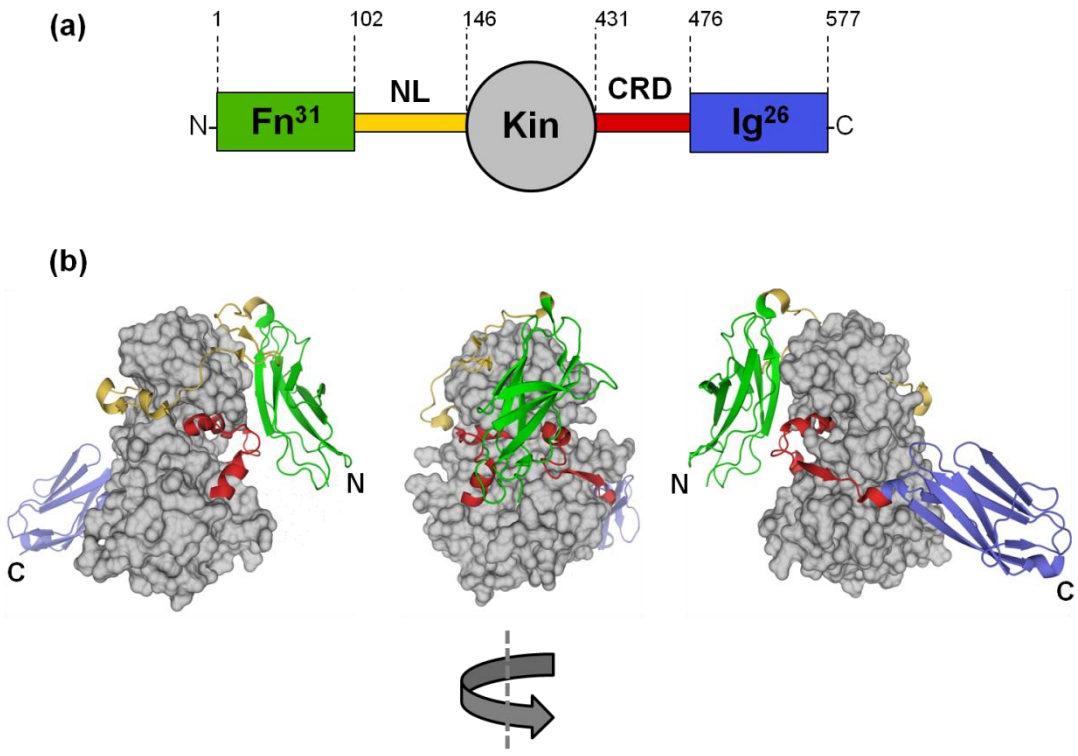


Fig 1.4.1 Crystal structure of the twitchin kinase region (TwcKR) from *C. elegans*. (a) schematic representation of the domain composition of TwcKR with domain boundaries denoted by residue numbers. **(b)** Crystal structure of TwcKR coloured as in (a) and shown in 3 rotated views. N- and C-termini are indicated. The CRD (red) blocks the active site cleft formed between the kinase lobes and the NL (gold) wraps around the kinase and occupies the hinge region (Protein Data Bank (PDB) ID: 3UTO) (von Castelmur *et al.*, 2012).

The mechanism for autoinhibitory tail removal in titin-like kinases is still not fully understood and protein activators that induce tail displacement have not been identified. Calmodulin is able to bind the CRD of TK and TwcK *in vitro* but does not lead to activation (Gautel *et al.,* 1995; Mayans *et al.,* 1998; Lei *et al.,* 1994). TwcK from *C. elegans* was shown to be activated ~1000 fold *in vitro* by S100, a calmodulin-like protein (Heierhorst *et al.,* 1996a), however the genome of *C. elegans* does not encode an S100 protein, ruling it out as a physiological activator of TwcK. TwcK interaction with the kinase MAK-1 is dependent on the CRD and it has been suggested that this interaction and possible phosphorylation of TwcK by MAK-1 may be an activation mechanism (Matsunaga *et al.,* 2015).

With the titin-like kinases shown to be involved in stretch-regulated signalling pathways in muscle, a model was proposed in which the build up of tension in the sarcomere leads to mechanically driven release of the CRD and a catalytically active kinase domain (Gräter *et al.,* 2005; Puchner *et al.,* 2008; Puchner & Gaub, 2010; Stahl *et al.,* 2011). In agreement with the hypothesis of stretch-activation, atomic force microscopy (AFM) and molecular dynamics (MD) simulations on human TK suggested that mechanical force could lead to unfolding of the CRD, freeing the active site for catalysis (Gräter *et al.,* 2005; Puchner *et al.,* 2008; Puchner and Gaub, 2010; Stahl *et al.,* 2011). Mechanically induced unfolding of the CRD was not detected for TwcK in AFM experiments, but due to experimental limitations it could not be concluded whether mechanically induced unfolding of the TwcK CRD occurs (Greene *et al.,* 2008). In steered MD simulations on TwcKR, the NL was suggested to be mechanically labile, with the force required for complete CRD release being high enough to compromise the TwcK catalytic domain (von Castelmur *et al.,* 2012). This suggests that the CRD may remain bound

to TwcK under stretch, possibly playing a stabilising role to protect the active site from mechanical deformation (von Castelmur *et al.*, 2012).

Structural differences between the CRD tails of TK and TwcK go some way to explain the lower propensity for CRD unfolding under stretch in TwcK. The CRD of TwcK is longer and more structured, consisting of 3 α-helices (αR1, αR2, αR3) and a β-strand (βR4) and has a greater number of interactions with the kinase domain **(Fig 1.4.2).** In the CRD of TK, the αR3 helix is substituted for a flexible loop. Mechanical release of the CRD may therefore not represent a common means of activation amongst the titin-like kinases. Two possible methods of mechanosensing have been proposed for the titin-like family. Firstly, removal of the NL could leave the kinase in a mechanically 'primed' state sufficient to permit catalysis, with the CRD remaining bound in a stabilising manner. Secondly, release of the CRD tail could expose the kinase surface, 'baring' it for protein interactions as part of a scaffolding role.

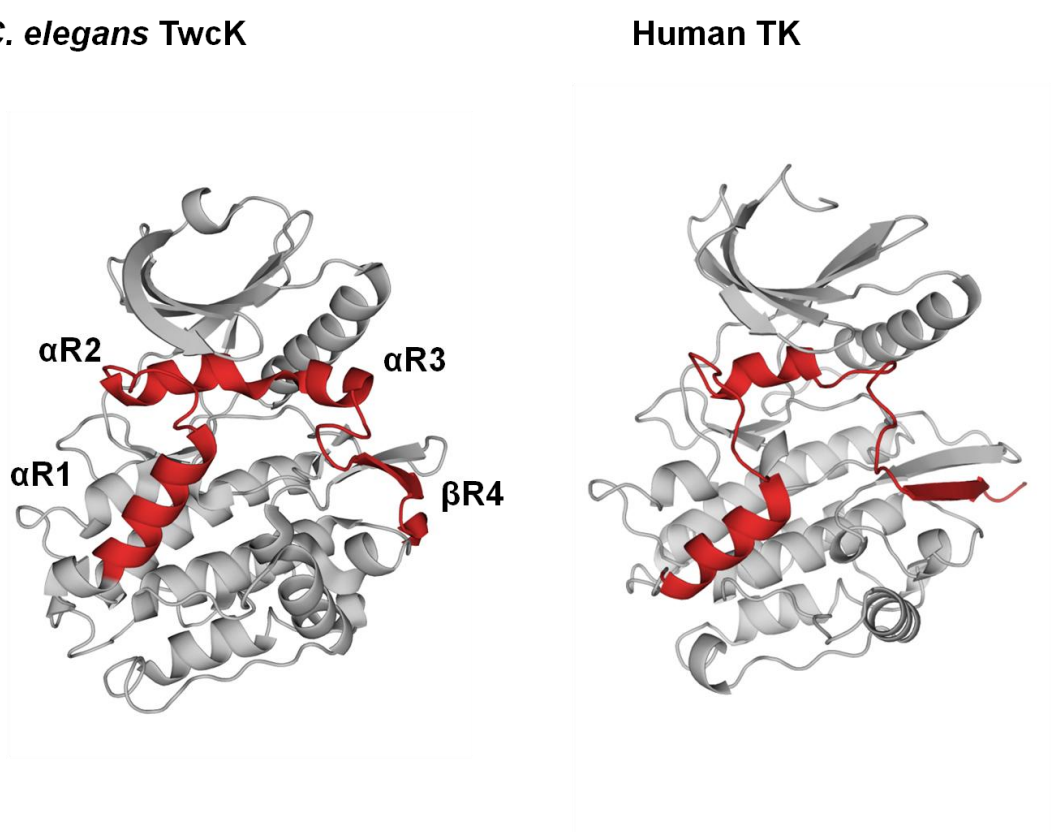


Fig 1.4.2 Comparison of the CRD in *C. elegans* TwcK and human TK. Twck (left) and TK (right). The kinase domain is shown in grey, with the CRD in red. Structural elements of the CRD are labelled for TwcK. PDB entries used: 3UTO (TwcKR) and 1TKI (TK).

With significant catalytic divergence amongst the titin-like family, including inactive kinases such as TK, the differing modes of mechanical activation may relate to the catalytic properties of the specific kinase domain. It could be envisaged that those catalytically active members of the family may undergo mechanical priming, with the CRD remaining bound to protect the kinase domain from mechanical deformation, thereby facilitating catalysis under stretch. For those kinases such as TK that lack catalytic activity, their role as a signalling scaffold may rely on baring the kinase domain for protein interactions, with a mechanical buffer to support catalysis not required.

Despite the differing catalytic properties, the highly conserved nature of the kinase locus of titin-like proteins hints at shared mechanistic principles within the titin-like kinase family. TwcK is currently the best characterised member of the titin-like kinase family and recombinant protein expression in *Escherichia coli* (*E. coli*) is well-established, along with the availability of crystal structures. TwcK was therefore chosen for the current mechanistic study and forms the basis of the work presented in this thesis.

1.5 Aims

This thesis aims to address outstanding gaps in the understanding of the catalytic principles of titin-like kinases and their autoregulation through both mechanical and biochemical means. Using TwcK as a representative of the titin-like family, the aims of the work presented here fall in to three main areas of study:

i) **TwcK regulation through autocatalysis**

While the understanding of intrasteric regulation in titin-like kinases is progressing, regulation by autocatalysis has not been studied. Here we investigate the capability of TwcK to undergo autophosphorylation as a means of regulating catalysis and how this fits with a potential mechano-chemical coupling of TwcK activity in the sarcomere.

ii) **Structural characterisation of the non-inhibited catalytic domain of TwcK in its active conformation and its substrate specificity**

All crystal structures of titin-like kinases to-date show the kinase domains in an inactive conformation, autoinhibited by their CRD and in the case of the TwcKR, both the NL and CRD autoinhibitory tails. There is currently no structural information available regarding the active conformation of titin-like kinases. Therefore, a major aim of this work is to obtain a crystal structure of the non-inhibited catalytic domain bound to a substrate or inhibitor that can reveal the activated state of TwcK . The active conformation of TwcK can provide valuable information on the conformational changes required for TwcK catalysis and by extension, help to further understand the role of the autoinhibitory tails in maintaining an inactive conformation.

iii) The mechanical properties of TwcK and the role of the autoinhibitory tails in stretch activation

For titin-like kinases to undergo mechanoactivation, these kinases must be capable of operating under stretch. We aim to explore the mechanical basis of activation in TwcK through stretching of the NL autoinhibitory tail using TwcK:DNA chimeras. This involves the attachment of a DNA spring to TwcK as a means of introducing mechanical tension. Crucially, we aim to address the question of whether the TwcK active site is still able to perform catalysis efficiently under stretch-induced mechanical deformation, a key premise of mechanoactivation in titin-like kinases.

Chapter 2

Twitchin kinase inhibition by autophosphorylation

This chapter explores the autocatalytic properties of twitchin kinase. We find that twitchin kinase undergoes autophosphorylation, leading to inhibition of catalysis. Four major sites of modification have been identified, with phosphorylation occuring in important mechanistic regions of the catalytic domain, suggestive of an additional mechanism for regulation of twitchin kinase activity.

2.1 Introduction

2.1.1 The protein kinase fold

Eukaryotic protein kinases consist of two lobes, a smaller N-terminal lobe (N-lobe) composed primarily of β-sheets and a larger C-terminal lobe (C-lobe), made up of α-helices. The active site cleft is formed between the N- and C- lobes and is the site of transfer of the γ phosphate of ATP to a phosphorylatable serine, threonine or tyrosine residue in the polypeptidic substrate. The protein kinase fold contains a number of highly conserved elements that are essential for catalysis. These are mapped onto the crystal structure of autoinhibited TwcK **(Fig 2.1.1.1)** and explained below. As TwcK is a serine/threonine kinase, this section will not discuss features that are applicable only to tyrosine kinases.

The N-lobe contains three major features that are responsible for nucleotide coordination during phosphotransfer. The glycine-rich loop is formed between the β1 and β2 strands of the N-lobe and helps to position the γ-phosphate of ATP for

phosphotransfer (Hemmer *et al.,* 1997; McNamara *et al.,* 2011). In the β3 strand, a lysine residue of the AxK motif (K185 in TwcK) chelates the β and γ phosphate groups of bound ATP and is essential for catalysis (Gibbs *et al.,* 1991; Carrera *et al.,* 1993; Iyver *et al.,* 2005). Mutation of the β3 lysine is a well-established means of abolishing catalysis whilst leaving the kinase capable of ATP binding (Carrera *et al.,* 1993; Iyver *et al.,* 2005). In the αC-helix, a central glutamate (E201 in TwcK) forms a salt-bridge with the β3 lysine that is essential for a correctly positioned αC-helix and subsequently an ordered active site (Huse & Kriyan, 2002; Taylor & Kornev, 2011).

The C-lobe forms the platform for peptide substrate binding and contains the majority of the conserved active site residues and regulatory elements. The catalytic loop contains the HxD motif, within which is the catalytic aspartate (D277 in TwcK) that acts as a base for deprotonation of the phosphorylatable side chain in the peptide substrate (Kenyon *et al.,* 2012). The DFG (Asp-Phe-Gly) motif of protein kinases (residues 297-299 in TwcK) is found in the Mg binding loop and is responsible for coordinating ATP, most commonly via bound Mg^{2+} ions (Knighton *et al.,* 1991).

The activation loop immediately follows the Mg binding loop. The activation loop is the most common site of phosphorylation-mediated activation of catalysis (Dorey *et al.,* 2001; Xu *et al.,* 2004; Roskoski Jr., 2004; Steichen *et al.,* 2012), with rearrangement of the activation loop facilitating correct positioning of the Mg binding loop and αC helix as well contributing to a favourable environment for peptide substrate binding (Nolen *et al.,* 2003; Steichen *et al.,* 2010). Following the activation loop, the P+1 loop acts as a platform for peptide substrate binding by positioning the hydroxyl group of the phosphorylatable substrate residue in close proximity to the catalytic aspartate (Johnson *et al.,* 2001; Endicott *et al.,* 2012).

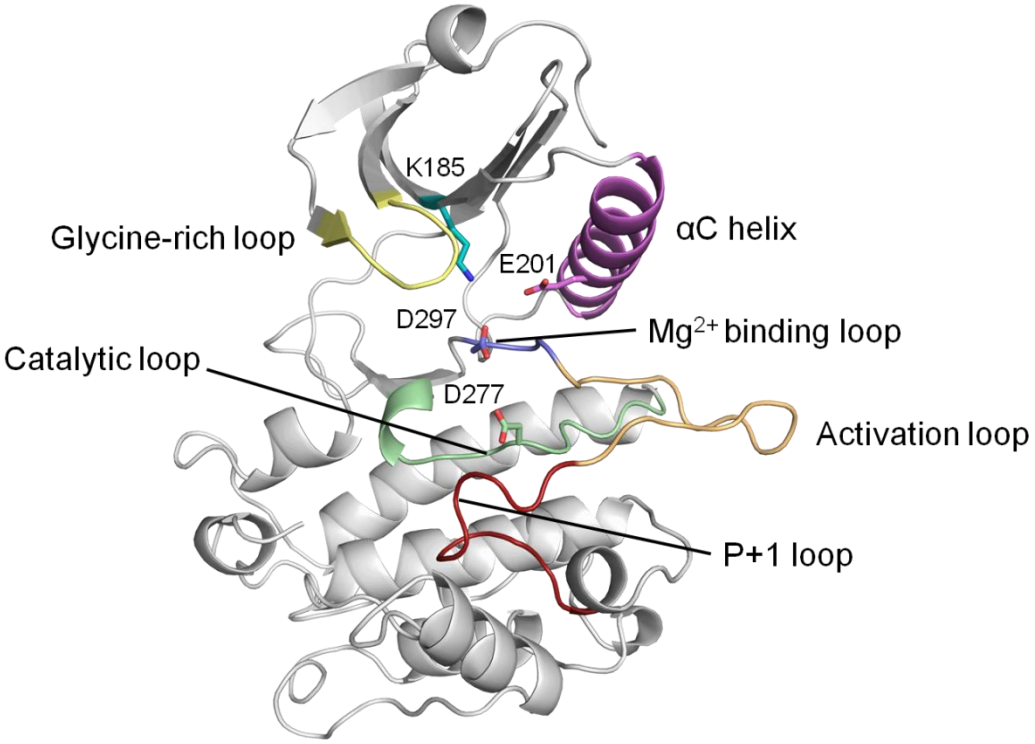


Fig 2.1.1.1 Key mechanistic regions of the protein kinase fold. Regions are mapped to the catalytic domain of TwcK. The crystal structure of TwcKR (PDB: 3UTO) is shown with regulatory tails removed.

Protein kinases catalyse a bi-substrate reaction and therefore contain binding sites for both ATP and for a peptide substrate. ATP is bound in the active site cleft between the N- and C-lobes. The ATP-binding pocket of protein kinases is formed by up to 38 residues (Vulpetti & Bosotti, 2004) that form favourable binding environments for each of the constituent ATP moieties; adenine, ribose and phosphate groups **(Fig 2.1.1.2).** The adenine ring of ATP makes non-polar interactions with hydrophobic side chains of residues from both the N- and C-lobes and hydrogen-bonds with the backbone of the interlobular hinge region (Zheng *et al.*, 1993; Vulpetti & Bosotti, 2004; McNamara *et al.*, 2009). A conserved polar residue at the N-terminus of the αD helix forms a hydrogen bond with a hydroxyl

group of the ribose moiety and the phosphate groups of ATP are coordinated the β3-lysine and αC glutamate as well as by Mg^{2+} ions bound by the DFG aspartate (Gibbs *et al.*, 1991; Carrera *et al.*, 1993; Iyver *et al.*, 2005; Huang *et al.*, 2012). The glycine-rich loop is also involved in positioning the **γ**-phosphate of ATP for phosphotransfer via backbone interactions (Hemmer *et al.*, 1997; McNamara *et al.*, 2011).

While the sequence conservation of the residues directly involved in ATP interactions is high, structural and sequence variation are observed in the solvent accessible opening of the ATP-binding pocket as well as in a buried region that sits deep within the cleft between the N- and C-lobes. The variability of these areas of the ATP-binding pocket has been exploited in work to produce ATP competitive, selective protein kinase inhibitors (reviewed by Müller *et al.*, 2015).

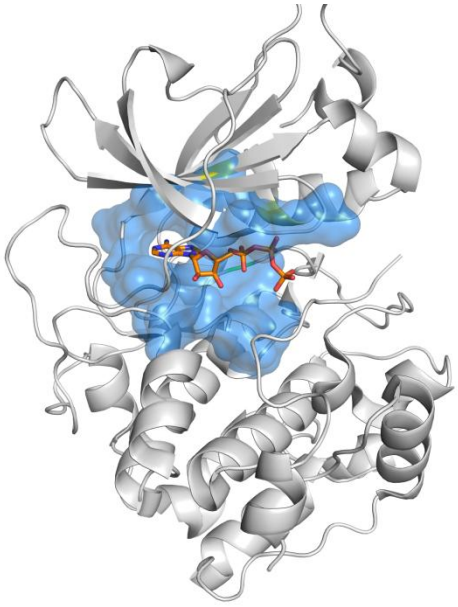


Fig 2.1.1.2 The kinase ATP-binding pocket. PKA (PDB: 4IAC) shown as surface representation with bound ATP analogue (AMP-PCP). Residues contributing to nucleotide binding are located in the N-lobe, C-lobe and interlobular hinge region.

Crystal structures of protein kinases in complex with a peptide substrate have been solved for several Ser/Thr kinases including PKA (Knighton *et al.,* 1991; Bossemeyer *et al.,* 1993; Gerlits *et al.,* 2013), protein kinase C (PKC) (Wang *et al.,* 2012), phosphorylase kinase (Lowe *et al.,* 1997) and cyclin-dependent kinase 2 (Cdk2) (Cook *et al.,* 2002), revealing that peptide substrates bind to the C-lobe in a largely extended conformation. The phosphorylatable residue (P0) is orientated to face directly in to the active site cleft where it interacts with the catalytic aspartate of the HxD motif (Kenyon *et al.,* 2012). Substrate recognition requires residues either side of P0, with free amino acids being a poor substrate for protein kinases (Adams, 2001). The substrate residues are sequentially numbered according to their position relative to P0, with P-1, P-2, P-3, etc. being N-terminal to P0 and P+1, P+2, P+3, etc. being C-terminal to P0.

Substrate specificity is conferred by a number of binding pockets that interact with the peptide substrate at residues close to the phosphorylation site. These binding pockets contribute to substrate specificity by favouring certain amino acid side chains at specific sites close to P0 (Pinna & Ruzzene, 1996; Kobe *et al.,* 2005; Ben-Shimon & Niv, 2011). For example, Ser/Thr kinases commonly require basic residues N-terminal to the phosphorylation site, with arginine frequently seen at the P-2, P-3 and P-5 positions in the peptide substrate. The preference for arginine in these positions has been mapped to a number of residues that form binding pockets (Zhu *et al.,* 2005; Ben-Shimon & Niv, 2011). A good example of this is seen in the substrate bound crystal structure of PKA (Gerlits *et al.,* 2013), showing coordination of the arginine at P-2 by highly conserved Glu170 and Glu230 residues **(Fig 2.1.1.3).** These residues form part of a P-2/P-5 pocket, that confers specificity for arginine at these positions in the peptide substrate (Ben-Shimon and Niv, 2011).

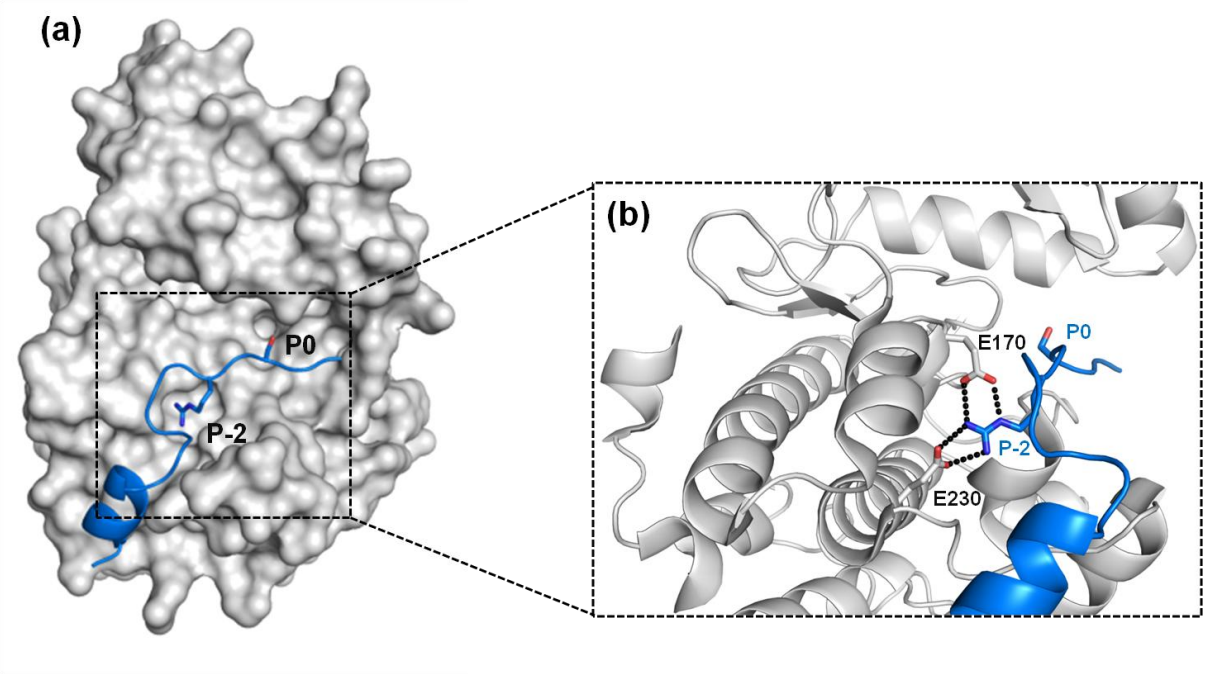


Fig 2.1.1.3 Kinase substrate binding pockets. Substrate specificity in protein kinases is conferred by binding pockets that accommodate specific residues at certain positions within the peptide substrate. **(a)** crystal structure of PKA (white) in complex with peptide substrate (blue) (PDB: 4IAC) with P0 and P-2 substrate residues shown as sticks. **(b)** close-up view of the P-2 binding pocket of PKA, with residues E170 and E230 hydrogen bonding with the P-2 arginine of the peptide substrate

2.1.2 Regulation of protein kinase activity

Protein kinases are key components of signalling pathways and act as signal transducers in the majority of eukaryotic cellular processes (Manning *et al.*, 2002). Protein kinases are involved in both intra- and inter-cellular processes, including cell growth, proliferation, differentiation and apoptosis (Zhang & Liu, 2002), transcription (Davis, 1995), cell cycle control (Pines, 1994) and rearrangement of the cytoskeleton (Leung *et al.*, 1996; Manning *et al.*, 2002). Aberrant protein kinase activity is associated with a multitude of disorders (Cohen, 2001), with perturbation of kinase-mediated signalling pathways a major feature of many cancers (Blume-

Jensen & Hunter, 2001). Strict spatial and temporal regulation of protein kinase activity is therefore critical. This regulation can be carried out in a number of ways and the mechanisms of control can be either activating or inhibitiory, employing phosphorylation, protein binding partners or autoinhibition (Endicott *et al.,* 2012). Several important examples of protein kinase regulation are discussed below:

Cyclin-dependent kinases (Cdks) are involved in control of the cell cycle and are activated through a two-step process, involving both a protein binding partner and phosphorylation by an upstream kinase as demonstrated for Cdk2 **(Fig 2.1.2.1)** (Pavletich, 1999). Prior to activation, Cdk2 is autoinhibited, with the catalytic aspartate positioned outside the active site and the activation loop occupying the active site cleft, blocking peptide substrate binding (Jeffrey *et al.,* 1995). A cyclin A subunit binds to Cdk2, inducing conformational changes in both the catalytic loop and the activation loop which moves the catalytic aspartate into a catalytically productive position and largely relieves the blockage of the catalytic cleft by the activation loop, leading to partial activation of Cdk2 (Jeffrey *et al.,* 1995). Full activation is then achieved by phosphorylation of a threonine residue within the activation loop of Cdk2 by Cdk-activating kinase (CAK) (Fisher & Morgan, 1994), with coordination of the phosphate group from arginine residues in the N-lobe, C-lobe and activation loop completing the reorganisation of the substrate binding site and leading to full activation (Pavletich, 1999).

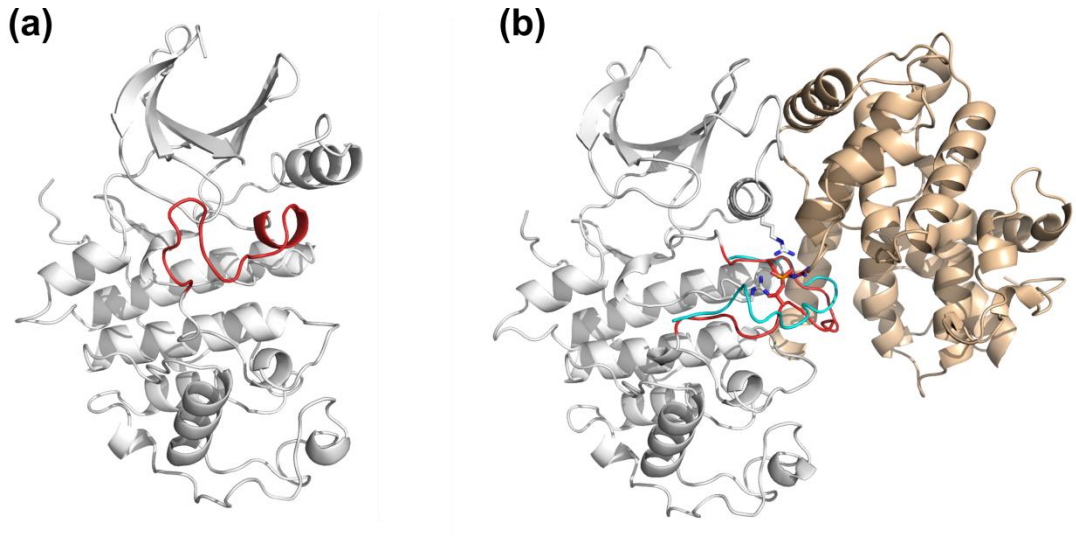


Fig 2.1.2.1 Two-step activation of Cdk2. (a) Inactive Cdk2 (grey) with activation loop (red) occupying the active site cleft. **(b)** Fully active Cdk2 (grey) in complex with cyclin A (beige) with phosphorylated activation loop (red). The unphosphorylated activation loop of partially activated Cdk2 (cyan) is superimposed on the fully active Cdk2 structure. PDB entries used: 1HCL (inactive), 1FIN (partially active) and 1JST (fully active).

The most frequently observed protein kinase activation mechanism involves phosphorylation within the activation loop. This leads to a local refolding and subsequent rearrangement of the activation segment, facilitating catalytic activity (Johnson & Lewis, 2001). Activation loop phosphorylation can occur by autophosphorylation (*cis* or *trans*) or performed by another kinase. The mechanism by which activation loop phosphorylation leads to kinase activation varies. The activation loop is often disordered in its unphosphorylated form or can adopt a conformation that occupies the active site and blocks catalysis (Bayliss *et al.*, 2015). Activation loop phosphorylation most commonly leads to changes in conformation of the active site, allowing the magnesium binding loop and αC helix to be correctly positioned for catalysis as well as creating a favourable environment for substrate binding (Johnson & Lewis, 2001; Nolen *et al.*, 2003; Steichen *et al.*, 2010).

Phosphorylation of at least one site within the activation loop is required for phosphotransfer in numerous Ser/Thr kinase including PKA **(Fig 2.1.2.2)** (Steichen *et al.*, 2012) and protein kinase C (PKC) (Xu *et al.*, 2004) as well as tyrosine kinases such as c-Abl (Dorey *et al.*, 2001) and c-Src (Roskoski Jr., 2004).

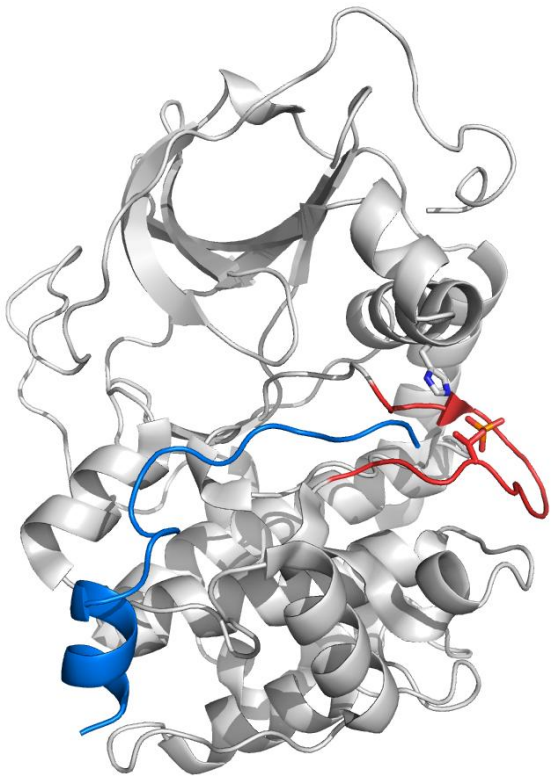


Fig 2.1.2.2 The phosphorylated activation loop of PKA. Crystal structure of PKA (white) with phosphorylated activation loop (red). The phosphorylated threonine residue in the activation loop is coordinated by a histidine residue (both shown as sticks) from the αC-helix. The phosphorylated activation loop forms a binding platform for the peptide substrate (blue). PDB entry used: 4IAF. In the unphosphorylated crystal structure of PKA, the activation loop is disordered and therefore not visible (PDB entry: 4DFY).

Titin-like kinases are regulated intrasterically by N- and C-terminal tails that pack against the catalytic kinase domain. This structural principle was first revealed by early crystal structures of human TK (Mayans *et al.*, 1998; Bogomolovas *et al.*, 2014) and TwcK from *C. elegans* (Hu *et al.*, 1994; Kobe *et al.*, 1996; von Castelmur *et al.*, 2012) and *Aplysia* (Kobe *et al.*, 1996) that showed the C-terminal regulatory domain (CRD) bound between the two kinase lobes, blocking the ATP and peptide substrate binding sites on the kinase. The concept was then extended by the more recent structure of *C. elegans* TwcK region that also included neighbouring domains flanking the kinase N- and C-terminally (von Castelmur *et al.*, 2012). This showed that the N-terminal linker (NL) wraps around the upper kinase lobe, occupying much of the interlobular kinase hinge region. Both NL and CRD regions were shown to be inhibitory, with the isolated catalytic domain of TwcK displaying maximal catalysis (von Castelmur *et al.*, 2012). It has been proposed that this dual instrateric inhibition is released in the active muscle by stretch-unfolding of the NL and CRD tails, which are pulled away from the catalytic kinase domain during the mechanical cycle of the sarcomere (Gräter *et al.*, 2005; Puchner *et al.*, 2008; von Castelmur *et al.*, 2012).

In contrast to the existent knowledge on titin-like kinase instrateric regulation, little is known on the autocatalysis of these kinases, which remains largely unstudied. Early work on TwcK catalysis showed that TwcK can perform low levels of phosphorylation (Lei *et al.*, 1994) at residues that were later mapped to the NL (von Castelmur *et al.*, 2012). Whether this autophosphorylation is functionally relevant however remains to be explored. Here, we investigate the capability of the catalytic domain of TwcK to autoregulate its phosphotransfer activity through phosphorylation. Our findings indicate that TwcK employs high-

levels of autophosphorylation to effectively inhibit its catalysis, so that inhibition occurs even in the absence of flanking regulatory domains. We discuss the possible implications of these findings in terms of the proposed role of TwcK in mechanotransduction of signalling in the sarcomere.

2.2 Methods

2.2.1 TwcK wild-type constructs

TwcK constructs **(Fig 2.2.1.1)** consisting of TwcKR, catalytic domain (Kin), NL-Kin-CRD, NL-Kin (all in the pETM-11 vector) and Kin-CRD (pETM-13 vector) were provided by Julius Bogomolovas and cloned via NcoI and KpnI restriction sites. Ligation in to pETM-11 plasmids yields a protein product containing an N-terminal His_6 affinity followed by a linker sequence and Tobacco Etch Virus (TEV) protease cleavage site for His_6 tag removal. Protein products from the pETM-13 vector possess only a C-terminal, non-cleavable His_6 affinity tag. Both pETM-11 and pETM-13 confer antibiotic resistance to kanamycin.

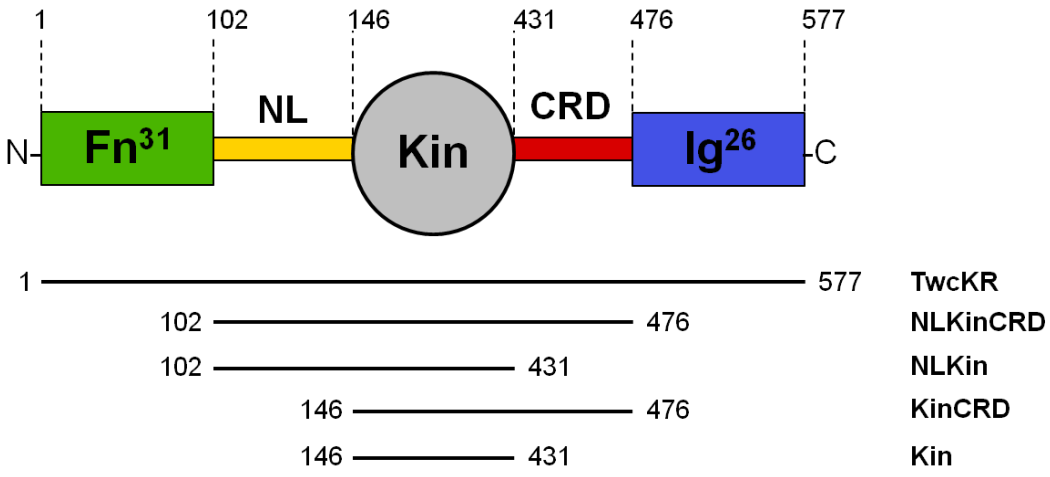


Fig 2.2.1.1 TwcK constructs and domain boundaries. Residue numbering is based on the crystal structure of TwcKR (PDB: 3UTO).

2.2.2 Generation of phosphorylation site point mutants by site-directed mutagenesis

Point mutations were introduced into the catalytic domain of TwcK (Kin) in order to generate the following constructs: Kin^{T212A}, Kin^{S290A}, Kin^{T301A}, Kin^{T316A}, Kin^{T401A}, Kin^{T412A}, Kin^{T212D}, Kin^{T316D} and Kin^{T401D}. Plasmid amplification was performed by PCR using Q5 DNA polymerase (NEB) with back-to-back primers containing the desired mutation in the forward primer only. A list of all primers used to generate Kin phosphorylation site mutants is given in **Table 2.2.2.1.** PCR amplification of the plasmid template was confirmed by agarose (1%) gel electrophoresis. Digestion of wild-type template DNA along with phosphorylation and linearisation of mutated DNA product was performed in a single reaction, containing 1 µL PCR product, 0.5 µL DpnI (NEB), 0.5 µL T4 polynucleotide kinase (NEB), 0.5 µL T4 DNA ligase (NEB), 1 µL 10x T4 DNA ligase buffer (NEB) and 6.5 µL ddH_2O. Incubation was for 30 minutes at 37^oC before transformation of the entire reaction into *E.coli* DH5-α. All mutations were confirmed by sequencing (Source Bioscience). The Kin^{S234A} mutant was generated by Quikchange mutagenesis, following the standard protocol provided by Agilent.

Table 2.2.2.1 List of mutagenic primers used for generation of Kin phosphorylation site mutants. The site of mutagenesis is underlined.

Primer Name		Sequence (5'-3')
T212A	forward	gagacacccg**g**cgcttgtgaa
	reverse	aaaactgacatggtttgaatctcttttc
S290A	forward	aacgaagaga**gc**caacgagctgaaattg
	reverse	gtgaacatgatgttctctg
T301A	forward	ttttggactc**g**ccgcccatct
	reverse	tcaatcaatttcagctcgttgc
T316A	forward	tacaacagga**g**ctgcccaatttg
	reverse	accttgacggattgttttg
T401A	forward	taccagaatg**g**ctattcatcaag
	reverse	tttggatcagcaagaagaag
T412A	forward	cccatggttg**g**caccaggaaa
	reverse	tgctccaaagcttgatgaatag
T212D	forward	gagacacccg**gat**cttgtgaa
T301D	forward	ttttggactc**ga**cgcccatct
T316D	forward	tacaacagga**ga**tgcccaatttg
T401D	forward	taccagaatg**ga**tattcatcaag
S234A	sense	caagtctgtaagggactc**a**gtcatatgcatgagaaca
	anti	tgttctcatgcatatgac**t**gagtcccttacagacttg

2.2.3 TwcK recombinant protein expression and purification

Expression and purification of various TwcK constructs was previously established in von Castelmur *et al*., 2012. All TwcK constructs were expressed in *E. coli* Rosetta (DE3) (Merck Millipore). Cultures were grown in LB media supplemented with 25 µg/mL kanamycin and 34 µg/mL chloramphenicol at 37^{o}C to an OD_{600} of 0.6-0.8. Protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by growth at 18^{o}C for approximately 18 hours. Cell pellets were harvested by centrifugation at 5000 *g* and stored at -80^{o}C until further use.

Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, pH 7.9) containing 20 µg/mL DNase I (Sigma Aldrich) and one complete EDTA-free protease inhibitor tablet (Roche) per litre of cell culture. Cell lysis was by sonication on ice followed by clarification of lysate by centrifugation at 39000 *g* for 45 minutes. Cell lysate was syringe filtered (0.22µm) and applied to a gravity flow column containing Ni^{2+}-NTA resin (Qiagen) equilibrated in lysis buffer containing 20 mM imidazole. Ni^{2+}-NTA resin was washed with 10 column volumes of lysis buffer containing increasing concentrations of imidazole and elution of bound sample was with 10 mL of lysis buffer containing 200 mM imidazole. Removal of imidazole from eluted samples was by buffer exchange into lysis buffer using PD-10 desalting columns (GE Healthcare). His_6-tag removal by His_6-tagged TEV protease cleavage was carried out overnight at 4^{o}C, followed by subtractive Ni^{2+} affinity purification, using Ni^{2+}-NTA resin equilibrated in lysis buffer.

Size exclusion chromatography (SEC) was performed on a Superdex 200 16/60 column (GE Healthcare) in 50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH

7.9. The resulting fractions were analysed by SDS-PAGE (12% acrylamide) and stored at 4^{o}C until further use. Staining for SDS-PAGE was by Coomassie in all cases. All protein concentrations were measured by absorbance at 280 nm using a nanodrop instrument (Thermo Fisher).

2.2.4 Measurement of TwcK catalytic activity by [γ^{32}P]-ATP phosphotransfer assay

Phosphotransfer assays were adapted from protocols published by Hastie *et al.,* 2006. All phosphotransfer assays were set up as 40 µL reactions containing 20 mM Tris-HCl, 10 mM $MgCl_2$, 0.2 mg/mL bovine serum albumin (BSA), 0.2 mg/mL peptide substrate and 200 µM [γ^{32}P] ATP (2 µCi per reaction) (Sigma Aldrich). Each reaction contained 30 ng of TwcK protein. For all assays, the peptide substrate used was derived from chicken gizzard myosin light chain (kMLC11-23) with amino acid sequence KKRARAATSNVFS as previously reported in Heierhorst *et al.,* 1996b and von Castelmur *et al.,* 2012. Peptide substrate was purchased from Biomatik (Delaware, USA) as a lyophilised trifluoroacetate (TFA) salt, purified by HPLC to >95% purity and was prepared by dissolving in 50 mM Tris, 50 mM NaCl, 1 mM DTT, pH 7.9 and final pH was adjusted to approx. pH 7.9 by addition of HCl.

Phosphotransfer assays were prepared before the experiment to contain all reaction components with the exception of ATP. Reactions were initiated by addition of 200 µM [γ^{32}P] ATP (2 µCi) and then briefly vortexed before incubation at 25^{o}C in a water bath for 20 minutes. Phosphotransfer reactions were halted by blotting on P81 phosphocellulose paper (Merck Millipore) and immediately submerged in 100 mM phosphoric acid. In order to remove residual [γ^{32}P] ATP, blotted reactions were subject to 4 washes with 100 mM phosphoric acid for a duration of 5 minutes per

wash. A final wash using acetone was performed and blotted reactions were then air-dried.

For each construct being assayed, background readings were measured from samples in which the peptide substrate was replaced with H_2O. Further negative controls containing peptide substrate with no kinase were also included in each experiment. All assays, background measurements and controls were carried out in duplicate. Quantification was performed by Čerenkov counting (Bem *et al.,* 1980) using a Beckman Coulter scintillation counter. Counting paramaters were as follows: nuclide 32P, count time of 1 minute per sample, coincidence time of 18 nsec. Duplicate samples containing a known quantity of [γ^{32}P] ATP were counted to allow calculation of molar quantities of ATP catalysed during phosphotransfer assays.

For plotting of phosphotransfer activity histograms (Microsoft Excel), counts for each assay were averaged and the average background count subtracted. Error bars were plotted by subtracting the standard deviation of background measurements from the standard deviation of assay measurements.

2.2.5 Dephosphorylation of twitchin kinase by lambda protein phosphatase

For dephosphorylation of TwcK, recombinantly produced TwcK samples were buffer exchanged following Ni^{2+} affinity purification into 50 mM Tris-HCl pH 7.9, 50 mM NaCl, 1 mM dithiothreitol, 1 mM $MnCl_2$ using PD-10 desalting columns (GE Healthcare) and incubated with 200-400 units of lambda protein phosphatase (NEB). Incubation with lambda protein phosphatase was carried out for 1-1.5 hours at room temperature before being addition of His_6-tagged TEV protease for His_6 affinity tag removal and incubation overnight at 4^{o}C. Further purification by

subtractive Ni^{2+} affinity followed by SEC was subsequently performed as described in section **2.2.3.**

2.2.6 Determination of TwcK phosphorylation state and identification of TwcK phosphorylation sites by mass spectrometry

All mass spectrometry was carried out by Dr. Mark Wilkinson (University of Liverpool, Institute of Integrative Biology).

To obtain an intact mass for the recombinant TwcK product, desalted TwcK was infused into the nano-electrospray source of a Waters Q-ToF micromass spectrometer at a flow rate of 50 µl/hour via a gas tight syringe. The positive ion mass spectrum of the sample was scanned in the range M/Z 80 to 2,000 using a scan time of 1 second and a data acquisition time of 5 minutes. MassLynx MaxEnt1 was then used to convert the summed multiply charged spectrum to a molecular mass spectrum.

In preparation for phosphorylation site identification, TwcK and its variants were digested in solution overnight in 25 mM Tris-HCl, 2 M urea, pH 7.8 using trypsin, Lys-C and Glu-C; either alone or in various protease combinations. MALDI-MS analysis was performed using a MALDI-Tof instrument (Waters-Micromass) on digests that had been enriched for phosphopeptides using either titanium oxide or Fe^{3+}-IMAC mini columns. Peptide mixtures were analysed using a saturated solution of alpha-cyano-4 hydroxycinnaminic acid in 50% acetonitrile/0.1% trifluoroacetic acid. Peptides were selected in the mass range of 1000 – 4000 Da. For ES-MS, the digests were desalted with C18 Zip Tips (Merck Millipore) and analysed as

described above for intact TwcK. Some individual peptides were subjected to ES-MS/MS for sequence analysis.

In-gel trypsin digests were also performed on TwcK samples fractionated by SDS-PAGE. Following an overnight incubation, the peptides were extracted from the gel pieces, desalted using C18 Zip Tips and analysed by MALDI-MS as described above.

2.2.7 Estimation of the effect of TwcK phosphorylation on ΔG of folding using FoldX

FoldX BuildModel was used to introduce phosphoryl-groups into the TwcK catalytic domain. The starting model was obtained using the TwcKR (PDB ID: 3UTO) trimmed of regulatory tails in order to represent the Kin construct used experimentally. FoldX Repair was then used to perform an energy minimisation on wild-type and phosphorylated structures (Schymkowitz *et al.*, 2005). The ΔΔG of folding (kcal/mol) compared to the unmodified structure was plotted as a histogram in Microsoft Excel.

2.3 Results

2.3.1 TwcK undergoes autophosphorylation

Chromatographic purification of the TwcK catalytic domain (Kin) was performed as described in **section 2.2.3**. SDS-PAGE analysis of purified TwcK revealed that a heterogeneous sample was obtained following size exclusion chromatography (SEC), with multiple bands that were only partially segregated **(Fig.2.3.1.1)**. These multiple bands seen by SDS-PAGE have an apparent molecular weight that is higher than the theoretical Kin molecular weight (32.6 KDa), with the species of lowest apparent molecular weight still appearing higher than the theoretical molecular weight. In an attempt to fully separate these multiple species, anion exchange chromatography (AEC) was performed following SEC. AEC elution profiles exhibited multiple distinct peaks, but SDS-PAGE revealed that the separation of species had not been achieved and that, in fact, SEC provided the best separation of species.

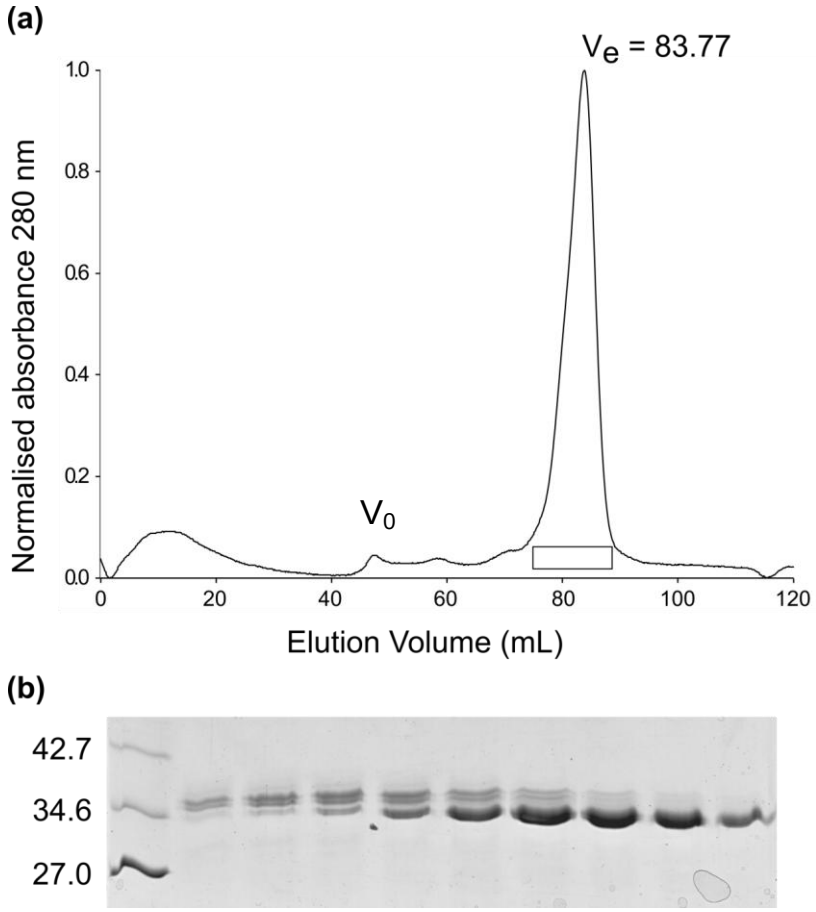


Fig 2.3.1.1 The recombinantly expressed catalytic domain of TwcK is chromatographically purified as multiple species. **(a)** SEC chromatogram of the TwcK catalytic domain (Kin). Exclusion volume (V_e) and column void volume (V_0) are indicated. Boxed area represents fractions for SDS-PAGE. **(b)** SDS-PAGE of Kin fractions following SEC. Molecular weight marker is shown to the left (values in kDa).

Various catalytically active TwcK constructs (von Castelmur *et al.*, 2012) were expressed and purified, revealing that the pattern of multiple species by SDS-PAGE was common throughout these active constructs, namely Kin, NLKin and KinCRD. Conversely, a fully autoinhibited TwcK construct, NLKinCRD, containing both N- and C-terminal autoinhibitory tails and shown to be inactive by phosphotransfer activity, was purified as an apparently single species following the

same purification protocol as for catalytically active TwcK constructs **(Fig 2.3.1.2)**. It was therefore concluded that the sample heterogeneity observed for TwcK was dependent on its catalytic activity.

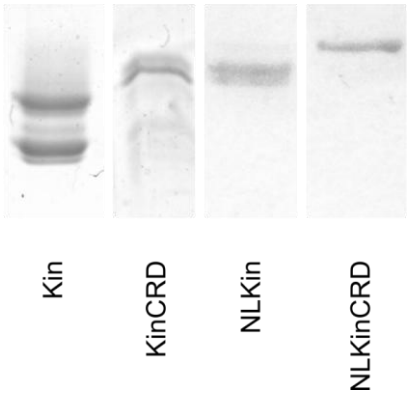


Fig 2.3.1.2 Catalytically active TwcK constructs are purified as heterogeneous samples. SDS-PAGE of SEC purified TwcK constructs, showing multiple species for catalytically active constructs (Kin, KinCRD, NLKin) and an apparently single species for the inactive NLKinCRD.

Due to the apparent dependence of TwcK heterogeneity on catalytic activity, it was hypothesised that phosphorylation was responsible, as a large number of protein kinases are known to undergo autophosphorylation (Lochhead, 2009). Electrospray-mass spectrometry (ES-MS) was performed in order to obtain an intact mass for purified Kin. ES-MS revealed three main masses; 32756, 32836 and 32916 Da **(Fig 2.3.1.3)**. The difference observed between the three main masses is approximately 80 Da, consistent with the molecular mass gained upon phosphorylation of serine or threonine residues. A number of smaller peaks at higher mass were also observed at mass intervals consistent with less abundant, higher phosphorylation states. Interestingly, the lowest mass observed (32756 Da) is in fact 152.4 Da larger than the predicted mass of Kin (32603.6). This therefore indicates

that the minimally phosphorylated species in recombinantly expressed Kin undergoes on average two phosphorylation events.

In order to confirm that the multiple Kin species observed by mass spectrometry were due to phosphorylation, purified wild-type Kin was subject to treatment by the non-specific, lambda protein phosphatase (lambda pp) prior to SEC purification. Treatment with lambda pp led to the collapse of the bands into a single species of lower apparent molecular weight, indicating that the sample heterogeneity observed in the untreated TwcK was indeed due to phosphorylation. Confirmation of the loss of heterogeneity in lambda pp treated Kin was obtained by ES-MS, showing a single main mass consistent with the theoretical molecular mass of unmodified Kin (32594.0 vs 32603.6 Da) **(Fig 2.3.1.3).**

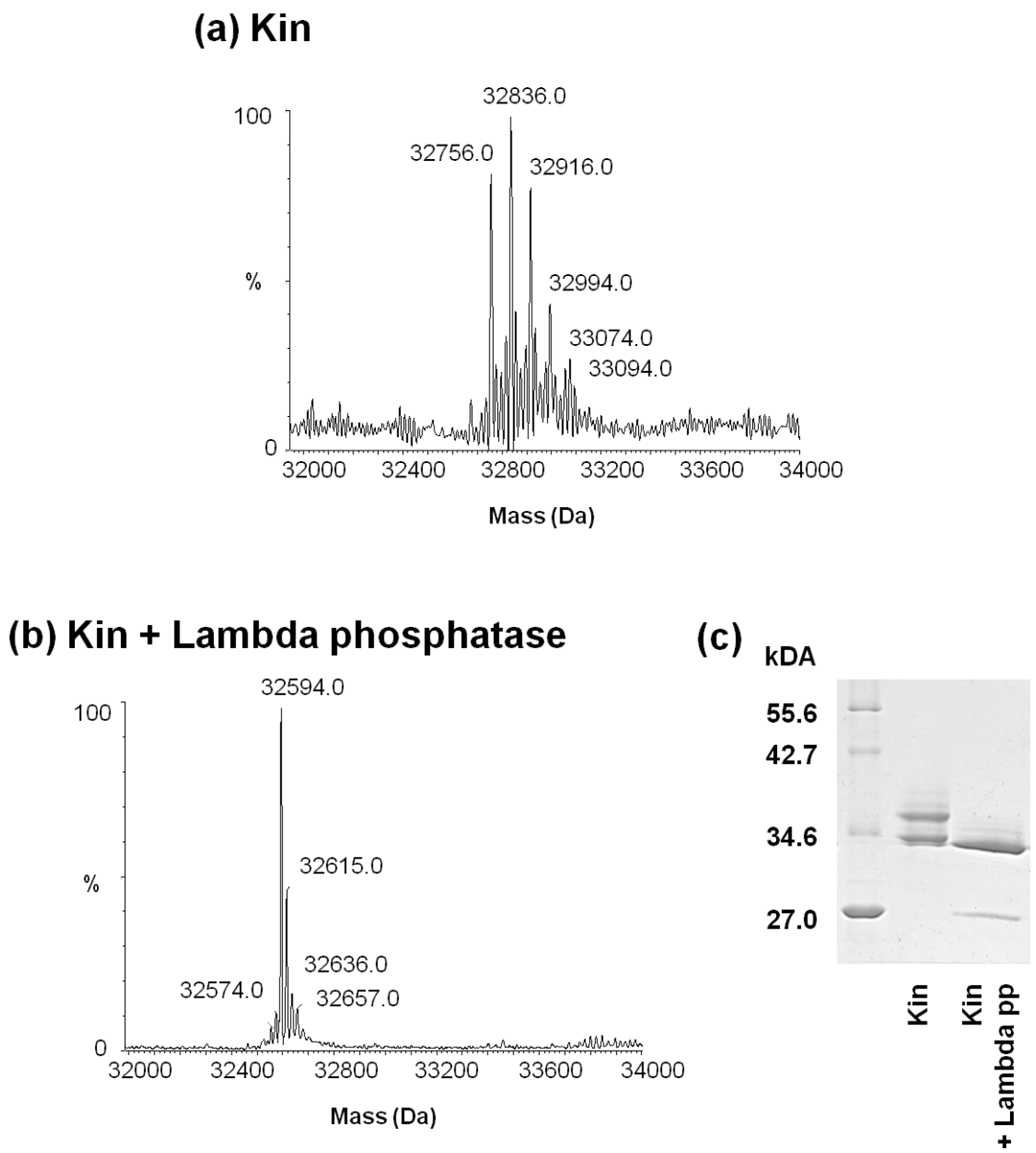


Fig 2.3.1.3 ES-MS intact mass reveals multiple phosphorylation states of TwcK. ES-MS intact mass spectra for **(a)** Kin and **(b)** Kin after Lambda phosphatase treatment. Untreated Kin shows a mixture of species, differing by the mass of a phosphoryl group or multiples thereof. All major species observed by ES-MS for untreated Kin are higher than the molecular mass of unmodified Kin calculated from amino acid sequence. Following Lambda phosphatase treatment, Kin ES-MS spectra consist of one major species corresponding to the theoretical molecular mass of unmodified Kin. Minor peaks of higher mass are likely due to incomplete Kin dephosphorylation. **(c)** SDS-PAGE showing the multiple species of purified Kin and collapse into a single species following Lambda phosphatase treatment (Kin Lambda pp). Molecular weight marker is shown adjacent.

To verify that the observed phosphorylation of TwcK was due to auto-phosphorylation, a Kin construct in which residue K185 was substituted for Alanine (Kin^{K185A}) was created. K185, a lysine residue in β3-strand, is a highly conserved residue in protein kinases located within the AxK motif. It is essential for phosphotransfer due to its role in the coordination of ATP, where it chelates the α- and β-phosphate groups. Substitution of this lysine residue to alanine leads to a catalytically inactive kinase. The use of such 'dead-kinase' mutants is well established in the protein kinase bibliography (Gibbs & Zoller 1991; Iyver *et al.,* 2005).

Kin^{K185A} was obtained by site-directed mutagenesis using the wild-type Kin construct as a template and subsequently recombinantly expressed and purified. Kin^{K185A} can be purified as a single species as judged by SDS-PAGE, with the apparent molecular weight of Kin^{K185A} being less than that of the lowest band of the wild-type Kin construct **(Fig 2.3.1.4)**. The apparent molecular weight of Kin^{K185A} as judged by SDS-PAGE is consistent with its theoretical mass (32603.6 Da) and runs lower than the wild-type Kin species of lowest apparent molecular weight but equivalent to lambda pp treated Kin.

Following expression and purification of Kin^{K185A}, phosphotransfer assays were performed to confirm the expected loss of catalysis due to the K185A point mutation. Phosphotransfer assays confirmed the abolition of phosphotransfer activity in Kin^{K185A} on the model peptide substrate, with only background levels of counts recorded for the experiment. Taking together the absence of heterogeneity in both lambda phosphatase treated and catalytically inactive Kin samples, can be concluded that the multiple observed species of TwcK are a result of autophosphorylation.

Results also indicate that bands of higher apparent molecular mass correspond to an increased phosphorylation state.

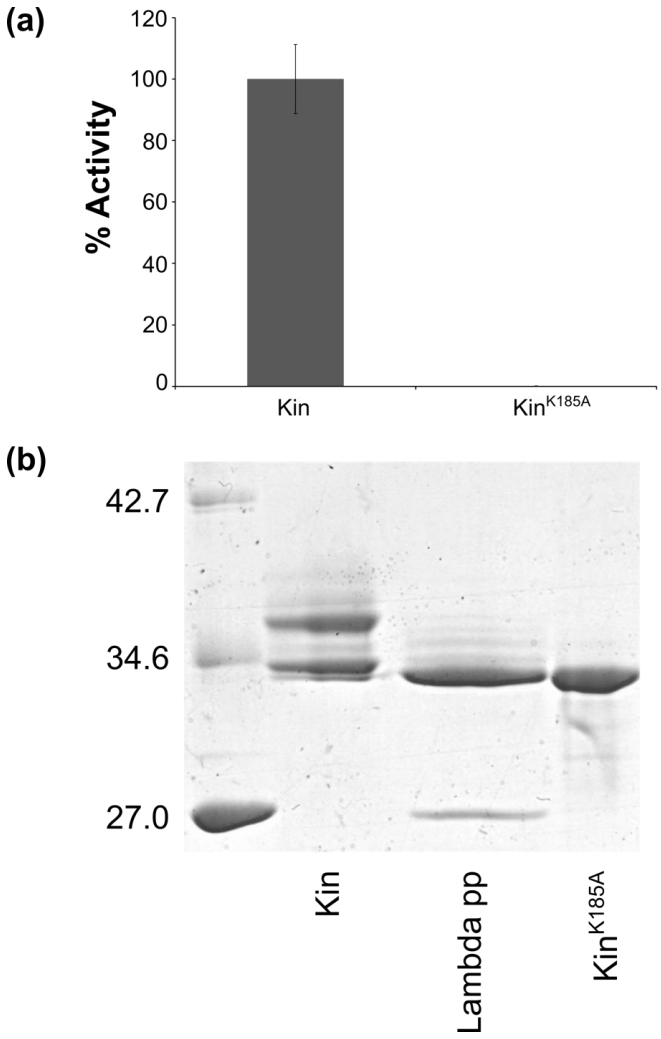


Fig 2.3.1.4 Kin^{K185A} is catalytically active and is equivalent in apparent molecular mass to dephosphorylated Kin. **(a)** Phosphotransfer assay for Kin (Hypo-P) and Kin^{K185A}, showing abolition of catalysis (<0.1%) following mutation of K185 to alanine. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate **(b)** SDS-PAGE comparison of Kin^{K185A} with wild-type Kin before and after Lambda phosphatase treatment, showing Kin^{K185A} runs as a single species with equivalent apparent molecular weight to dephosphorylated Kin (Lambda pp). Molecular weight marker is shown adjacent with values in kDa.

2.3.2 Hyperphosphorylation leads to the inactivation of TwcK

With TwcK confirmed to undergo phosphorylation, work was then directed to ascertain the effect of autophosphorylation on TwcK catalysis. Protein kinase phosphorylation has been the subject of extensive research, revealing phosphorylation as a widespread mode of regulation, acting by a variety of mechanisms (Cohen, 2000; Endicott *et al.,* 2012). The most frequently observed means of phosphorylation-mediated protein kinase regulation is through activating phosphorylation events in the activation loop. This most commonly leads to conformational changes in the active site, allowing the magnesium binding loop and αC helix to be correctly positioned for catalysis as well as creating a favourable environment for substrate binding (Nolen *et al.,* 2003; Steichen *et al.,* 2010).

To assess the differing levels of phosphorylation on catalysis, purified Kin species were partially segregated by SEC as shown in **Fig 2.3.1.1**. Three samples were classified based on the band composition seen by SDS-PAGE. These were 'Hypo-P' (consisting primarily of those species with bands of lowest apparent molecular mass, corresponding to minimally phosphorylated species), 'Hyper-P' (highest apparent molecular mass, maximally phosphorylated) and ‘50:50’ (an approximately equal proportion of all species). These partially segregated species were then subject to phosphotransfer assay. Strikingly, species with higher levels of phosphorylation were seen to have drastically reduced catalytic activity (approximately 5%) when compared to the minimally phosphorylated sample. Sample containing a roughly proportional mix of phosphorylation states showed catalytic activity at a level intermediate between the two **(Fig 2.3.2.1)**.

A catalytically active but partially autoinhibited TwcK construct, Kin-CRD was then purified and partially segregated into three major species as described for Kin. Phosphotransfer activity assays were performed in order to assess whether the pattern of phosphorylation mediated inhibition was consistent with that observed for Kin. Once again, minimally phosphorylated species displayed approximately 95% greater catalytic activity than those with the highest levels of catalysis **(Fig 2.3.2.1)**. The large reduction in catalysis in hyper-phosphorylated species demonstrates the ability of TwcK to negatively regulate catalytic activity by autophosphorylation.

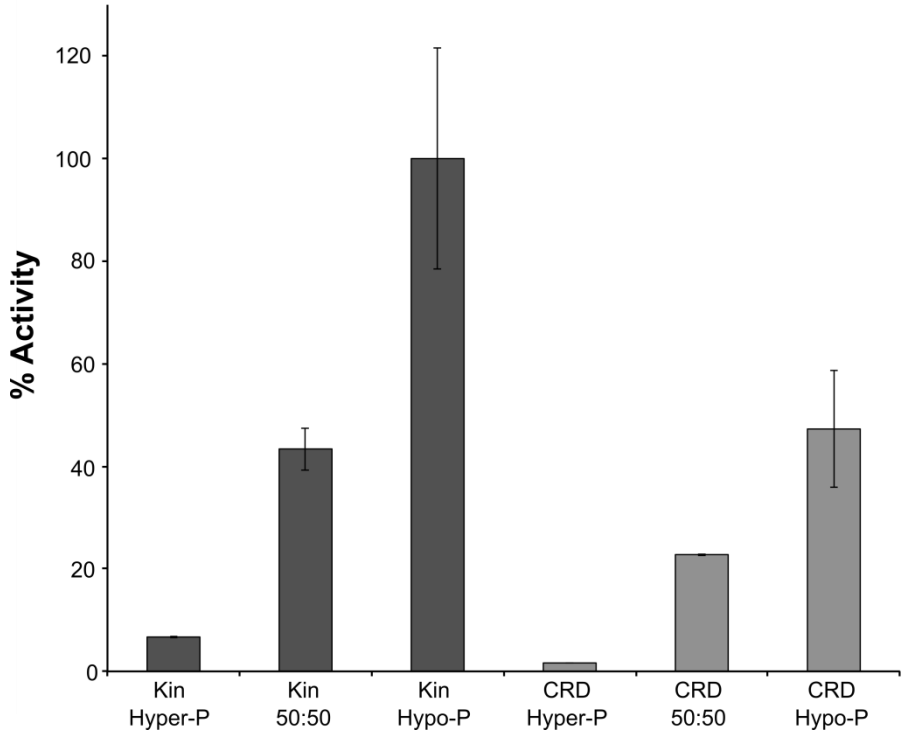


Fig 2.3.2.1 Hyper-phosphorylated TwcK species have dramatically reduced catalytic activity. Phosphotransfer assay on partially segregated Kin and KinCRD (CRD) species, showing low levels of catalysis for hyper-phosphorylated (Hyper-P) species and highest activity for hypo-phosphorylated (Hypo-P) species. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

2.3.3 Phosphorylation of TwcK is not required for activation

Mass spectrometry indicated that the most active, minimally phosphorylated species of purified Kin is likely to be doubly phosphorylated. Lambda phosphatase treated Kin was subject to phosphotransfer assays in order to determine the levels of Kin catalysis in the absence of phosphorylation. Prior to assays, removal of phosphate groups was again confirmed by ES-MS, providing a single major mass corresponding to unmodified Kin. For phosphotransfer assays, minimally phosphorylated Kin ('Upper') was used as a positive control for the assay and designated a value of 100% activity with Kin^{K185A} included as a negative control.

Phosphotransfer assays revealed that Lambda pp treated Kin shows an increase in activity of approximately 35% compared to the most active untreated Kin sample **(Fig 2.3.3.1)**. Although Lambda phosphatase treated Kin exhibits higher activity when compared to the untreated Kin sample, it must be considered that reporting of the activity of untreated Kin may be an underestimate due to the inability to completely segregate the various phosphorylation states. The activity of dephosphorylated Kin being equal to or higher than any of the phosphorylated Kin species shows that all observed phosphorylation events are either neutral or inhibitory with regards to catalytic activity rather than activating.

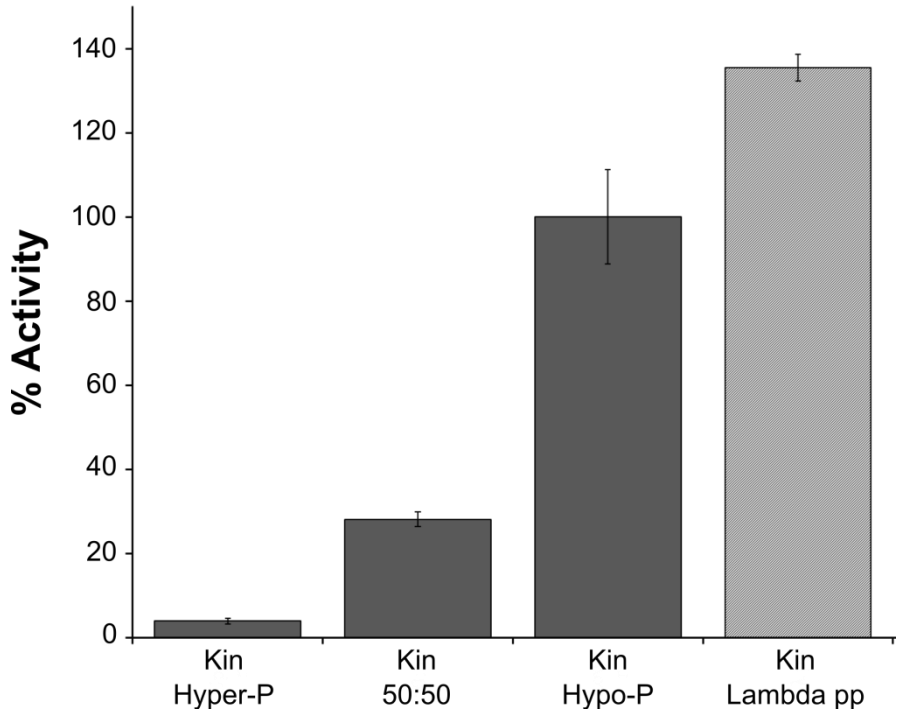


Fig 2.3.3.1 Dephosphorylated TwcK has higher catalytic activity than minimally phosphorylated species. Phosphotransfer assay on partially segregated Kin species and dephosphorylated Kin (Lamda pp) Minimally phosphorylated Kin (Hypo-P) was designated as 100% activity, with dephosphorylated Kin showing ~35% increased activity. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

2.3.4 Identification of multiple auto-phosphorylation sites in TwcK by mass spectrometry

Following confirmation of TwcK autophosphorylation, purified Kin samples were subject to proteolytic cleavage with Trypsin and/or Glu-C, followed by enrichment for phosphopeptides. MALDI-MS was then performed on the resulting proteolytic fragments in order to identify phosphorylated peptides. Due to the incomplete phosphorylation of Kin samples, phosphopeptides were seen as pairs of peaks differing by an *m/z* value corresponding to one or more phosphorylation events. Initial MALDI-MS experiments showed pairs of peaks for the peptides S290-

D305 and S290-K307 (missed cleavage) in which the peptide mass was increased by 80 and 160 Da compared to their predicted mass. With the only phosphorylatable residues within this sequence being S290 and T301, it was apparent that both S290 and T301 can be subject to phosphorylation. Another peptide pair was subsequently identified corresponding to L295-K307 with a mass increase of 80 Da. With the peptide L295-K307 containing only a single phosphorylatable residue, T301, it was thereby confirmed that T301 undergoes autophosphorylation. Phosphorylation of S290 was commonly not detected during subsequent MALDI-MS experiments, leading to the conclusion that S290 is able to be phosphorylated but is likely a secondary site of low occupancy. T212 was also identified as a phosphorylation site from the presence of a peak with mass equivalent to the peptide H210-E225, containing only T212 as a phosphorylatable residue within this sequence.

Following initial MALDI-MS on Trypsin/Glu-C digested Kin samples, Kin samples were subject to digestion with Trypsin and Lys-C. ES-MS analysis on the resulting proteolytic fragments revealed three pairs of peaks, corresponding to peptides H210-K241 [M^+3H^+], T313-R363 [M^+4H^+] and M400-R419 [M^+2H^+] along with a corresponding singly phosphorylated peak for each **(Fig 2.3.4.1).** The H210-K241 peptide contains the previously identified H210 residue, helping to confirm T210 as a site of TwcK phosphorylation, although the H210-K241 peptide also includes S234 as a potential site of modification. The identity of phosphorylation sites in peptides T313-R363 and M400-R419 were more ambiguous as the peptide masses contained more than one phosphorylatable residue. These are T313, T14 and T316 for the peptide T313-R363 and T401 and T412 for the peptide M400-R419. Mass spectrometry therefore lead us to propose major TwcK phosphorylation sites at T212, T301, one of T313, T314 or T316 and T401 or T412 as well as a minor site at

S290.

The catalytically inactive Kin^{K185A} mutant was also purified and subject to analysis by mass spectrometry in order to confirm that the peptide pairs observed for the wild-type Kin were due to autophosphorylation. For each pair of peaks observed in the wild-type Kin, only those corresponding to the non-modified peptides were observed for Kin^{K185A}, confirming that the peptide pairs seen for wild-type Kin arose from phosphorylated and non-phosphorylated species.

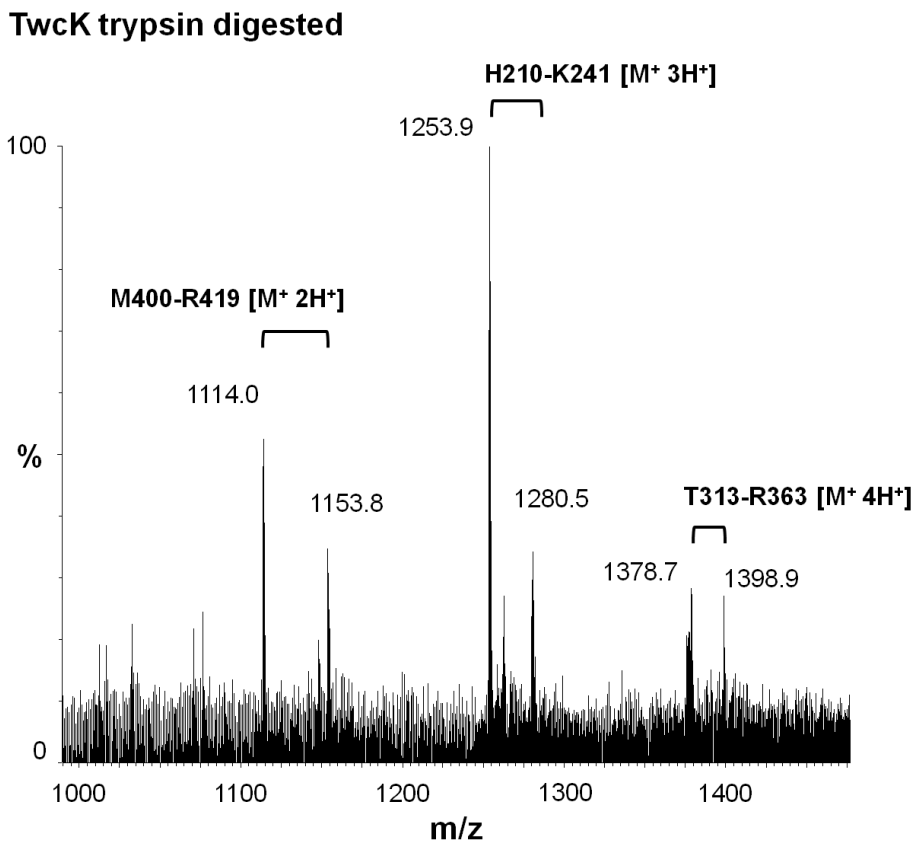


Fig 2.3.4.1 Identification of TwcK phosphorylation sites by mass spectrometry. MALDI-MS spectra following tryptic digest of TwcK and phosphopeptide enrichment. Pairs of peaks corresponding to non-phosphorylated and phosphorylated peptides are labelled.

2.3.5 Validation of phosphorylation sites by site-directed mutagenesis

Following analysis of phosphorylated peptides by mass spectrometry, it remained to validate and confirm the identity of phosphorylation sites due to the presence of multiple Ser/Thr residues within several of the identified peptides. A series of Kin mutants were generated by site-directed mutagenesis in which the sites identified by mass spectrometry were substituted for a non-reactive alanine residue. The mutants produced were as follows: T212A, S234A, S290A, T301A, T316A, T401A and T412A. T316 was chosen as the most likely site of substitution over T313 and T314 due to its proximity to the active site, directly facing the catalytic aspartate of TwcK as seen in the TwcKR structure (von Castelmur *et al.,* 2012). Additionally, a structurally equivalent residue to T316 was identified as a site of autophosphorylation in ZIPK (Zipper-interacting kinase), a member of the Ca^{2+}/calmodulin-dependent kinase family showing 40% sequence identity to the TwcK catalytic domain (Graves *et al.,* 2005).

Upon expression of Kin phosphorylation site mutants, substitution to alanine at T401 rendered the protein completely insoluble. A further mutant was therefore made in which T401 was substituted for aspartate (T401D). This T401D mutant exhibited no decrease in solubility when compared to the wild-type Kin construct.

As shown previously, SDS-PAGE can be used as an indicator of the level of phosphorylation present within a TwcK sample. Kin phosphorylation site mutants were recombinantly expressed, purified **(Fig 2.3.5.1a)** and analysed by SDS-PAGE, with a change in band composition interpreted as a change in phosphorylation state. Based on SDS-PAGE, T212A, T301A, T316A and T401D mutants appeared to show differences in band composition when compared to wild-type Kin **(Fig**

2.3.5.1b). In the case of T212A and T316A mutants, the difference in SDS-PAGE profile appeared most stark, with the loss of species of higher apparent molecular weight and therefore higher phosphorylation state seen in both. The remaining species for T212A and T316A mutants were equivalent in apparent molecular weight to the minimally phosphorylated species of wild-type Kin. In the cases of T301A and T401D mutants, an apparent loss of a component of the upper, highly phosphorylated species was seen along with a corresponding increase in intensity of a species of intermediate apparent molecular mass.

Purified samples of each phosphorylation site mutant were subject to tryptic digest and analysed by ES-MS as previously described. For all mutants, mass spectrometry showed loss of the phosphorylated peptide from the previously observed pairs of phosphorylated and non-phosphorylated peaks in wild-type Kin samples. It was therefore confirmed that each substituted position was indeed a site of phosphorylation in wild-type Kin. Samples from T212A, T301A and T401D mutants were seen to retain phosphorylation at the remaining sites, whereas phosphorylation at all sites was seen to be abolished in the T316A mutant, suggestive of disruption of catalysis in Kin^{T316A}.

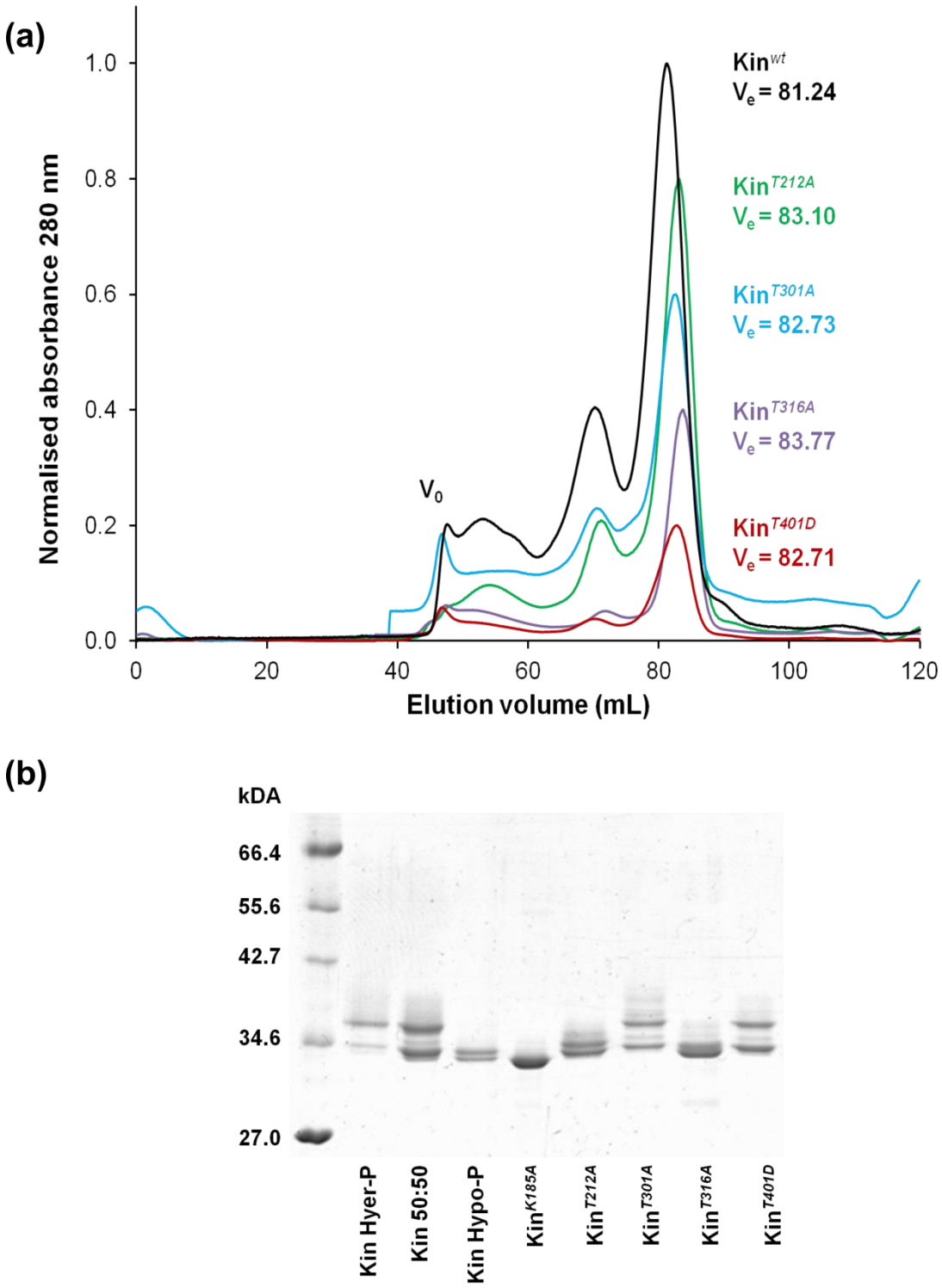


Fig 2.3.5.1 Purification of TwcK phosphorylation site mutants and SDS-PAGE analysis. (a) SEC chromatogram of the TwcK phosphorylation site mutants and wild-type Kin. Exclusion volume (V_e) and column void volume (V_0) are indicated. **(b)** SDS-PAGE following SEC. Molecular weight marker is shown to the left.

2.3.6 Twck autophosphorylation occurs at mechanistically important regions for catalysis

After identification of major phosphorylation sites, the corresponding residues were mapped onto the existing crystal structure of TwcK (PDB: 3UTO). When mapped onto the crystal structure, it became clear that three of the four identified phosphorylation sites were located within mechanistically important regions of the kinase fold **(Fig 2.3.6.1)**.

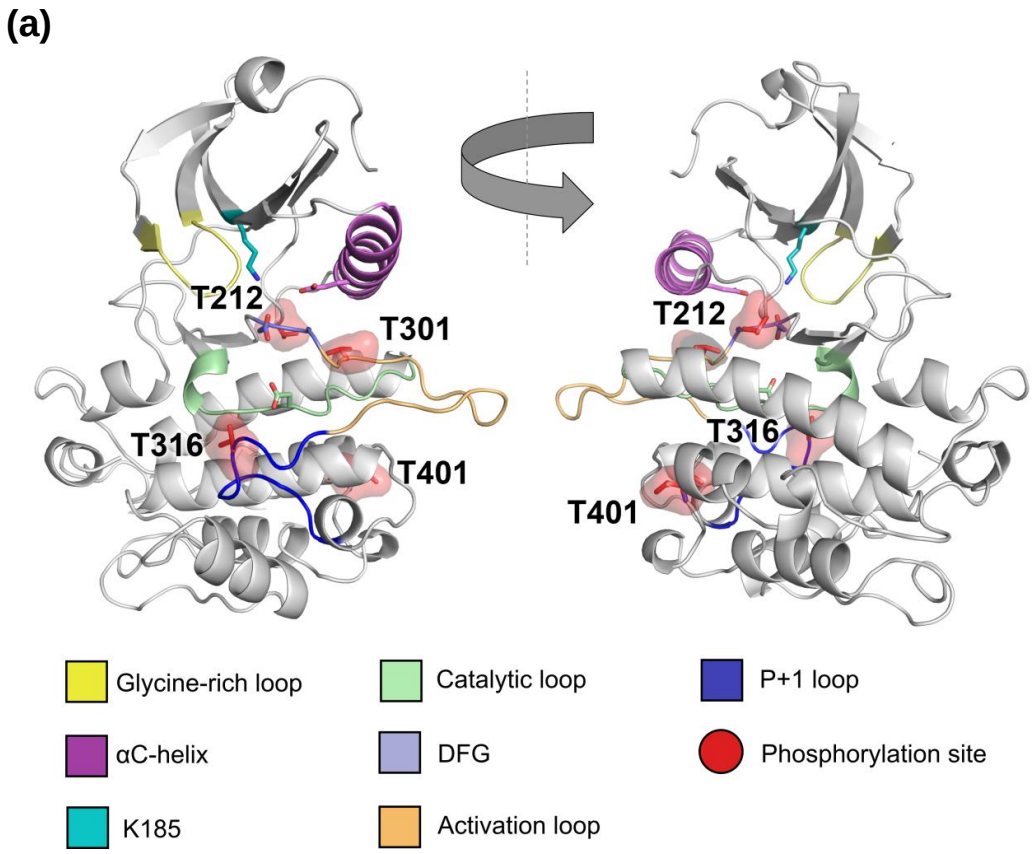
T212 is located at the base of the αC-helix within the αC-β4 loop. In protein kinases, this loop is generally seen to contain a conserved HxN motif. With interlobular motions required during catalysis, this HxN motif is a site of contact between the N- and C- lobes, anchoring the αC-helix and acting as a hinge, allowing the transition of the αC-helix between its active and inactive conformations. The HxN motif is therefore considered a possible regulatory element in eukaryotic protein kinases (Kannan & Neuwald, 2005; Kannan *et al.,* 2008). The conserved asparagine of the HxN motif is replaced by T212 in *C. elegans* TwcK.

T301 is found immediately following the highly conserved DFG motif (D297-G299). The DFG motif is essential for catalysis, with the aspartate residue coordinating Mg^{2+} ions from the ATP/Mg^{2+} substrate.

T316 is located in the P+1 loop of TwcK. The P+1 loop of protein kinases forms a critical part of the substrate-binding region, helping to position the phosphorylatable substrate residue for catalysis (Johnson *et al.,* 2001; Endicott *et al.,*

2012). In TwcK, T316 faces directly in to the active site in close proximity to the catalytic aspartate.

T401 is the only one of the four identified TwcK phosphorylation sites that is not located within a known region of mechanistic importance to protein kinase catalysis. T401 forms the N-terminal cap of the penultimate helix within the bottom of the C-terminal lobe. Despite its distance from the active site, this region has been identified as a potential site of protein-protein interactions in a computational study of the eukaryotic protein kinase fold and its subdomains (McClendon *et al.*, 2014).



(b)

NL

```
102       110       120       130       140       150       160
DERKRRRG YDVDEQGKIV RGKGTVSSNY DNYVFDIWKQ YYPQPVEIKH DHVLDHYDIH EELGTGAFGV

170       180       190       200       210       220       230
VHRVTERATG NNFAAKFVMT PHESDKETVR KEIQTMSVLR HPTLVNLHDA FEDDNEMVMI YEFMSGGELF

Kin
240       250       260       270       280       290       300
EKVADEHNKM SEDEAVEYMR QVCKGLCHMH ENNYVHLDLK PENIMFTTKR SNELKLIDFG LTAHLDPKQS

310       320       330       340       350       360       370
VKVTTGTAEF AAPEVAEGKP VGYYTDMWSV GVLSYILLSG LSPFGGENDD ETLRNVKSCD WNMDDSAFSG

380       390       400       410       420       430       440
ISEDGKDFIR KLLLADPNTR MTIHQALEHP WLTPGNAPGR DSQIPSSRYT KIRDSIKTKY DAWPEPLPPL

CRD
450       460       470
GRISNYSSLR KHRPQEYSIR DAFWDRS
```

Fig 2.3.6.1 Location of TwcK phosphorylation sites in the TwcK molecule. (a) Identified autophosphorylation sites are shown in the crystal structure of the catalytic domain of TwcK extracted from PDB: 3UTO. Catalytically important regions of the kinase fold are highlighted. **(b)** TwcK amino acid sequence with NL, catalytic domain (Kin) and CRD indicated. Identified phosphorylation sites are underlined in red.

To assess the potential structural consequences of phosphorylation at the identified sites, analysis of folding energy (ΔΔG) was performed using FoldX. A positive ΔΔG of more than 0.5 kcal/mol is considered to be destabilising. FoldX was used to estimate the destablising effect of each phosphorylated residue individually, with phosphorylation at T301 and T316 predicted to be most poorly tolerated based on large increases in ΔΔG of approximately 6 and 4 kcal/mol respectively **(Fig 2.3.6.2)**.

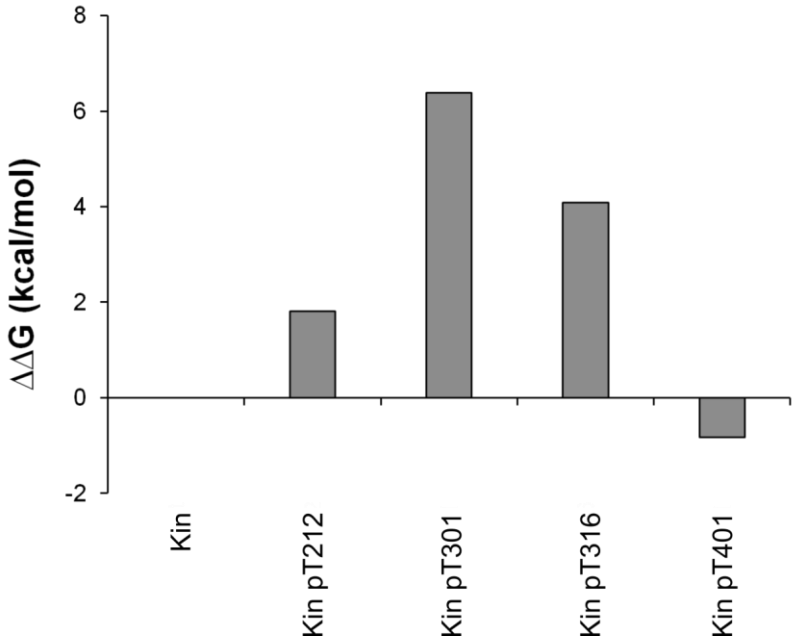


Fig 2.3.6.2 FoldX calculations predict a strongly destabilising effect of phosphorylation at T301 and T316. FoldX was used to calculate ΔΔG of folding for single phosphorylation events at each site. Phosphorylation and T301 and T316 were predicted to be the most destabilising, with increasing of ~6 and ~4 kcal/mol respectively.

The destabilisation that likely occurs following phosphorylation of T301 and T316 is in keeping with their location within the TwcK catalytic domain. T301 is adjacent to the DFG motif which is important for correct arrangement of the kinase active site, including the αC-helix and activation loop. T316 is also in close

proximity to the active site. Additionally, autophosphorylation of the other identified sites was abolished in the T316A mutant, suggesting that this region is highly sensitive to modification.

2.3.7 The effect of TwcK phosphorylation on catalytic activity

To study the effect of individual phosphorylation events on catalysis, the Kin phosphorylation site mutants (Kin^{T212A}, Kin^{T301A}, Kin^{T316A} and Kin^{T401D}) were subject to phosphotransfer assays **(Fig 2.3.7.1)**. Results of phosphotransfer assays revealed that the T301A and T401D substitutions had little effect on TwcK catalysis, with only a relatively small reduction in activity observed (approximately 64% and 80% of maximal wild-type Kin activity respectively). Kin^{T301A} and Kin^{T401D} were correspondingly those mutants that showed the least dramatic change in band composition as based on SDS-PAGE. In contrast, phosphotransfer assays for Kin^{T212A} and Kin^{T316A} mutants showed marked decreases in activity compared to wild-type, with Kin^{T212A} exhibiting approximately 7% of wild-type activity and Kin^{T316A} showing <1% of wild-type activity.

Following phosphotransfer assays on the initial phosphorylation site alanine mutants, another set of mutants were generated to replace T212 and T316 with phosphomimics in the form of aspartate. Substitution of phosphorylatable residues for negatively charged aspartate or glutamate residues is often used to mimic the negative charge introduced by phosphorylation (Dissmeyer & Schnittger, 2011). Phosphotransfer assays on these phosphomimics showed that in the case of Kin^{T212D}, a strong recovery in activity to approximately 70% of wild-type can be seen upon

replacement of T212 with aspartate. However, Kin^{T316D} showed no increase, with residual activity (<1% of wild-type) as with Kin^{T316A} **(Fig 2.3.7.1).**

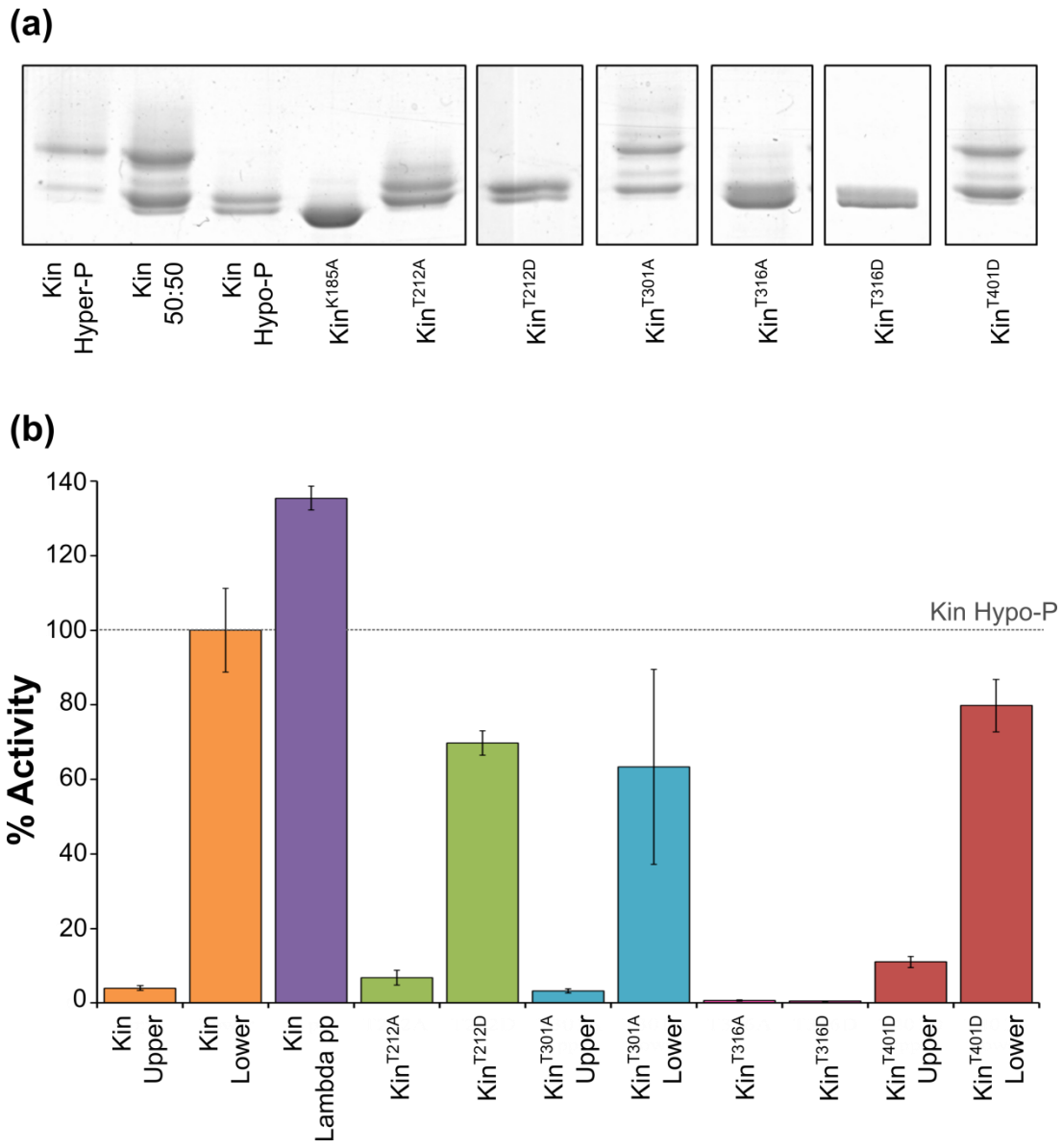


Fig 2.3.7.1 Analysis of TwcK phosphorylation site mutants. (a) SDS-PAGE analysis of TwcK phosphorylation site mutants. **(b)** Phosphotransfer assay on TwcK phosphorylation site mutants. The various wild-type TwcK states are shown as reference. Values (as %) are relative to the hypo-phosphorylated Kin species. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

With significant recovery of catalytic activity in the Kin^{T212D} mutant, the reduction in activity in the Kin^{T212A} was likely to be due to the loss of side-chain in

the HxN hinge-region of TwcK disrupting catalysis. The loss of activity in the Kin^{T316A} mutant was not recoverable by T316D mutation and was the only mutant in which autophosphorylation of all other sites was also abolished. It is therefore likely that T316 is the major site in terms of inhibition by autophosphorylation in TwcK, with the other sites having a cumulative effect on catalysis.

2.4 Discussion

Auto-phosphorylation is a common mechanism of kinase regulation. When the auto-phosphorylation occurs in the catalytic kinase domain, this is often within the activation loop, which transitions from a "closed" to an "open" conformation inducing the associated rearrangement of the active site. This allows the magnesium binding loop and αC-helix to be correctly positioned for catalysis as well as creating a favourable environment for the binding of the peptide substrate (Nolen *et al.,* 2003; Steichen *et al.,* 2010).

Although examples of kinase inhibition by auto-phosphorylation are found, such events commonly occur in accessory, regulatory domains. This is the case for the tyrosine kinase c-Src (cellular Src Kinase), in which negative regulation of catalytic activity occurs by phosphorylation of Y527 located within a C-terminal extension (Thomas *et al.,* 1991). Phosphorylation of Y527 is carried out by C-terminal Src kinase (CSK) (Murphy *et al.,* 1993) and the phospohorylated Y527 interacts with the SH2 (Src homology 2) domain of c-Src, which along with the c-Src SH3 (Src homology 3) domain is able to bind the catalytic domain, exerting allosteric inhibition on c-Src catalysis (Sicheri & Kuriyan, 1997). Inhibitory autophosphorylation of CAMK (Ca^{2+}/calmodulin-dependent protein kinase) has also been shown as a mechanism of negative autoregulation (Hanson & Schulman, 1992). CAMK is activated by binding of Ca^{2+}/calmodulin and autophosphorylation at T305 and T306 in the C-terminal calmodulin-binding domain of CAMK prevents Ca^{2+}/calmodulin binding and activation (Hanson & Schulman, 1992). Hyperphosphorylation of Raf-1 by ERK negatively regulates Raf-1 activity and

desensitises Raf-1 to additional stimuli. Dephosphorylation by PP2A (protein phosphatase 2) is required to return Raf-1 to an active state (Dougherty *et al.,* 2005).

In titin-like kinases, the activation loop and catalytic groups (including the DFG motif) seem to adopt an active conformation and appear primed for productive substrate binding and catalysis (Kobe *et al.,* 1996; Mayans *et al.,* 1998; von Castelmur *et al.,* 2012; Bogomolovas *et al.,* 2014). Thus, regulation by autophosphorylation was not expected. Interestingly, we find here that hyper-, auto-phosphorylation in TwcK serves as a mechanism of catalytic silencing. Phosphorylation events are identified in the P+1 loop (T316), the HxN motif at the base of the αC helix (T212), adjacent to the DFG motif (T301) and in a site distant from the active site that has been implicated in protein-protein interactions (T401). Of these, the modification of residue T316 in the P+1 loop appears to act as a main inhibitory event. The P+1 loop is part of the peptide substrate binding cleft (Johnson *et al.,* 2001; Endicott *et al.,* 2012), so its alteration influences substrate binding properties. With the P+1 loop crucial for substrate binding, disruption of activity in both $TwcK^{T316A}$ and $TwcK^{T316D}$ mutants is likely a consequence of reduced affinity for peptide substrate.

Phosphorylation of the P+1 loop is also observed in other kinases at residues structurally equivalent to T316 in TwcK. For example, phosphorylation occurs in an equivalent threonine residue in this loop (T180) in Zipper Interacting Protein Kinase (ZIPK), a member of the DAPK family (Graves *et al.,* 2005), although the role of this phosphorylation event in regulation of catalytic activity remains unclear. As in TwcK, ZIPK containing TtoA or TtoD mutations at this site did not display catalytic activity (Graves *et al.,* 2005). The microtubule affinity regulating kinase 2 (MARK2) also undergoes phosphorylation in the P+1 loop. Activation of MARK2 requires

phosphorylation within the activation loop by the up-stream kinase MARKK (microtubule affinity regulating kinase-kinase). However, MARK2 can also be phosphorylated by MARKK at S212 in the P+1 loop, leading to its total inactivation (even the basal activity seen in unphosphorylated MARK2 is lost) (Timm *et al*., 2003). As in TwcK and ZIPK, MARK2 activity is completely abolished in S212A and S212E mutants, demonstrating the sensitivity of this site to modification. Thus, in TwcK, MARK2 and possibly ZIPK, phosphorylation of the P+1 loop acts as a means of negative regulation. Interestingly, P+1 loop phosphorylation has also been described for the mitotic kinase Bub1, where it causes instead a change in substrate specificity (Lin *et al.*, 2014). Crystal structures of both the P+1 unphosphorylated (PDB: 4R8Q) and P+1 phosphorylated forms (PDB: 4QPM) show that phosphorylation at S969 forces the loop out from the active site and changes its conformation. This change in Bub1 enables discrimination between two *in vivo* substrates, H2A and cdc20. Phosphorylation at S969 greatly increased H2A phosphorylation but had no effect on cdc20 phosphorylation, indicating that the rearangement of the P+1 loop resulting from phosphorylation influenced substrate specificity (Lin *et al.*, 2014). In the case of TwcK, only a single peptide substrate has been identified, namely the chicken gizzard myosin light chain-derived peptide, kMLC11-23 (Heierhorst et al., 1996b). It cannot thus be currently determined whether phosphorylation of T316 in the P+1 loop of TwcK is able to modulate substrate specificity in a similar manner to that described for Bub1 (Lin et al., 2014). Exploration of potential peptide substrates is therefore necessary to better understand TwcK (auto-) catalysis as well as in working towards identification of an *in vivo* substrate for TwcK. This would be best undertaken by using peptide arrays. Peptide arrays are constructed by immobilising multiple peptides on a surface such as a glass

side, on which assays can be performed by spotting reactions on to the surface. Imaging is then performed (such as radioimaging for phosphotransfer assays using 32P ATP) to identify those peptides towards which the kinase shows phosphotransfer activity. Specificity and preference for the position of phosphorylatable residues, basic residues etc. can be studied by using overlapping peptides. Large arrays of peptides can be constructed using thousands of potential substrates, for example *C. elegans* muscle proteins (Shigaki *et al.*, 2007; Böhmer & Uecker, 2009; Thiele *et al.*, 2011).

The phosphorylation site T401 is not located in a known regulatory or catalytic region of TwcK however the equivalent residue to T401 in ZIPK has been shown to be phosphorylated (Graves *et al*., 2005). This region of the kinase fold has been identified as a possible site for protein-protein interactions (McClendon *et al.*, 2014) and may be a site of recognition or interaction with other proteins in the muscle fibre.

The findings of this study suggest that the control of TwcK activity *in vivo* might involve mechanical activation followed by a subsequent desensitization (inactivation) by phosphorylation. Although TwcK is capable of autophosphorylation during expression in *E. coli*, due to the ordered nature of the sarcomere, TwcK is spatially constrained so that some of these sites may be inaccessible for autophosphorylation *in vivo*. In the case of T316, due to its close proximity to the active site, facing directly inwards towards the catalytic aspartate, it could be envisaged that TwcK would be capable of *cis*-autophosphorylation at T316 *in vivo*. Indeed, twitchin from *Aplysia* is also likely to undergo autophosphorylation (Heierhorst *et al*., 1994; Butler & Siegman, 2011). In the case of T212, T301 and T401, it is unclear whether autophosphorylation would be achievable *in vivo.*

Phosphorylation of TwcK may therefore add to the instrasteric regulation by the NL and CRD tails to form a triple auto-inhibitory mechanism. The possibility of a triple auto-inhibitory mechanism suggests that TwcK activity in the sarcomere must be both spatially and temporally, tightly regulated, with punctual activity followed by rapid desensitisation.

It has been recently demonstrated that *C. elegans* TwcK binds the nematode MAPKAP kinase 2 orthologue, MAK-1, and is phosphorylated by MAK-1 *in vitro,* as well as evidence to suggest that a number of other related kinases interact with TwcK (Matsunaga *et al.,* 2015). This hints at the possibility of a feedback mechanism acting alongside mechanical activation, involving inactivation by phosphorylation followed by resensitisation by dephosphorylation.

Regulation of TwcK catalysis by phosphorylation implies that the intervention of a phosphatase is required to revert TwcK to its ground state. This would allow for TwcK catalysis to be inhibited while still in a mechanically primed state due to stretch. In *C. elegans,* the kinase domains from the titin-like protein UNC-89 bind SCPL-1, a phosphatase that appears to have a role in muscle regulation, with SCPL-1 knockouts causing defects in egg-laying muscles (Qatoda *et al.,* 2008). It is possible that TwcK may also have a phosphatase binding partner that could function to resensitise TwcK following phosphorylation mediated inactivation.

With titin-like kinases acting as signalling nodes in stretch-activated pathways, phosphorylation could potentially mediate interactions with protein binding partners. Phosphorylation-mediated binding has been demonstrated in a number of cases, including SH2 domains that recognise proteins phosphorylated at tyrosine residues (Koch *et al.,* 1991) and 14-3-3 proteins that bind in a phosphorylation-dependent manner (Ngok *et al.,* 2013). This could allow for

recruitment of protein binding partners, in keeping with the scaffolding role of titin-like kinases in signalling pathways.

Modification of the identified sites in TwcK may occur in other members of the titin-like kinase family. There is significant conservation of the sites in other titin-like kinases: T316 and T401 are strictly conserved in the TwcKs (sequence alignments in Mayans *et al.*, 2013) and T316 is often a serine among the vertebrate obscurin PK1s. T301 is frequently a serine in invertebrate UNC-89 PK1. T401 is largely serine in vertebrate obscurin PK1 and often threonine (nematode) or aspartate (insect) in invertebrate UNC-89 PK1. T401 is also threonine in human TK. Phosphorylation at these sites could occur as a means of regulating catalytic activity or to modulate protein-protein interactions. Work must now be carried out to determine whether these sites are modified *in vivo* and if so, how these sites are functionally relevant.

Chapter 3

The active conformation of twitchin kinase

In this chapter, we present the crystal structure of the twitchin kinase catalytic domain in its active conformation, the first structure of a titin-like kinase in the absence of its autoinhibitory tails. We assess the conformational changes undergone by twitchin kinase to access the active conformation, explore the possible role of the CRD as a pseudosubstrate and test twitchin kinase catalysis on a potential physiological substrate.

3.1 Introduction

3.1.1 The protein kinase active conformation

In their functional cycle, protein kinases transition between open, inactive conformational states and closed, active forms (Huse & Kuriyan, 2002; Kornev *et al.,* 2006). A number of essential features of the kinase fold undergo rearrangement in order to permit the adoption of the catalytically productive active conformation, with a delicate interplay between the various protein kinase regions acting in concert to create a favourable environment for catalysis.

The elucidation of the motions and conformational changes involved in protein kinase catalysis have been facilitated by crystal structures of a number of protein kinases in active and inactive conformations. The first crystal structure of a protein kinase-substrate complex was solved for PKA with a 20 amino acid

substrate-analogue peptide inhibitor, in what is now known as the closed conformation (Knighton *et al.,* 1991). Much of the subsequent work on the structural principles of protein kinase catalysis has also involved PKA. The crystal structure of PKA with no nucleotide bound (Cox *et al.,* 1994) revealed an open conformation of a protein kinase, showing that while the open conformation is important to allow access to ATP and for release of ADP following catalysis, lobe closure is required to bring the active site into a catalytically productive conformation for phosphotransfer (Cox *et al.,* 1994; Johnson *et al.,* 1996).

The inactive conformations amongst kinases are diverse, whereas the active conformation is seen to be similar throughout the kinase family (Bayliss *et al.,* 2015). The major conformational motions undergone in protein kinases to enter the active conformation are as follows:-

(i) Movement of the αC-helix is facilitated by the interlobular hinge region, positioning the conserved αC glutamate residue (E201 in TwcK), which along with the β3-lysine in the AxK motif (K185 in TwcK) chelates the β and γ phosphate groups of bound ATP (Gibbs *et al.,* 1991; Carrera *et al.,* 1993; Iyver *et al.,* 2005; Huang *et al.,* 2012).

ii) The flexible glycine-rich loop, formed between the β1 and β2 strands of the N-lobe closes down over the active site cleft and helps to position the **γ**-phosphate of ATP for phosphotransfer (Hemmer *et al.,* 1997; McNamara *et al.,* 2011). The closure of the glycine-rich loop is a distinctive and easily recognisable feature of a protein kinase in its active conformation.

iii) In some kinases, the DFG motif can occupy an 'out' conformation in the inactive state in which the aspartate responsible for Mg^{2+} coordination is flipped

away from the active site (Treiber & Shah, 2013). In addition to coordinating ATP via bound Mg^{2+} ions, hydrophobic interactions involving the DFG motif with both the αC-helix and HxD motif in the catalytic loop contribute to the correct positioning of a number of active site residues. The positioning of the DFG motif is itself influenced by the conformation of the activation loop.

iv) The activation loop is the most common site of phosphorylation-mediated protein kinase activation (Dorey *et al.*, 2001; Xu *et al.*, 2004; Roskoski Jr., 2004; Pike *et al.*, 2008; Steichen *et al.*, 2012). The mechanism by which activation loop phosphorylation leads to kinase activation varies. Most commonly it leads to changes in conformation of the active site, allowing the magnesium binding loop and αC helix to be correctly positioned for catalysis as well as creating a favourable environment for substrate binding (Nolen *et al.*, 2003; Steichen *et al.*, 2010).

In the active conformation, the conserved elements of the protein kinase fold are assembled and anchored to the F-helix by two hydrophobic spines (Taylor & Kornev, 2011) (**Fig 3.1.1.1**). Revealed by surface analysis of active and inactive kinase (Kornev *et al.*, 2006), the regulatory spine (R-spine), is composed of four hydrophobic residues; one from the β4 strand and αC-helix of the N-lobe and one from the activation loop and catalytic loop in the C-lobe (Kornev *et al.*, 2006; Taylor & Kornev, 2011). The mobile nature of the activation loop and αC-helix mean that the R-spine can be dynamically assembled and disassembled, corresponding to active and inactive conformations (Huse & Kuriyan, 2002; Taylor & Kornev, 2011).

The catalytic spine (C-spine) is also assembled from residues in both the N- and C-lobes, including the alanine of the β3 AxK motif which is seen to directly contact the adenine ring of ATP (Taylor & Kornev, 2011). This means that full assembly of the catalytic spine is completed by binding of ATP.

Interestingly in TwcK, the autoinhibited, inactive form of the kinase seen in the TwcKR crystal structure (PDB: 3UTO) exhibits several features characteristic of an active kinase conformation. The αC-helix, activation loop, DFG motif and catalytic loop appear to be in conformations permissible for catalysis. The hydrophobic spines that traverse the protein kinase fold and act as anchors for the catalytic machinery (Kornev *et al.*, 2006; Taylor & Kornev, 2011) are also assembled, save for completion of the C-spine by ATP binding. The glycine-rich loop is however seen to adopt an open conformation in the TwcKR crystal structure, with the CRD occupying the active site cleft and not allowing for full closure, as well as the blocking of the interlobular hinge region by the NL.

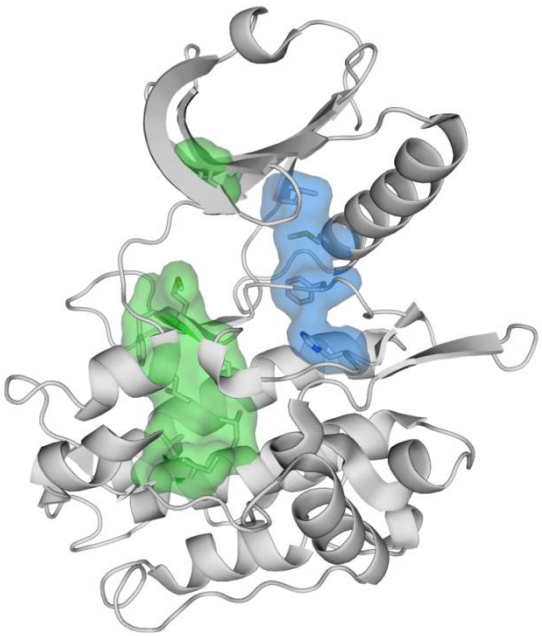


Fig 3.1.1.1 TwcK hydrophobic spines. The catalytic domain of the autoinhibited TwcKR crystal structure trimmed of regulatory tails, showing assembled hydrophobic spines. The regulatory spine (blue) is complete, with the catalytic spine (green) requiring only the binding of ATP for completion.

With the active site and hydrophobic spines of TwcK seen to be mostly assembled, it would appear that large scale conformational changes may not be required in order to access a catalytically active state. This is in keeping with the hypothesis of mechanoactivation in which removal of the NL leaves TwcK in a mechanically primed state (von Castelmur *et al.,* 2012).

3.1.2 TwcK as a myosin light chain kinase paralogue

The most closely related protein kinases to titin-like kinases based on sequence homology are the myosin light chain kinases (MLCKs). The catalytic domain of TwcK is 52% identical to that of chicken smooth muscle MLCK and 44% identical to rabbit skeletal muscle MLCK (Benian *et al.*, 1989). There is no evidence for MLCK in *C. elegans* (Wormbase genetic database, www.wormbase.org/species/c_elegans) or in molluscan muscle (Butler & Siegman, 2011) and it has long been hypothesised that TwcK could be the invertebrate paralogue of MLCK. This hypothesis is formed based on functional and catalytic comparisons between TwcK and MLCK as well as sequence similarities between TwcK and MLCK that suggest common regulatory principles that will be explained below.

Tissue specific MLCKs are expressed in both muscle and non-muscle cells and are involved in numerous processes. In smooth muscle, MLCK phosphorylation of N-terminal terminal regions of myosin regulatory light chains (RLCs) is the mainpathway for regulation of muscle contraction (Michnoff *et al.,* 1986; Ikebe, 1989; Itoh *et al.,* 1992; Getz *et al.,* 2011; Kamm & Stull, 2011). In skeletal and cardiac muscle, MLCK regulates the potentiation of muscle contraction (Sweeney *et*

al., 1993). Non-muscle MLCK is involved in a wide-range of processes such as cell migration, cytokinesis (Kamm & Stull, 2001), fibroblast contraction (Kolodney & Elson, 1993) and platelet contraction (Getz *et al.,* 2011). The *in vivo* substrate of MLCK is the N-terminus of myosin II RLC, which MLCK phosphorylates at Thr18 and Ser19, allowing for the myosin-actin crossbridge and initiation of contraction (Michnoff *et al.,* 1986; Ikebe, 1989; Itoh *et al.,* 1992; Getz *et al.,* 2011).

As with the titin-like kinases, MLCKs contain an Ig-Ig-Fn-linker-Kinase-tail-Ig domain arrangement. MLCKs are autoinhibited by a C-terminal extension that blocks RLC substrate binding. The autoinhibitory segment of MLCKs was identified through mutational analysis as comprising the region that links the C-terminus of the catalytic domain to a Ca^{2+}/CaM binding domain (Pearson *et al.,* 1988; Ikebe *et al.,* 1989; Olson *et al.,* 1990; Ito *et al.,* 1991; Knighton *et al.,* 1992; Gallagher *et al.,* 1993; Allen & Walsh, 1994; Krueger *et al.,* 1995). Through neutron scattering studies it was shown that Ca^{2+}/CaM dependent activation of MLCK occurs by displacement of the C-terminal autoinhibitory segment, freeing the active site for RLC substrate binding (Krueger *et al.,* 1997; Krueger *et al.,* 1998). Human TK and TwcK have been shown to bind Ca2+/CaM *in vitro* through the CRD, although this binding does not lead to activation of catalysis Gautel *et al.,* 1995; Mayans *et al.,* 1998; Lei *et al.,* 1994). TwcK kinase was activated by S100, a calmodulin-like protein *in vitro* (Heierhorst *et al.,* 1996a), although the absence of an S100 protein in *C. elegans* means this is not the activation of mechanism of TwcK *in vivo.*

An additional MLCK was recently identified in cardiac muscle (cMLCK) that lacks the autoinhibitory C-terminal segment (Chang *et al.,* 2016). The crystal structure of cMLCK, the first of any MLCK, shows the active conformation of the kinase with bound small-molecule inhibitor and reveals a short C-terminal helix,

dubbed the "pseudoregulatory helix" (PSH) that is insufficient to inhibit catalysis (Chang *et al.,* 2016).

The autoinhibitory C-terminal extension of MLCK acts as a pseudosubstrate inhibitor (Kemp *et al.,* 1987). Inhibition of kinase activity by a pseudosubstrate peptide is a well-established mode of inhibition and involves a peptide of similar sequence to that of the substrate that lacks a phosphorylatable residue. Binding of the pseudosubstrate to the catalytic domain blocks substrate binding and thereby inhibits catalysis (Pearce *et al.,* 2010). The crystal structure of PKA in complex with the pseudosubstrate inhibitor PKI revealed the conformation and binding mode of this pseudosubstrate peptide, forming an α-helix that occupies a groove on the C-lobe of the catalytic domain, followed by an extended structure that binds in the active site cleft **(Fig 3.1.2.1)** (Zheng *et al.,* 1993). Sequence analysis and NMR data revealed the similarity between the pseudosubstrate region of MLCK and the PKI pseudosubstrate of PKA, implying that the MLCK substrate is likely to interact with the catalytic domain in a similar manner to that seen in PKA (Barden *et al.,* 1996).

The crystal structure of TwcKs from *C. elegans* and *Aplysia californica* revealed the CRD, bound to the catalytic domain and blocking the active site cleft (Hu *et al.,* 1994; Kobe *et al.,* 1996). The αR1 helix at the N-terminus of the CRD was seen to occupy a similar position to that of the helical region of the PKA pseudosubstrate inhibitor **(Fig 3.1.2.1)**. With the high sequence similarity between TwcK and MLCK, it was thought that the CRD as seen in TwcK could represent a common mode of inhibition by the C-terminal domains of TwcK and MLCK. Catalytic data for TwcK however showed that recombinantly produced TwcK constructs containing only the first 36 residues of the CRD (αR1, αR2 and aR3 helices) were not inhibited, suggesting that the most important residues for

autoinhibition may lie within the last 24 residues of the CRD (Lei *et al.*, 1994). These data are in contradiction to the MLCK substrate, where the N-terminal region of the C-terminal domain, corresponding to the TwcK αR1 helix, has been shown to be autoinhibitory (Knighton *et al.*, 1992; Barden *et al.*, 1996). Thus, the exact determinants of CRD autoinhibition remained unclear and the pseudo-substrate mechanism in these kinase controversial.

The mechanoactivation hypothesis of TwcK regulation suggests that the αR1 helix of the TwcK CRD could remain bound to the catalytic domain, with steered molecular dynamics simulations suggesting it could act as a buffer to protect against mechanical deformation (von Castelmur *et al.*, 2012), implying a non-inhibitory role for the αR1 helix. For the αR1 helix to remain bound as a mechanical buffer, it must be compatible with substrate binding in order to allow catalysis. The αR1 helix as a mechanical buffer would therefore preclude a pseudosubstrate role as in MLCK. To date however, the question of the inhibitory properties of the N-terminal region of the CRD remains unanswered.

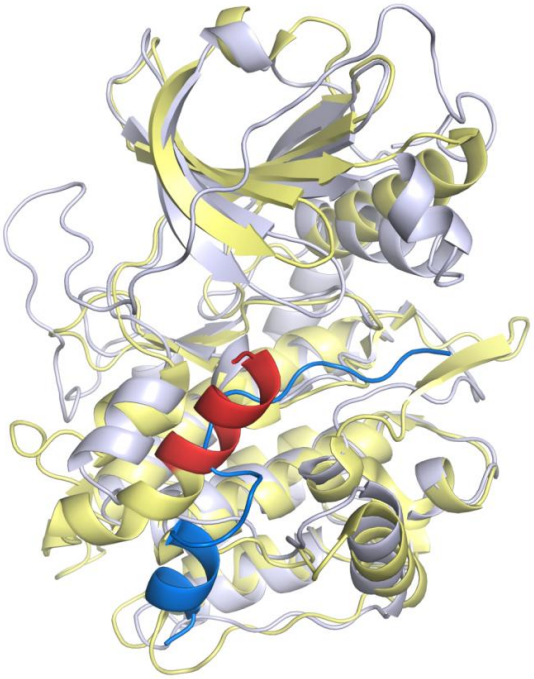


Fig 3.1.2.1 The TwcK αR1 helix as a potential pseudosubstrate. Superimposition of the crystal structures of the TwcKR kinase domain (yellow) with αR1 helix (red) and PKA (light purple) with bound pseudosubstrate inhibitor (blue). The TwcK αR1 helix occupies a similar position to part of the PKA pseudosubstrate. PDB entries: 1ATP and 3UTO.

In vivo data suggest that TwcK may play a similar role to MLCK in invertebrates. In molluscs, twitchin has been shown to be involved in maintaining the 'catch' state in smooth muscle (Siegman *et al.,* 1998; Funabara *et al.,* 2005; Yamada *et al.,* 2001; Butler and Siegman, 2011). The catch state in molluscan smooth muscle allows for maintenance of tension for prolonged periods of time with minimal energy consumption. Twitchin becomes phosphorylated at several sites upon release of the catch state and dephosphorylation of twitchin is required to re-enter the catch state (Funabara *et al.,* 2005). Dephosphorylation of twitchin allows for F-actin binding whereas twitchin phosphorylation decreases F-actin affinity (Funabara *et al.,* 2005). With twitchin shown to bind thick filaments, it is thought that twitchin connects thick and thin filaments in a phosphorylation-dependent manner to regulate the myosin cross-bridge cycle (Siegman *et al.,* 1998; Yamada *et*

al., 2001; Funabara *et al.,* 2005; Butler & Siegman, 2011). Furthermore, the catch state of smooth muscle in *Mytilus edulis* is regulated by a force-activated kinase that phosphorylates a TwcK peptide substrate, with a 10% increase in stretch corresponding to a two-fold increase in phosphorylation of this substrate, although it has not been definitively confirmed that this force-activated kinase is TwcK (Butler & Siegman, 2011). In the mollusc *Aplysia californica,* TwcK has been shown to phosphorylate myosin RLC at Thr15 (Heierhorst *et al.,* 1995), the first *in vivo* substrate discovered for any of the titin-like kinases. Thr15 of *Aplysia* myosin RLC is analogous to the phosphorylation site of vertebrate MLCK, further supporting the role of TwcK in invertebrate muscle contraction by regulation of the myosin cross-bridge cycle.

Early work on the function of twitchin in *C. elegans* showed that mutant strains that lack twitchin (encoded by the *unc-22* gene) exhibit a phenotype in which muscles are unable to perform normal contraction, with uncoordinated, transient contraction of individual sarcomeres without contraction of the adjacent sarcomere (Waterston *et al.,* 1980). All mutations outside of the *unc-22* gene that were capable of suppressing the phenotype caused by a lack of expressed twitchin were found in myosin heavy chain of the body wall muscle (Moerman *et al.,* 1982; Dibb, *et al.,* 1985), suggesting that twitchin may act via myosin in regulating muscle contraction.

In recent work carried out in collaboration with our lab, it has been shown that the role of TwcK activity in *C. elegans* is to regulate the contraction/relaxation cycle in the sarcomere (Matsunaga *et al.,* unpublished - **section 7.1.2**). This was achieved by mutation of the β3 lysine of TwcK (K185) to alanine, abolishing the activity of TwcK (**section 2.3.1)**. The K185A mutation was introduced in to *C. elegans* twitchin by CRISPR/Cas9 gene editing (Friedland *et al.,* 2013), resulting in

a mutant strain (KtoA) that expressed only the kinase-dead twitchin. *C. elegans* lacking TwcK activity were shown to undergo normal development, lack the twitching phenotype characteristic of many other mutations to twitchin (unc-22) (Waterston *et al.*, 1980), have regular sarcomeric organisation and unaffected life span and brood size. The mutant strain however showed increased speed of both swimming and crawling and had faster rates of muscle contraction/relaxation.

These findings suggest that the role of TwcK *in vivo* is to inhibit muscle contraction, meaning that its cellular pathway is ultimately involved in regulation of the motor proteins in the sarcomere. Despite the increased speed of movement in the KtoA strain, wild-type *C. elegans* appear to show greater competitive fitness than the KtoA mutant strain, indicating that TwcK catalysis *in vivo* is not required for development and viability but may be required for competitive survival (Matsungaga *et al.*, unpublished - **section 7.1.2**). The *in vivo* substrate of TwcK is assumed to be a sarcomeric component due to the localisation of TwcK to the sarcomere, although this cannot be concluded with certainty and thus the *in vivo* TwcK substrate must be identified to fully conclude on the role of TwcK in regulating muscle contraction/relaxation.

3.1.3 *C. elegans* TwcK substrates

TwcK from *C. elegans* phosphorylates vertebrate myosin RLC-derived peptides (Heierhorst *et al.*, 1996b), although a physiological substrate has yet to be found. Myosin RLC I from *C. elegans* has been previously tested as a potential TwcK substrate, with results showing that the tested myosin RLC is a poor TwcK

substrate (Heierhorst *et al.*, 1996b). This is unsurprising as myosin RLCs in *C. elegans* are highly divergent and the substrate in question contained no consensus sequence based on the known phosphorylation sites of MLCK and *Aplysia* TwcK **(Table 3.1.3.1)** (Michnoff *et al.*, 1986; Ikebe, 1989; Itoh *et al.*, 1992; Heierhorst *et al.*, 1996b). The current model peptide for catalytic studies of *C. elegans* is derived from chicken gizzard myosin light chain (Heierhorst *et al.*, 1996b) and contains substitutions to alanine at positions equivalent to residues 14, 15 and 23 of chicken gizzard myosin light chain. Peptide substrates derived from the identified myosin RLC phosphorylation site for *Aplysia* also proved to be poor substrates for *C. elegans* TwcK (Heierhorst *et al.*, 1996b). The identification of a physiological substrate for *C. elegans* TwcK is required to understand the signalling pathway(s) in which TwcK intervenes.

Table 3.1.3.1 TwcK and MLCK substrate peptides. TwcK model peptide susbtrate derived from residues 11-23 of chicken gizzard myosin light chain (kMLC 11-23), *Aplysia* TwcK rMLC and Vertebrate rMLC II are seen to be highly similar in sequence, containing an N-terminal cluster of basic residues and a conserved ATSNVF motif. The *C. elegans* rMLC I-derived peptide as tested in Heierhorst *et al.*, 1996b is highly divergent and was seen to be a poor substrate for *C. elegans* TwcK. Residues known to undergo phosphorylation are highlighted in red.

Substrate name	Sequence
kMLC 11-23 TwcK model substrate	KKRARAATSNVFS
Aplysia TwcK myosin RLC substrate	KKGRSGRA**T**SNVFA
Vertebrate myosin RLC II substrate	KKRPQRA**TS**NVFS
C. elegans myosin RLC I	MSKIIAKKKSSNKKRSGSEAAQ

3.1.4 Specific aims

This work in this chapter aims to elucidate the catalytic principles of TwcK in terms of its active conformation as well as its substrate specificity and potential role as an invertebrate paralogue of MLCK. These aims can be summarised as follows:

i) **The active conformation of TwcK –** we aim to solve the crystal structure of TwcK in the absence of its regulatory tails to gain an understanding of any conformational changes required to access the active conformation.

ii) **The role of the CRD in the pseudosubstrate mechanism -** investigation into whether the αR1 helix is autoinhibitory as in MLCKs and would need to be removed from the kinase domain in order to allow catalysis, or whether it is in fact compatible with substrate binding and therefore compatible with catalysis if it remains in place upon activation.

iii) **Investigation of a potential physiological substrate for TwcK -** with the potential role of TwcK as an invertebrate paralogue of MLCK as described herein, we have identified a potential physiological substrate derived from a *C. elegans* myosin light chain and aim to test its suitability as a TwcK substrate.

In addressing these questions, we aim to learn more about the principles of TwcK catalysis and autoinhibition, helping us to better understand the *in vivo* role of TwcK activity.

3.2 Methods

3.2.1 Cloning of Kin-αR1 construct

TwcK Kin-αR1 (Uniprot Q23551 residues 6254-6548), corresponding to a 9 amino acid extension to the TwcK caytalytic domain **(Fig 3.2.1.1)**, was subcloned by PCR using the larger NLKinCRD as template. PCR amplification was performed using Q5 DNA polymerase (NEB) with the following primers:

	NcoI
Kin forward primer	5'- gcgccatggaaatcaagcatgatcatgttctgg-3'

	KpnI
αR1 reverse primer	5'- agtggtaccttaatcgtatttggtcttgattga-3'

Digestion used NcoI and KpnI restriction enzymes (NEB) with the restriction sites indicated above. Digested PCR product was ligated into the pETM-11 vector using T4 DNA ligase (NEB). Ligation into pETM-11 plasmids yields a protein product containing an N-terminal His_6 affinity tag followed by a linker sequence and Tobacco Etch Virus (TEV) protease cleavage site for His_6-tag removal. Cloning was verified by sequencing (Source Bioscience).

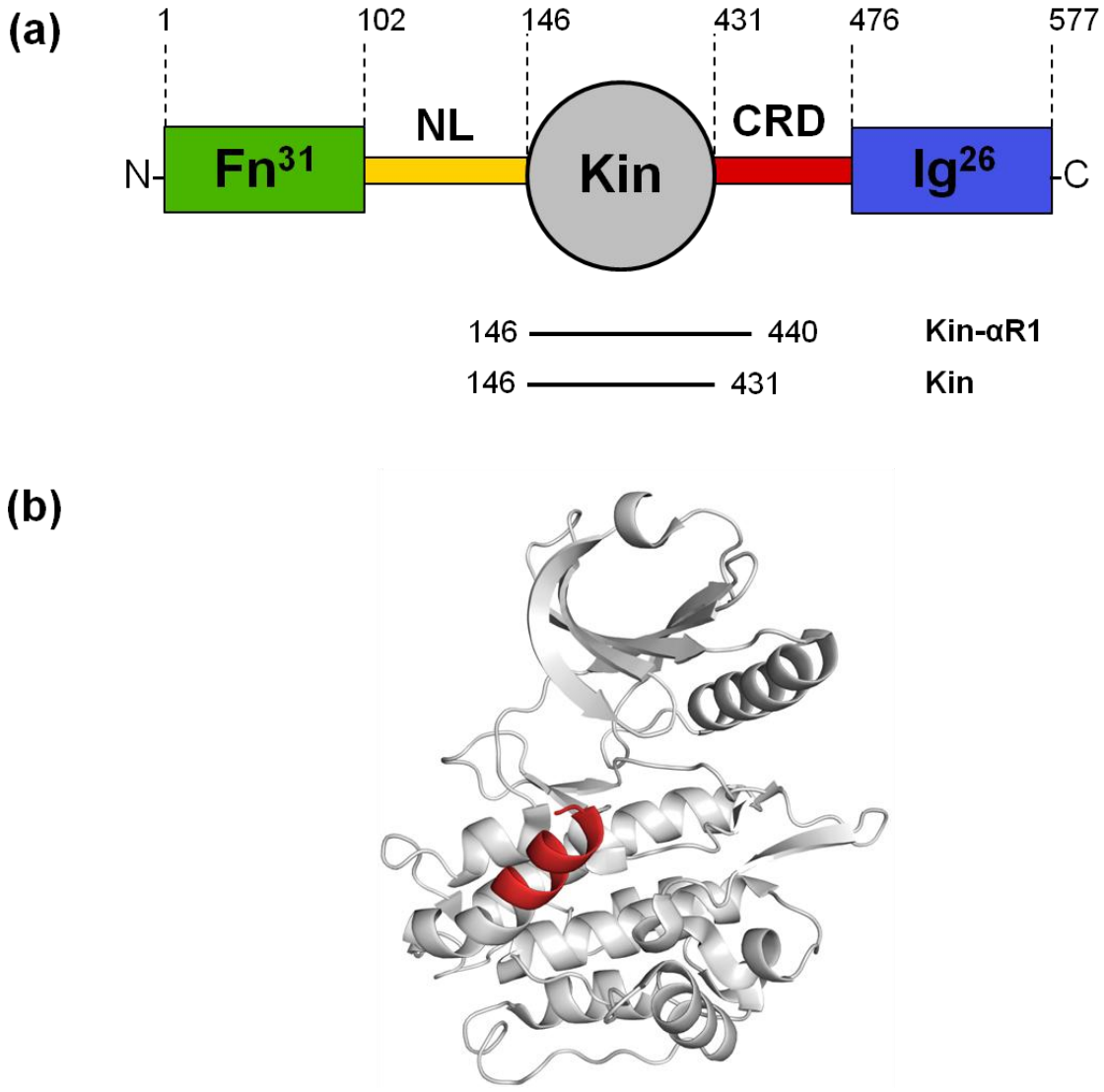


Fig 3.2.1.1 The TwcK Kin-αR1 construct. The TwcK Kin-αR1 construct comprises the catalytic domain and the αR1 helix of the CRD, representing a 9 amino acid extension to the Kin construct. **(a)** schematic representation of the Kin and Kin-αR1 constructs within the TwcK region **(b)** TwcKR crystal structure (PDB: 3UTO) showing the extent of the catalytic domain (grey) and the αR1 helix of the CRD (red).

3.2.2 TwcK Kin and Kin-αR1 protein expression and purification

TwcK Kin and Kin-αR1 were overexpressed in *E.coli* Rosetta (DE3) (Merck Millipore) as described in section **2.2.3**. Purification of Kin and Kin-αR1 in

preparation for phosphotransfer assays was also performed as described in section **2.2.3.**

For crystallisation of Kin-αR1, a modified protein purification protocol was used based on the purification of TwcK established by von Castelmur *et al.*, 2012 but incorporating a dephosphorylation step to remove phosphate groups due to autophosphorylation, as follows:

Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, pH 7.9) containing 20 µg/mL DNase I (Sigma Aldrich) and 1 complete EDTA-free protease inhibitor tablet (Roche) per litre of cell culture. Cell lysis was by sonication on ice followed by clarification of lysate by centrifugation at 39000 *g* for 45 minutes. The lysate was syringe filtered (0.22µm) and applied to a 5 mL HisTrap HP column connected to an Åkta FPLC system (GE Healthcare) equilibrated in lysis buffer containing 20 mM imidazole. Elution of His_6-tagged protein was by continuous imidazole gradient.

For dephosphorylation, affinity purified Kin-αR1 samples were buffer exchanged into dephosphorylation buffer (50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, 2 mM $MnCl_2$ pH 7.9) using PD-10 desalting columns (GE Healthcare). Kin-αR1 was then incubated with 400 units of Lambda protein phosphatase (NEB) for 1 hour at room temperature for dephosphorylation, before addition of TEV protease for His_6-tag removal and incubation overnight at 4^oC. Subtractive Ni^{2+} affinity purification was then performed using a 5 mL HisTrap HP column connected to an Åkta FPLC system (GE Healthcare) equilibrated in dephosphorylation buffer.

Anion exchange chromatography was carried out using a 5 mL QFF column on an Åkta FPLC system (GE Healthcare). Prior to sample loading, QFF columns

were washed with 5 column volumes AEC buffer A (50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT) followed by 5 column volumes AEC buffer B (50 mM Tris-HCl, 1 M NaCl, 1 mM DTT) and finally 5 column volumes AEC buffer A. Elution was with a continuous gradient of AEC buffer B over a 10 column volume length.

Following AEC, Kin-αR1 was concentrated in preparation for size exclusion chromatography (SEC) using centrifugal filter units with MWCO 10,000 (Amicon). SEC was carried out using a Superdex 200 16/60 column (GE Healthcare) in 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM TCEP, pH 7.9. The resulting fractions were analysed by SDS-PAGE, with those fractions showing sample heterogeneity indicative of incomplete dephosphorylation being removed. The remaining homogeneous SEC fractions were pooled for crystallisation.

3.2.3 Crystallisation of TwcK Kin-αR1 with Staurosporine

The use of staurosporine in crystallisation of protein kinases was first described in Prade *et al.*, 1997. Purified Kin-αR1 protein was concentrated to 9.5 mg/mL using centrifugal filter units with MWCO 10,000 (Amicon). Staurosporine (Thermo Fisher) was added to the concentrated protein sample from a 5mM stock (in DMSO) to a final molar ratio of 1:1.5 (protein:staurosporine) and incubated for 1 hour on ice before being centrifuged at 13,000 g for 10 minutes. Final concentration of Kin-αR1 protein following addition of staurosporine was 8.6 mg/mL.

Crystal trials were set up in 96 well Intelliplates (Art Robbins) using the sitting drop, vapour diffusion format and commercial crystallization screens. Trials were set using the Gryphon (Art Robbins) nanolitre dispensing robot. Drops

consisted of 150 nL Kin-αR1-Staurosporine mixture and 150 nL reservoir solution with reservoirs containing 70 µL reservoir solution. The following commercial screens were set up: JCSG+ Eco, PEG suite I, PEG suite II, Structure and Morpheus (all from Molecular Dimensions). All screens were incubated at 18^{o}C. Crystals were manually harvested using litho loops (Hampton Research) and flash cooled in liquid N_2. Crystals obtained from JCSG+ Eco screen were cryoprotected in mother liquor supplemented with 30% glycerol prior to flash cooling. Crystals obtained from Morpheus screen did not require supplementing with an additional cryoprotectant.

3.2.4 TwcK Kin-αR1 X-ray diffraction and data processing

X-ray data collection for TwcK Kin-αR1 crystals was carried out at beamline PXI of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) equipped with an Eiger 16M X detector (Dectris, Switzerland). Data was collected using a wavelength of 1 Å with 0.1^{o} oscillation per frame (0.1 second exposure time per frame).

Data processing used the XDS/XSCALE package (Kabsch, 2010). Molecular replacement was performed with Phaser (McCoy *et al.,* 2007), using the existing crystal structure of TwcKR (PDB: 3UTO) trimmed of flanking domains and regulatory tails as search model (von Castelmur *et al.*, 2012). The resulting model was refined using phenix.refine (Afonine *et al.,* 2012) and manual model building was in Coot (Emsley *et al.,* 2010). Molprobity (Chen *et al.,* 2010) was used for structure validation.

3.2.5 Determination of Kin-αR1 activity and Kin substrate specificity by [γ^{32}P]-ATP phosphotransfer assays

Phosphotransfer assays were performed as described in **section 2.2.4**, with the exception of those assays used to assess the substrate specificities of TwcK. For assays testing for peptide substrate specificity, control reactions were carried out with the model peptide substrate with sequence KKRARAATSNVFS as previously reported by Heierhorst *et al*., 1996b and von Castelmur *et al.,* 2012 and as described in **section 2.2.4**. Phosphotransfer assays against *C. elegans* myosin light chain 4 derived substrate (Uniprot Q09510 residues 11-22) contained peptide substrate with amino acid sequence RQRPQRATSNVFA. For reactions in which metal ion specificity was assessed, control samples contained 10 mM $MgCl_2$ as previously described which was replaced with 10 mM $MnCl_2$ or an equivalent volume of ddH_2O when required.

Peptide substrates were purchased from Biomatik (Delaware, USA) as lyophilised trifluoroacetate (TFA) salts, purified by HPLC to >95% purity. Peptides were dissolved in 50 mM Tris, 50 mM NaCl, 1 mM DTT, pH 7.9 and pH was adjusted to approx. pH 7.9 by addition of HCl.

3.3 Results

3.3.1 Production of recombinantly expressed Kin-αR1

Kin-αR1 was overexpressed in *E. coli* and purified chromatographically, resulting in a yield of 10 mg of soluble, pure protein per L culture. As previously described for catalytically active TwcK constructs **(section 2.3.1)** Kin-αR1 eluted from SEC as a heterogeneous sample based on SDS-PAGE. As with Kin samples, Kin-αR1 showed a characteristic SDS-PAGE banding pattern indicative of significant autophosphorylation within the sample and suggestive of a catalytically active TwcK sample **(Fig 3.3.1.1).**

With sample heterogeneity being undersirable for crystallisation and a likely contributory factor to previous failed crystallisation attempts, the method of purification for Kin-αR1 was modified to produce a more homogeneous Kin-αR1 sample. Guided by findings presented in **section 2.3.1** on TwcK autophosphorylation, a dephosphorylation step by Lambda protein phosphatase was introduced. Following SEC on Lambda protein phosphatase treated Kin-αR1, a more homogeneous sample was obtained as judged by SDS-PAGE, although some heterogeneity was still present **(Fig 3.3.1.1).** Those fractions containing multiple species, likely due to incomplete dephosphorylation, were removed before pooling the remaining fractions for crystallisation trials.

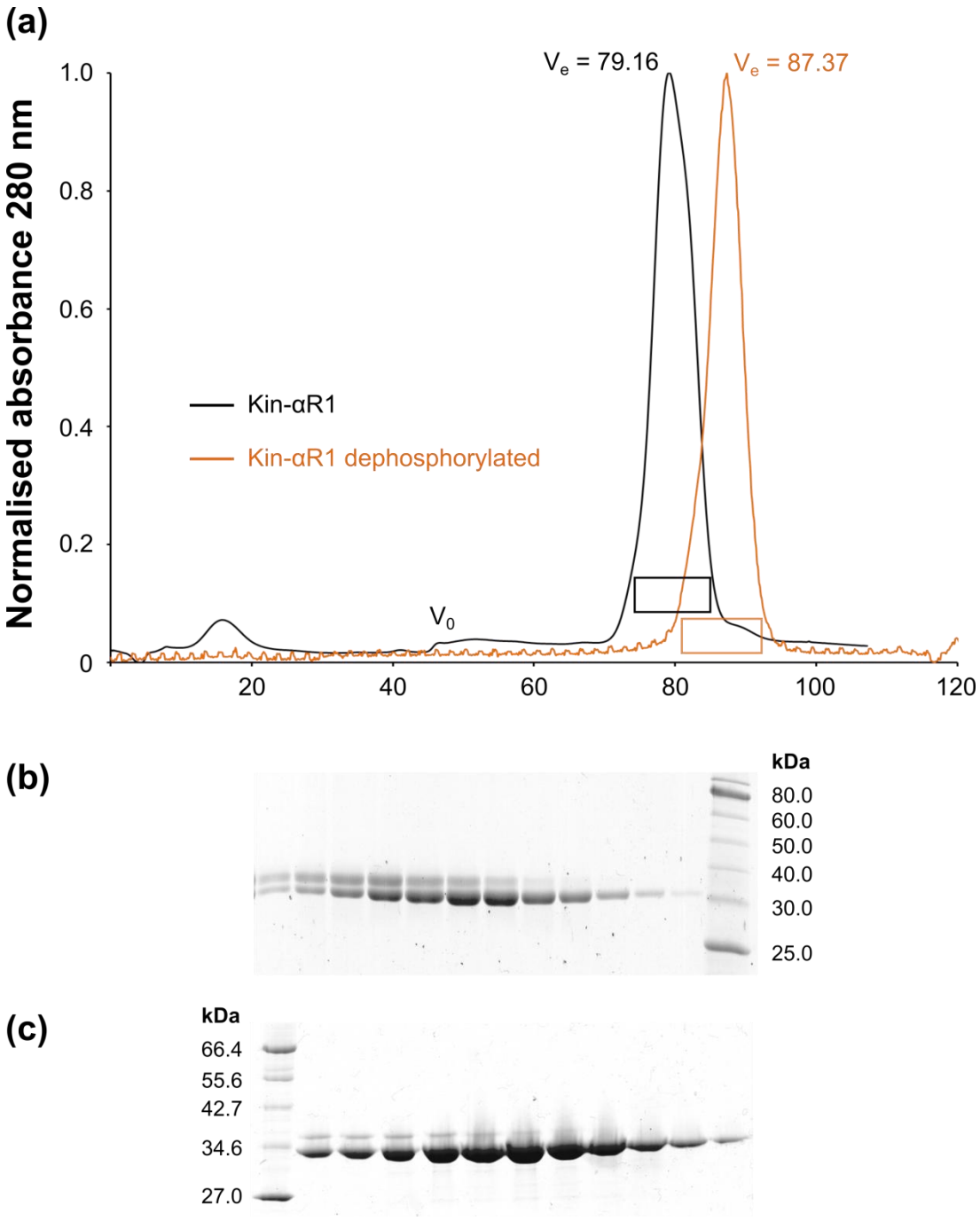


Fig 3.3.1.1 Purfication of recombinantly expressed Kin-αR1 (a) SEC chromatogram of recombinantly expressed Kin-αR1 (black) and Lambda protein phosphatase treated Kin-αR1 (orange) from a Superdex 200 16/60 column. The exclusion volumes (V_e) and column void volume (V_0) are indicated. Boxed regions were assayed in SDS-PAGE (displayed) **(b)** SDS-PAGE for Kin-αR1 and **(c)** Lambda protein phosphatase treated Kin-αR1. Molecular weight markers are shown adjacent (Broad Range 10-250 kDa in **(a)** and Broad Range 2-212 kDa in **(b)**, both from NEB).

3.3.2 Crystallisation of TwcK Kin-αR1 with staurosporine

Previous attempts had been made to crystallise the Kin-αR1 apo-enzyme, but all trials had failed to produce crystals. The innately flexible nature of kinases domains due to the interlobular motions required for catalysis can be an obstacle to overcome in crystallisation. Substrate analogues and inhibitors that can trap the kinase in a closed conformation can assist in crystallisation.

Therefore, crystallisation was here trialled using the dephosphorylated Kin-αR1 sample in complex with the kinase inhibitor, staurosporine. Staurosporine is a broad-range, non-selective protein kinase inhibitor (Meggio *et al.,* 1995) that occupies the ATP binding pocket (Prade *et al.,* 1997) and has been used in facilitating crystallisation of a large number of protein kinases. Examples include the ribosomal kinase RSK1, in which crystallisation with staurosporine was used as a basis to inform specific inhibitor design (Ikuta *et al.*, 2007). LIMK1 (LIM kinase 1), a protein kinase implicated in tumour mestastasis (Yoshioka *et al.,* 2003) was also first solved in complex with staurosporine as a gateway to design of specific ATP-competitive inhibitor compounds (Beltrami *et al.,* 2011).

From the 5 commercial screens set up for Kin-αR1 with staurosporine, two conditions produced crystal hits. Condition A: 0.2 M lithium sulphate, 0.1 M Bis Tris pH 5.5, 25% [w/v] PEG 3350 (JCSG+ Eco, condition 95). Condition B: 0.06 M divalents, 0.1 M buffer 2 pH 7.5, 50% precipitant 1 (Morpheus, condition 5). Crystals in both conditions grew with a thin plate morphology to approx. dimensions of 50x20x10 μm^3 for condition A and 100x40x10 μm^3 for condition B over a time period of three days **(Fig 3.3.2.1).**

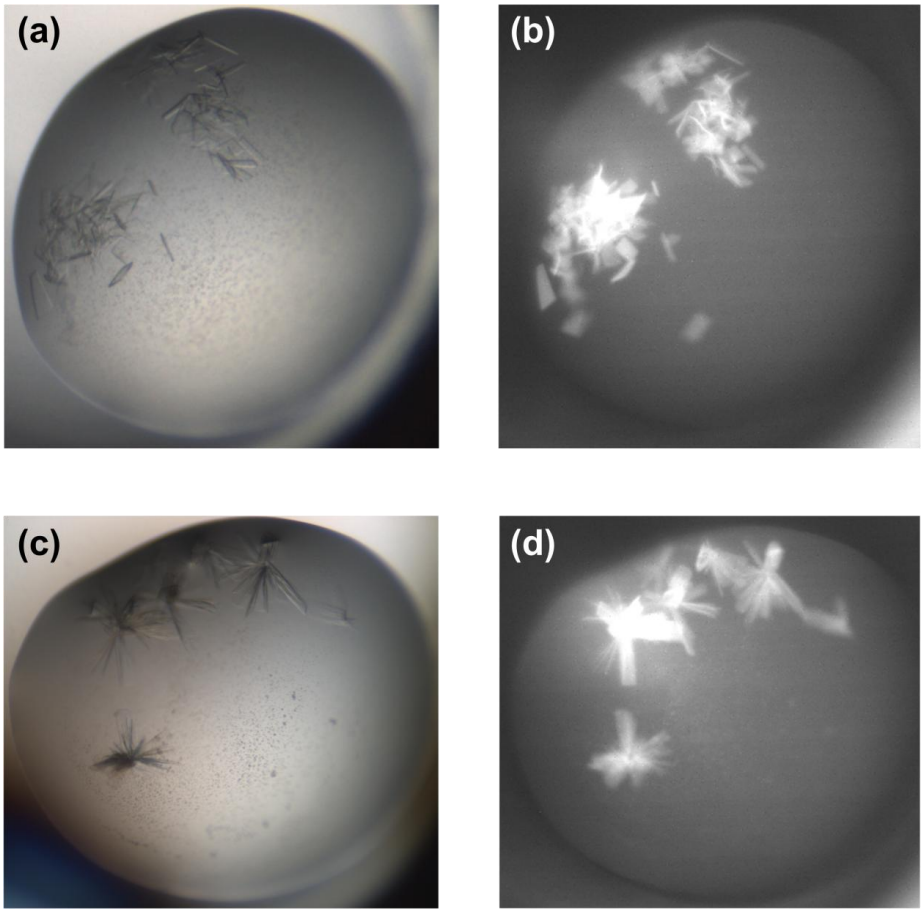


Fig 3.3.2.1 Kin-αR1 crystal hits. (a) Visible light and **(b)** UV images of crystals grown in 0.2 M lithium sulphate, 0.1 M Bis Tris pH 5.5, 25% [w/v] PEG 3350 (JCSG+ Eco, condition 95). **(c)** Visible light and **(d)** UV images of crystals grown in 0.06 M divalents, 0.1 M buffer 2 pH 7.5, 50% precipitant 1 (Morpheus, condition 5). All crystals grew in a thin plate morphology.

Crystals from both conditions were harvested and mounted for X-ray data collection. Upon irradiation, crystals quickly deteriorated and collection of data sets was not possible for any crystals from condition B. A single crystal from condition A was used for structure solution.

3.3.3 Crystal structure of TwcK Kin-αR1 in complex with staurosporine

The crystal structure of Kin-αR1 **(Fig 3.3.3.1a)** was solved to a resolution of 2.75 Å. Data processing statistics and model refinement parameters are given in **Table 3.3.3.1.** Diffraction data was collected to 2.4 Å resolution however for refinement, a resolution cut-off of 2.75 Å was used due to a large decrease in completeness and redundancy of the data beyond this point. The crystal structure presented here is therefore considered to be preliminary and work is ongoing to optimise crystal growth in an effort to improve data quality.

The crystal form in this study contained two molecular copies of Kin-αR1 per asymmetric unit. Staurosporine was bound within the ATP binding pocket of TwcK in both molecular copies **(Fig 3.3.1.b)**. The interactions established by staurosporine with TwcK **(Fig 3.3.3.1c)** involve hydrophobic interactions between the staurosporine indole carbazole rings and the side chain of residues L162, G163, V170, A183, Y230, F232, M284 and I296 in TwcK. Additionally, a hydrogen bond is formed between the hydroxyl group of the central indole carbozole ring and the main chain nitrogen of M233.

Table 3.3.3.1 Kin-αR1 X-ray data processing and model refinement statistics. Values in parentheses are for the highest resolution bin.

Diffraction data		
X-ray source	PXI (PSI)	
Detector	Eiger 16M X (Dectris)	
Wavelength (Å)	1.000	
Space group	P 1 2_1 1	
Unit cell dimensions	a = 71.48 Å	$\alpha = 90.00^o$
	b = 57.67 Å	$\beta = 106.33^o$
	c = 86.09 Å	$\gamma = 90.00^o$
Molecules per a.u/solvent content	2/53%	
Resolution (Å)	29.48-2.75 (2.848-2.75)	
Unique reflections	16898 (1401)	
R_{meas} (%)	9.0 (32.7)	
Redundancy	3.23 (2.15)	
I/σ(I)	11.06 (3.54)	
Completeness (%)	95.29 (80.66)	
Refinement and model statistics		
Reflections in working/free set	16362/505	
R-factor/R_{free} (%)	22.25/27.47	
Protein residues	539	
Protein atoms	4240	
Ligand atoms	70	
Water molecules	24	
rmsd bond/angle (^{o})	0.004/0.980	
Structure validation (Molprobity)		
Ramachandran favoured (%)	91.1	
Ramachandran outliers (%)	2.1	
C-beta outliers	0	
Clashscore	14.06	
Overall score*	2.6	

*Overall score as given by Molprobity denotes the expected resolution to obtain the current model quality. Overall score is expected to be lower than the actual data resolution.

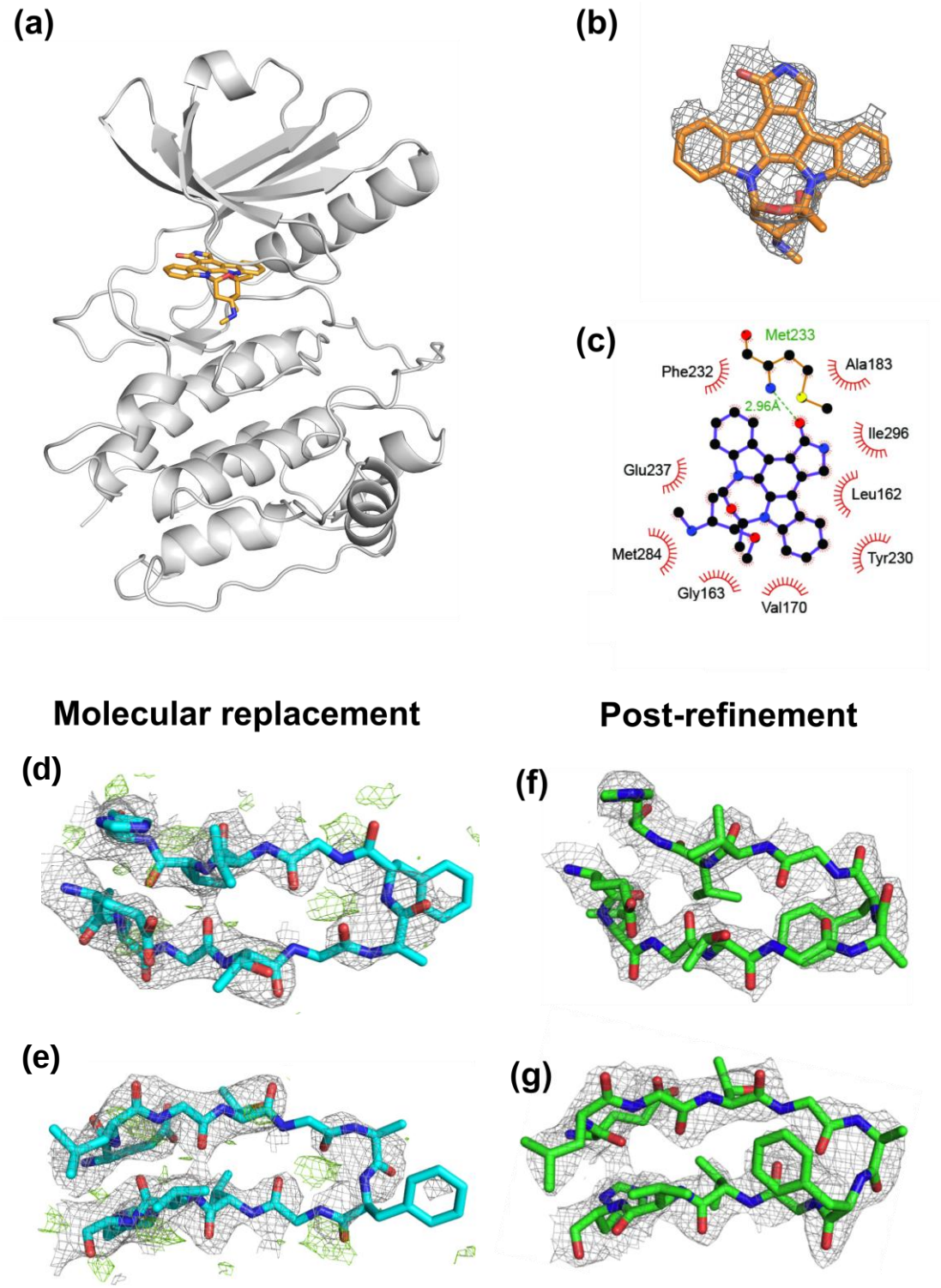


Fig 3.3.3.1 The crystal structure of Kin-αR1 in complex with staurosporine (a) cartoon representation with staurosporine shown as sticks **(b)** | F_o-F_c | density (grey) contoured at 2.5σ before building of staurosporine (orange) **(c)** diagram of TwcK-staurosporine interactions generated with LigPlot+ (Wallace *et al.*, 1994). Hydrophobic interactions are indicated by red lines and hydrogen bonding with green dashed lines. **(d)** | $2F_o$-F_c | electron density map contoured at 1.3σ (grey), | F_o-F_c | density (green) contoured at 2.5σ and glycine-rich loop of TwcK (cyan) countoured at 1.3σ, immediately following molecular replacment and **(e)** rotated 180°. **(f)** | $2F_o$-F_c | electron density map (grey) with glycine-rich loop of TwcK (green) countoured at 1.3σ, after final round of refinement and **(g)** rotated 180°.

The binding of staurosporine appears to be in agreement with other crystal structures of protein kinases in complex with staurosporine **(Fig 3.3.3.2).** Comparison with the crystal structures of DAPK, MAP KAP kinase 2 and PKA (in complex with staurosporine show that staurosporine bound to TwcK has a highly similar orientation of the indole carbazole rings and the TwcK residues interacting with staurosporine are equivalent to those in the aforementioned crystal structures (PDB entries: DAPK, 1WVY; MAP KAP kinase 2, 1NXK; PKA, 1STC).

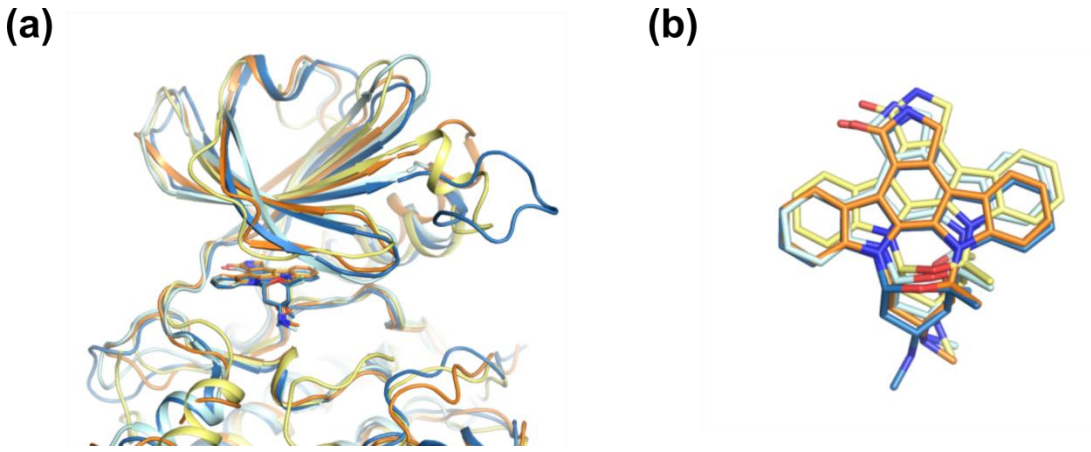


Fig 3.3.3.2 Comparison of staurosporine-bound kinase structures. (a) Superimposed crystal structures of TwcK (orange) DAPK (blue), MAP KAP kinase 2 (cyan) and PKA (yellow) bound to staurosporine showing a shared mode of binding. **(b)** The orientation of bound staurosporine is highly similar for each crystal structure. PDB entries used for figure: DAPK, 1WVY; MAP KAP kinase 2, 1NXK; PKA, 1STC.

The crystal structure of the TwcK-staurosporine complex reveals a closed conformation of TwcK, with the glycine-rich loop seen to close over the active site-cleft **(Fig 3.3.3.3a)**. Subsequently, K185 is now seen to interact with E201 in the αC-helix at a distance of 2.9 Å, forming the polar interaction necessary for coordination of ATP **(Fig 3.3.3.3b)**. This interaction is not in place in the autoinhibited structure of TwcKR, with K185 and E201 at a distance of 4.9 Å. Aside, from the movement within the N-lobe leading to the changes seen in the glycine-rich loop and the Lys185-Glu201 pair, the other catalytic elements of TwcK are seen to undergo little movement. The DFG-motif, catalytic loop, activation loop, and the αC-helix are all seen to occupy highly similar orientations to those seen in the autoinhibited structure of TwcKR **(Fig 3.3.3.3b)**. If the crystal structure of TwcK in complex with staurosporine is representative of the active conformation of TwcK, this suggests that the kinase domain of TwcK does not require significant rearrangement in order to access an active conformation.

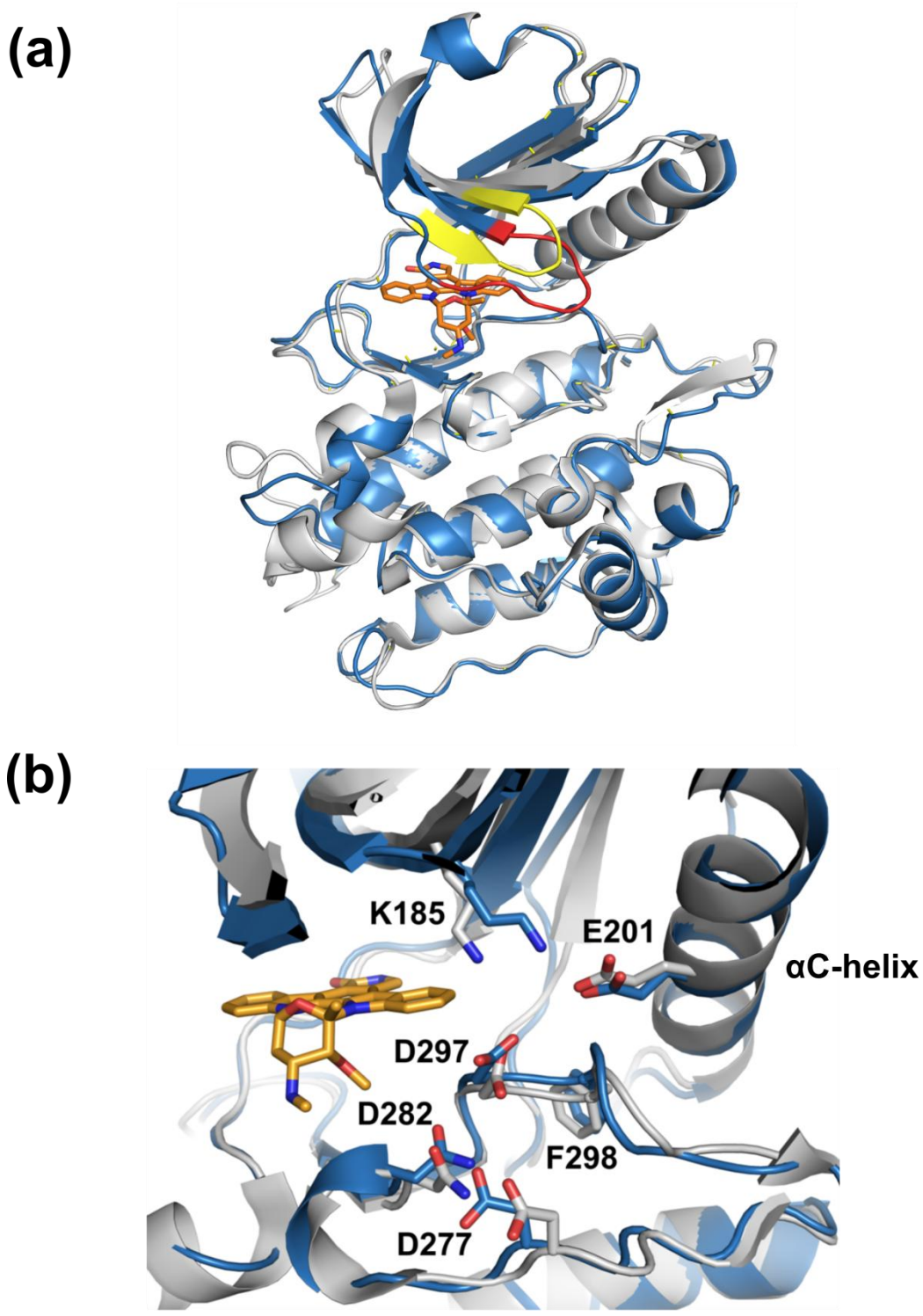


Fig 3.3.3.3 **Comparison of open and closed TwcK crystal structures.** Crystal structures of Kin-αR1 in complex with staurosporine (blue) and the catalytic domain from the crystal structure of the autoinhibited TwcKR (PDB: 3UTO) were superimoposed. **(a)** overall structure showing the closure of the glycine-rich loop in the presence of bound staurosporine (orange). The glycine-rich loop is highlighted in red for Kin-αR1 and yellow for TwcKR **(b)** Close-up view of the TwcK active site showing catalytic residues (sticks). The K185-E201 salt bridge important for ATP coordination is formed in the staurosporine bound structure while the remaining active site residues are seen to undergo little change upon lobe closure.

In order to rule out the effect of contacts within the crystal lattice being responsible for the conformational change observed in the glycine-rich loop, symmetry mates were generated to 50 Å. From this, it can be seen that no crystal contacts are formed by the glycine-rich loop, confirming that the closure of the glycine-rich loop is not an artefact of crystallisation.

3.3.4 The αR1-helix is disordered in the crystal structure of the TwcK-staurosporine complex

The TwcK catalytic domain is mostly well ordered, however the C-terminal sequence, making up the αR1 helix and the preceding helix and loop, was disordered in the crystal structure and not visible in electron density maps from residue N415 onwards. The extent of the disorder with respect to the expected structure based on the crystal structure of the TwcKR is shown in **Fig 3.3.4.1.**

The disordered αR1 helix in the crystal structure of TwcK in complex with staurosporine is a surprising result. The CRD of titin-like kinases is seen to have a stabilising effect, with human TK unable to be produced as a truncated sample that lacks the complete CRD but can be expressed when the αR1 helix is present. Similarly, the recombinantly expressed TwcK kinase domain results in a higher yield of soluble protein and increased stability when it includes the αR1 helix. The disordered αR1 helix in the TwcK-staurosporine structure was therefore unexpected.

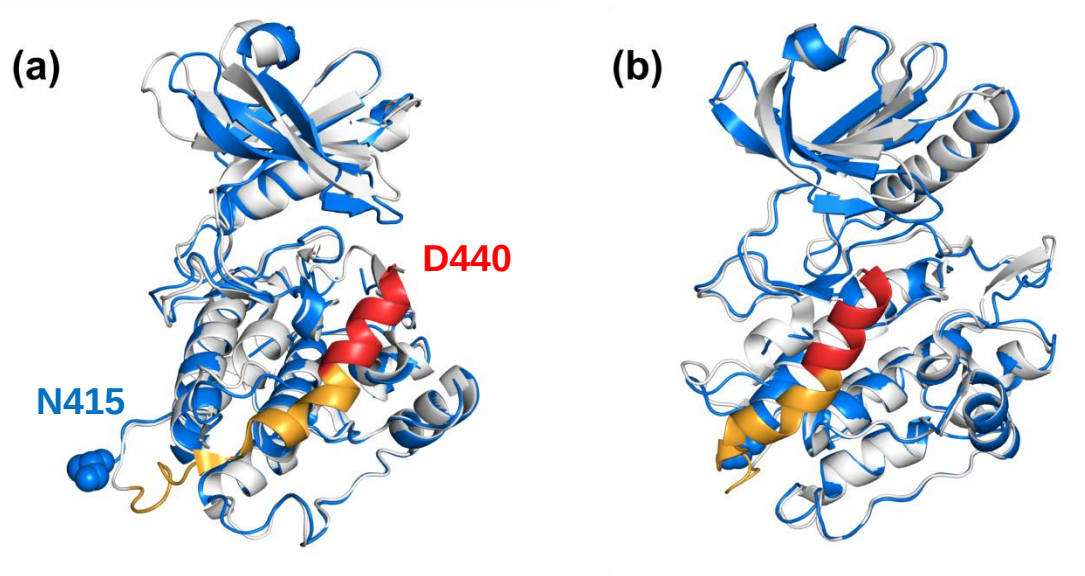


Fig 3.3.4.1 The αR1 helix is disordered in the TwcK-staurosporine crystal structure. Residues N415 to D440 (labelled) in the crystal structure of Kin-αR1 (blue) in complex with staurosporine are disordered and not visible in electron density. The extent of the missing residues are shown by comparison to the catalytic domain of TwcKR (PDB: 3UTO) (white) with regulatory tails removed. **(a)** side view and **(b)** front view. The C-terminal residue of TwcK as visible in the Kin-αR1 structure is shown as spheres, with the C-terminal region of the catalytic domain as visible in the TwcKR structure shown in orange. The additional αR1 helix as included in the Kin-αR1 construct is shown in red.

The fact that the αR1 helix is completely disordered in the TwcK-staurosporine crystal structure indicates that the region is not folded, rather than simply not packing against the catalytic domain. This means that conclusions about the nature of the αR1 helix cannot be drawn in terms of its role as a pseudosubstrate inhibitor.

3.3.5 No inhibition of catalysis is observed in Kin-αR1

The crystal structure of TwcK in complex with staurosporine indicates that the αR1 helix is unlikely to contribute to autoinhibition of TwcK catalysis in the current construct. To validate this observation and determine whether the αR1 helix had an inhibitory effect on catalysis in the Kin-αR1 construct, phosphotransfer assays were performed. The characteristic SDS-PAGE banding pattern of autophosphorylated species was also observed here and it was similar to that of the catalytic domain (Kin) (described in **section 2.3.1).**

The hyper-phosphorylated Kin-αR1 species (as judged by SDS-PAGE) showed very low levels of catalysis (approximately 1% maximal activity). The minimally phosphorylated Kin-αR1 species (as judged by SDS-PAGE) displayed levels of catalysis that were essentially identical to that of the Kin control (approximately 96%). Results of phosphotransfer assays on Kin-αR1 **(Fig 3.3.5.1)**, therefore show that the Kin-αR1 construct is not catalytically inhibited by the presence of the αR1 helix from the CRD. This result is in agreement with the crystal structure that reveals that the αR1 helix is not natively folded and not bound to the catalytic domain of TwcK, therefore being unable to participate in autoinhibition.

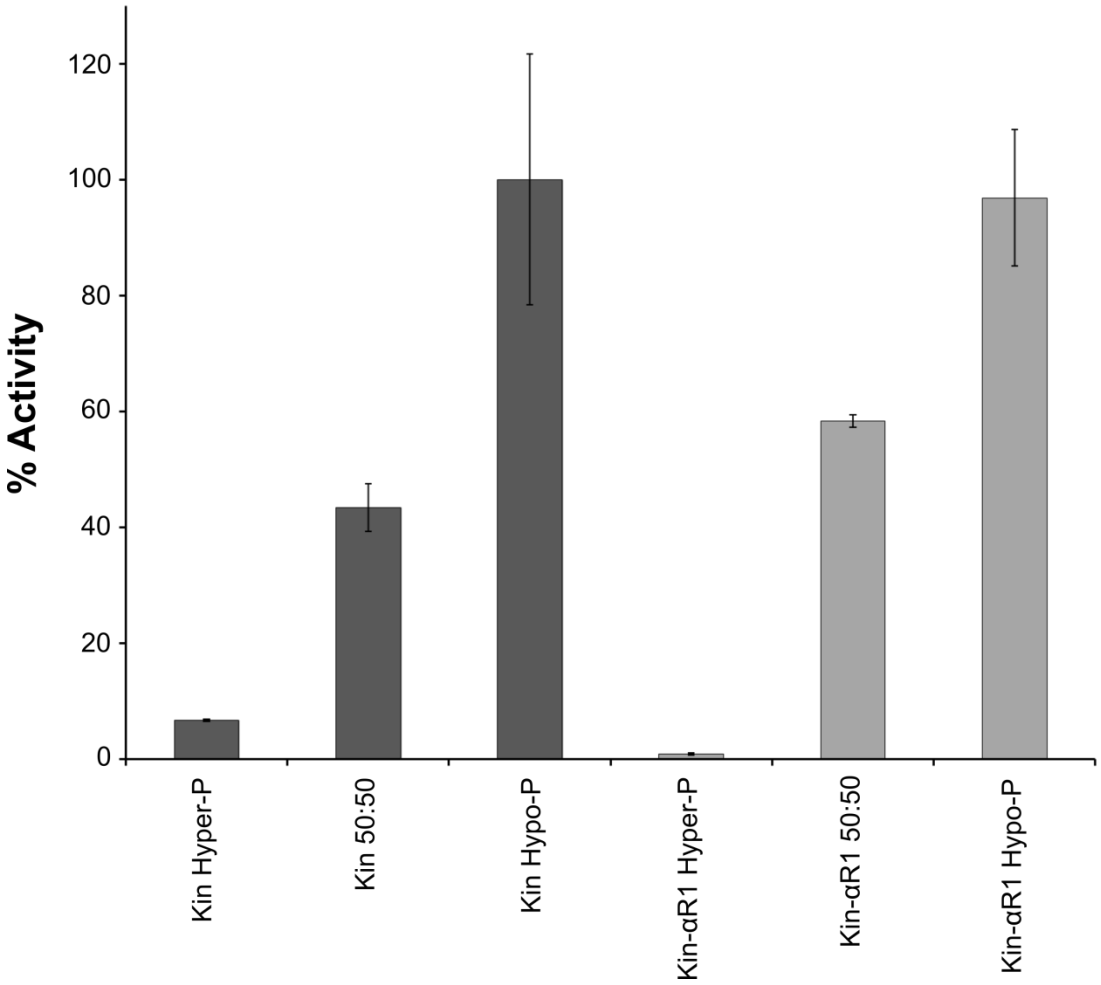


Fig 3.3.5.1 Phosphotransfer assays on Kin and Kin-αR1. The TwcK catalytic domain (Kin) was used as a positive control and designated as 100% activity, with the activity of Kin-αR1 given as % of Kin activity. Hyper-P denotes the hyper-phosphorylated species, Hypo-P the hypo-phosphorylated and 50:50 an approximately equal mixture of phosphorylation states. Kin-αR1 has essentially identical levels of catalysis to Kin, with maximal activity at 96%, showing that the αR1-helix is not inhibitory in the Kin-αR1 construct. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

3.3.6 Myosin light chain 4 peptide from *C. elegans* is a poor substrate for TwcK

In order to describe the function of TwcK *in vivo* and the cellular pathway in which it participates, it is necessary to identify a physiological substrate. With the

improvement in sequencing coverage of the *C. elegans* genome came the identification of additional myosin RLCs (The *C. elegans* sequencing consortium). Through searching for sequences similar to the MLCK consensus substrate in the Wormbase repository for the *C. elegans* genome, myosin light chain-4 (mlc-4) was identified as containing a sequence at its N-terminus with high sequence similarity to the current TwcK model peptide substrate and most importantly, to the cognate MLCK substrate (**Table 3.3.6.1**). Crucially, mlc-4 has been shown to be phosphorylated at T17 and S18, which is required for myosin activity (Gally *et al.*, 2009). We therefore identified the N-terminus of *C. elegans* mlc-4 as a good candidate for a physiological TwcK substrate. With TwcK associating with thick filaments and proposed as a potential MLCK paralogue in invertebrates (**section 3.1.3)**, the identification of a potential TwcK substrate in mlc-4, a constituent protein of the sarcomere, presents a good candidate for a functionally relevant physiological substrate for TwcK in *C. elegans*.

Phosphotransfer assays were conducted in order to assess the ability of TwcK to phosphorylate a peptide substrate derived from *C. elegans* mlc-4 (residues 11-22) **(section 3.2.5)**. Phosphotransfer assays were carried out using the Kin construct, following the same protocol as all previously conducted TwcK assays. Assays revealed that catalytic activity in the presence of *C. elegans* mlc-4 derived peptide substrate was vastly reduced, with approximately 4% of the activity of the model substrate **(Fig 3.3.6.1)**.

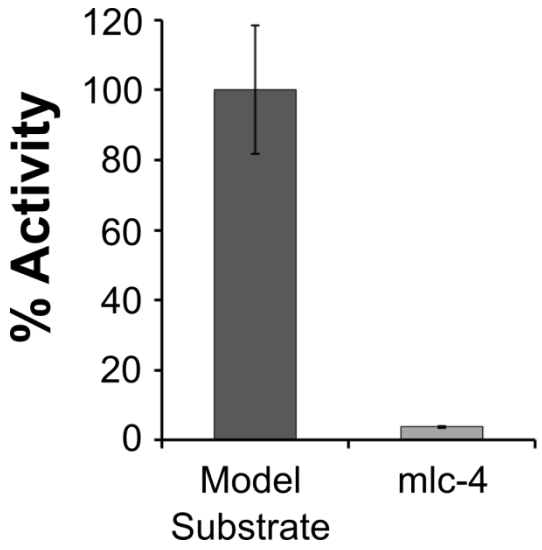


Fig 3.3.6.1 Phosphotransfer assays on Kin with model peptide substrate and mlc-4 derived peptide substrate. Activity was reduced to approx. 4% in samples containing the mlc-4 substrate compared to the model substrate. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

These data show that TwcK phosphorylates the mlc-4 peptide at considerably lower levels than the model peptide substrate. In order to rationalise this decreased activity, a sequence comparison of the peptide substrates from MLCK, *Aplysia* TwcK, the *C. elegans* TwcK model substrate and the mlc-4 derived substrate was carried out (**Table 3.3.6.1**). In all of the peptide substrates, an ATSNVF motif is conserved. In this ATSNVF, the Thr and Ser residues correspond to the phosphorylated Thr18 and Ser19 of the MLCK substrate (rMLC II) and Thr15, the rMLC phosphorylation site for TwcK in *Aplysia*. Arginine is also seen at the P-5 position. The specificity of MLCKs towards phosphorylation of Thr18 or Ser19 has been shown to be strongly dependent on the position of the substrate P-3 arginine residue (Kemp & Pearson, 1985). With this arginine located at position 16 of the peptide substrate **(as in Fig 3.3.6.2)**, the phosphorylated residue is Ser19. If Arg16 is

located instead at residue 15, specificity is seen to shift to phosphorylation of Thr18 (Kemp & Pearson, 1985).

Table 3.3.6.1 TwcK and MLCK peptide substrate comparison. The conserved ATSNVF motif is in bold with phosphorylatable residues within this sequence highlighted in red. Confirmed *in vivo* phosphorylation sites are underlined: Thr15 for *Aplysia* rMLC and Thr18/Ser19 for vertebrate MLCK substrate. P-5 and P-2/23 Arg are shown in blue. Regions of identical sequence between the *C. elegans* mlc-4 and vertebrate rMLC II substrates are boxed.

Substrate name	Sequence
kMLC 11-23 TwcK model substrate	KK**R**A**R**A**ATSNVF**S
C. elegans mlc-4 substrate	RQ**R**PQ**R****ATSNVF**A
Aplysia TwcK myosin RLC substrate	KKG**R**SG**R****A**<u>**T**</u>**SNVF**A
Vertebrate myosin RLC II substrate	KK**R**PQ**R****A**<u>**T**</u>**SNVF**S

Several substrate preferences have been identified for TwcK from both *C. elegans* and *Aplysia.* Disruption of the basic cluster of residues at the N-terminus of the peptide substrate has little effect on catalysis (Heierhorst *et al.,* 1996b), in contrast to MLCK in which this basic cluster is required for maximal catalysis (Kemp & Pearson 1985). TwcK from *C. elegans* and *Aplysia* both have a preference for threonine phosphorylation as well as for arginine at the P-3 position (Heierhorst *et al.,* 1996b), with the ideal phosphorylation motif identified as RxφT, where x is any amino acid and φ is hydrophobic (Heierhorst *et al.,* 1996b). Additionally, the TwcK autophosphorylation sites identified in this work (**section 2.3.4)** are all at threonine residues, further indicating a preference for threonine at P0.

In the model substrate used for TwcK catalytic assays, phosphorylation at threonine would mean that the P-3 position contains arginine, fulfilling the reported TwcK substrate specificity (Heierhorst *et al.,* 1996b). For threonine phosphorylation in the mlc-4 peptide, this arginine would instead be located at the P-2 position. Two possibilities therefore may explain the reduced activity shown by *C. elegans* TwcK towards the mlc-4 peptide substrate:

i) In order to position arginine at P-3, specificity could switch to phosphorylation of serine, for which TwcK shows a reduced preference and therefore lower levels of phosphorylation are observed.

ii) The preference of TwcK for threonine at P0 could be stronger than its preference for arginine at P-3. In this scenario, TwcK could continue to phosphorylate threonine but due to the shift of arginine from P-3 to P-2, substrate affinity would be reduced, leading to lower levels of catalysis.

The P-2/-5 substrate binding pocket as described for Ser/Thr kinases (Zhu *et al.,* 2005; Ben-Shimon & Niv, 2011; Gerlits *et al.,* 2013) contains isoleucine (I345) in *C. elegans* TwcK and valine (V242) in *Aplysia* TwcK. In MLCK, methionine is located at this position. While all three kinases show some preference for arginine at the P-3 position, both *Aplysia* TwcK and MLCK phosphorylate substrates *in vivo* containing arginine at P-2 and all substrates contain arginine at P-5. With the hydrophobic side-chains of these residues likely to be incompatible with arginine binding in this region, this suggests that the substrate binding mode of TwcK and MLCK may not utilise the same binding pockets as have been previously described. The crystal structure of TwcK in complex with a peptide substrate is therefore

required to decipher the structural determinants of substrate specificity in these kinases.

3.3.7 TwcK catalysis is Mg^{2+} dependent

To fully investigate the substrate binding of TwcK, the crystal structure of TwcK in complex with staurosporine can be used as a basis from which to pursue further crystallisation experiments. In particular, the crystal structure of TwcK in complex with its peptide substrate would provide highly useful information regarding the substrate-binding mode of TwcK, allowing us to more productively explore the substrate-specificity of TwcK based on structural features of the peptide-binding site. Although the crystal structure of TwcK in complex with staurosporine has now been solved, we also sought additional possibilities for co-crystallisation of TwcK in complex with both its ATP and peptide substrates whilst preventing catalysis. For this, possibilities include the use of non-hydrolysable ATP analogues as well as the available TwcK K185A mutant, in which mutation of the AxK lysine in the β3 strand renders TwcK catalytically inactive **(section 2.3.1)**.

A factor to consider for crystallisation of protein kinases in complex with nucleotide substrates is the presence of divalent metal ions. The majority of protein kinases require divalent metal ions in order to coordinate the ATP substrate and perform phosphotransfer. The requirement of protein kinases is largely for Mg^{2+} although a small number of kinases have been shown to bind ATP in the presence of other metal ions such as Mn^{2+} or even without a metal counterion (Bailey *et al.*,

2015). The question of what the preference of TwcK is towards metal ions is an important one in guiding further crystallisation of TwcK.

To test the metal dependency of TwcK, phosphotransfer assays were performed using the minimally autophosphorylated TwcK kinase domain. Assays were set up containing Mg^{2+} (as in previous assays), Mn^{2+} and in the absence of divalent metal ions. Phosphotransfer assays revealed that TwcK had a strong preference for Mg^{2+} over Mn^{2+}, with activity in the presence of Mn^{2+} ions reduced to approx. 4% compared to with Mg^{2+} **(Fig 3.3.7.1)**. In the absence of divalent metal ions, TwcK was seen to have only residual activity (<1%). Results of phosphotransfer assays to test for metal dependency therefore show that TwcK catalysis is dependent on the presence of Mg^{2+} ions and will likely be required for future work to elucidate the crystal structure of TwcK in complex with its substrates. It should however be noted that lack of catalysis in the presence of Mn^{2+} or in the absence of metal ions might indicate a decreased affinity for ATP. Further experiments such as isothermal titration calorimetry (ITC) could be used in order to properly explore the metal dependency of ATP binding in TwcK.

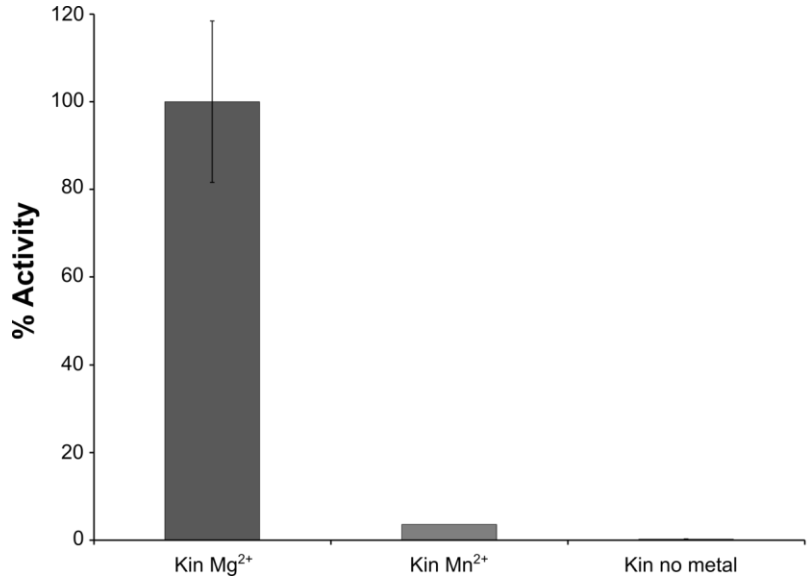


Fig 3.3.7.1 Phosphotransfer assays on TwcK with Mg^{2+}, Mn^{2+} and in the absence of divalent metal ions. Assays show that TwcK catalysis is Mg^{2+} dependent. Kin-Mg^{2+} was seen to have maximal activity and is designated as 100%, with 4% of maximal activity seen in the presence of Mn^{2+} and only residual activity (<1%) with no divalent metal ions added. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

3.4 Discussion

In this work, we have shown the crystal structure of the TwcK catalytic domain in complex with staurosporine, the first structure of a titin-like kinase in a closed conformation. Whilst the TwcK conformation seen here shows closure of the glycine rich loop over the active site cleft, this is not necessarily indicative of an active protein kinase conformation. Crystallisation of PKA with staurosporine was shown to induce a conformation that is intermediate between the open, inactive conformation and the closed, active conformation (Prade *et al.,* 1997). In this intermediate conformation, while the active site was assembled, the N-lobe was seen to be in a more open conformation compared to the conformation of PKA in complex with ADP, ATP and ATP analogues (Prade *et al.,* 1997).

To determine whether the conformation seen in the crystal structure of TwcK in complex with staurosporine is likely to be representative of the active conformation, the TwcK-staurosporine crystal structure was compared to that of Death-activated protein kinase (DAPK), a kinase of the Ca^{2+}/calmodulin-regulated family that has 38% sequence identity to the TwcK catalytic domain. Crystal structures of DAPK are available in complex with ADP (PDB: 3F5U) (McNamara *et al.,* 2009), AMP-PNP (PDB: 3EHA) (McNamara *et al.,* 2009) and staurosporine (PDB: 1WVY) and were structurally superimposed along with the crystal structure of TwcK-staurosporine **(Fig 3.4.1)**. It has previously been shown that conformational changes upon nucleotide binding in DAPK are localised to the glycine-rich loop (McNamara *et al.,* 2009). The conformation of DAPK in complex with staurosporine is seen to be the same as the conformation in the presence of nucleotides, with the glycine-rich loop occupying the same position and a fully assembled active site.

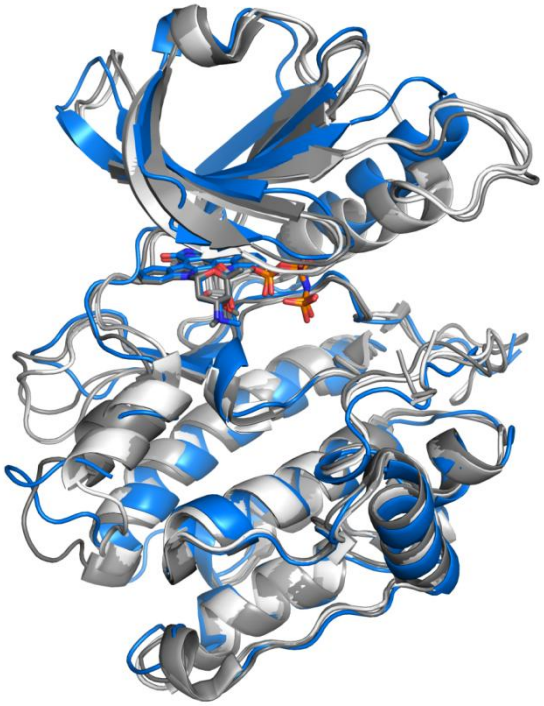


Fig 3.4.1 The crystal structure of TwcK-staurosporine adopts an active conformation. Staurosporine-bound TwcK (blue) and DAPK (grey) in complex with ADP (PDB: 3F5U), AMP-PNP (PDB: 3EHA) and staurosporine (PDB: 1WVY) with the TwcK-staurosporine complex. Changes in all structures upon nucleotide or inhibitor binding are localised to the glycine-rich loop which adopts a highly similar position in all, suggesting that the conformation of TwcK in complex with staurosporine is representative of the active conformation.

The crystal structure of TwcK-staurosporine is seen to adopt a highly similar conformation to the nucleotide bound and staurosporine bound structures of DAPK. With conformational changes localised to the glycine-rich loop in both DAPK and TwcK, it would appear that the crystal structure of TwcK bound to staurosporine is representative of the active TwcK conformation. Although the αC-helix of TwcK is in a raised position relative to its DAPK counterpart, the K185-E201 salt bridge that is crucial for phosphotransfer is in place, indicating that the staurosporine bound conformation of TwcK is likely to be a catalytically active conformation.

The crystal structure of the TwcK catalytic domain revealed the closure of the glycine-rich loop over the active site cleft as the only apparent conformational change undergone by TwcK in order to access the active conformation. The lack of apparent conformational changes in the other catalytic elements confirms that the crystal structure of the autoinhibited TwcKR (PDB:3UTO) (von Castelmur *et al.,* 2012) shows TwcK in a conformation in which significant rearrangement of the catalytic machinery is not required to reach a catalytically productive conformation. This appears to be in agreement with the model of mechanical activation proposed by von Castelmur *et al*., in which removal of the NL from the interlobular hinge region by mechanical force leaves TwcK in a mechanically primed state (von Castelmur *et al*., 2012).

The αR1 helix of TwcK is expected to remain bound to the kinase domain, providing a buffer against mechanical stretch and helping to maintain the stability of the TwcK catalytic domain, which may not be able to function under mechanical deformation (von Castelmur *et al.,* 2012). The hypothesis of the αR1 helix as a mechanical buffer was an appealing one due to the mechanical regime of helices. The main chain hydrogen bonds within helices are aligned with the helical axis, which itself is often aligned with the force vector during mechanical stretch. When stretched, helices can therefore undergo longitudinal shearing. This longitudinal shearing means that the force required to unfold a helix is equivalent to one hydrogen bond multiplied by the number of hydrogen bonds within the helical motif. Helices thereby act as constant force springs and can undergo force-induced folding-refolding transitions, contrary to Hooke's law which describes force rising with extension in entropic springs (Schwaiger *et al.,* 2002; Berkmeier *et al.,* 2011; Wolny *et al.,* 2014). The αR1 helix was however not visible in the crystal structure of TwcK

in complex with staurosporine, having not folded into a helix. We were therefore unable to conclude on whether the αR1 helix is compatible with substrate binding or whether it contributes to autoinhibition through a possible pseudosubstrate mechanism as seen in the C-terminal extension of MLCK (Knighton *et al.,* 1992; Barden *et al.,* 1996).

The unfolded αR1 helix does shed light on previous work in which the final 24 residues of the CRD were claimed to be responsible for autoinhibition (Lei *et al.,* 1994), thereby dismissing the N-terminal region of the CRD as autoinhibitory and with it the likelihood of a pseudosubstrate mode of inhibition as seen in MLCK. Our data show that truncation of the CRD can lead to unfolded structural elements. It is therefore likely that the truncated CRD constructs used to determine the autoinhibitory region of the TwcK CRD (Lei *et al.,* 1994) contained an unfolded or improperly folded CRD, explaining why autoinhibition was only observed in the presence of the complete CRD.

In working towards identification of a physiological substrate for *C. elegans* TwcK, an mlc-4 derived peptide substrate was tested in phosphotransfer assays. Results showed that this mlc-4 peptide is a poor substrate for TwcK *in vitro*, likely due to the presence of an arginine at the P-2 position being poorly compatible with the substrate specificity of *C. elegans* TwcK. While the activity towards the mlc-4 peptide is non-negligible but low, high levels of activity are thought to be a key trait of TwcK in its *in vivo* function. TwcK is embedded in the semi-crystalline lattice of the muscle sarcomere in invertebrates and rather than being diffusible within the cell is evenly distributed throughout the thick filaments at a ratio of approximately one TwcK per 50 myosin molecules (Moerman *et al.,* 1988; Vibert *et al.,* 1993). It has therefore been proposed that attachment to the thick filament in this manner would

likely mean only a relatively small number of myosin RLC molecules would be accessible to each TwcK domain, requiring high catalytic activity for phosphorylation under short bursts of activation (Lei *et al.*, 1994). The low level of phosphorylation seen on the mlc-4 peptide substrate suggests that mlc-4 may not be a suitable target for phosphorylation by *C. elegans* TwcK *in vivo*.

In work carried out on the *in vivo* function of TwcK catalysis in *C. elegans*, TwcK activity in body wall muscle was shown to inhibit the rate of contraction/relaxation, causing an increased speed of movement (Matsunaga *et al.*, unpublished - **section 7.1.2**). In body wall muscle of *C. elegans*, mlc-1 and mlc-2 isoforms are expressed that do not contain a sequence that bears any resemblance to the TwcK phosphorylation motif. Mlc-4 is not expressed in body wall muscle but is expressed in pharyngeal muscle (G. Benian, personal communication) where mutants lacking TwcK activity were shown to have an increased rate of contraction/relaxation, opposite to the effect of loss of TwcK activity in body wall muscle (Matsunaga *et al.*, unpublished - **section 7.1.2**). TwcK may therefore have a dual role in regulation of muscle contraction/relaxation in *C. elegans* that appears to correlate with expression of mlc-4. The direct presence of twitchin in pharyngeal muscle is however yet to be confirmed, so knockout of TwcK activity might have an indirect effect on pharyngeal muscle, perhaps due to a developmental effect. The possibility therefore remains that mlc-4 could be one of more possible *in vivo* substrates for *C. elegans* TwcK despite relatively low levels of phosphorylation *in vitro* but further work is required to identify the *in vivo* TwcK substrate(s).

3.5 Future perspectives

Further work is needed to determine the relevance of mlc-4 as a possible physiological TwcK substrate. In this work, it is shown that an mlc-4 derived peptide is phosphorylated at low levels by TwcK *in vitro*. Firstly, a true negative control is needed for phosphotransfer assays to ascertain whether the observed phosphorylation of mlc-4 by TwcK is genuine or simply due to promiscuous activity by TwcK. Secondly, short peptides may not represent the fold seen in the context of the native protein from which they are derived (Ho & Dill, 2006), which could lead to insolubility, aggregation or poor substrate recognition of the peptide. Phosphotransfer assay conditions could be optimised, for example through use of small amounts of detergent to aid with solubility and aggregation as well as varying the length of the mlc-4 peptide substrate. Establishing recombinant expression of the full length mlc-4 protein would provide an excellent way to study the potential of mlc-4 as a TwcK substrate.

With crystallisation of the TwcK catalytic domain in its active conformation now established, we hope to utilise this knowledge to further explore the substrate binding properties of TwcK. For this, peptides of varying length will be designed based on the TwcK model peptide substrate for use in co-crystallisation experiments. The crystal structure of TwcK in complex with its peptide substrate would allow for a clear understanding of the substrate specificity of TwcK by revealing the mode of peptide binding and the peptide binding pockets that confer specificity, informing further work on identification of a physiological TwcK substrate. Additionally, revealing the substrate binding mode of TwcK would help significantly to clarify

whether the αR1 helix of the CRD is compatible with substrate binding or whether it acts as a pseudosubstrate as seen in MLCK.

Chapter 4

Using DNA as a molecular spring to study twitchin kinase mechanoactivation

In this chapter, we construct a twitchin kinase:DNA chimera, establishing the use of DNA as a molecular spring to study twitchin kinase mechanoactivation and its ability to perform catalysis under mechanical deformation. This work provides a method for directly monitoring kinase catalysis while under mechanical stretch, a key requirement for exploring the mechanosensory properties of titin-like kinases.

4.1 Introduction

4.1.1 Mechanoactivation of TwcK

Titin-like kinases have been shown to be autoinhibited by regulatory tails that appear to be mechanically labile according to MD simulations. With titin-like kinases shown to be involved in mechanotransduction in the sarcomere (Lange *et al.*, 2005; Butler & Siegman, 2011) and the identity of *in vivo* activators remaining elusive, it was hypothesised that titin-like kinases would be directly activated by tension in the sarcomere (Gräter *et al.*, 2005; Puchner *et al.*, 2008; Puchner & Gaub, 2010; Stahl *et al.*, 2011; von Castelmur *et al.*, 2012). Stretch unfolding of the CRD was thought to be the likely mechanism for mechanoactivation across the titin-like family, however the identification of the mechanically labile NL in TwcK (von Castelmur *et al.*, 2012) revealed an alternative mechanism of mechanoactivation.

Thus, activation mechanisms throughout the titin-like family of kinases are likely to vary, based on the composition of these regulatory tails (Mayans *et al.,* 2013).

Mechanoactivation in titin-like kinases has been been studied by AFM and MD simulations. In the case of human TK, AFM studies and force-probe MD simulations suggest that the CRD can be removed from the active site by stretch-induced unfolding, with ATP binding occurring in the stretched TK (Gräter *et al.,* 2005; Puchner *et al.,* 2008; Puchner & Gaub, 2010; Stahl *et al.,* 2011). In the case of TwcK and TTN1 (titin 1) kinases from *C. elegans*, stretch-induced unfolding of the CRD is not observed in AFM data due to technical limitations, so that it remains unknown whether the CRD can undergo stretch-induced unfolding in these kinases (Greene *et al.,* 2008).

The crystal structure of the *C. elegans* TwcKR and steered MD simulations indicated that the NL is mechanically labile, whilst the force required to mechanically remove the CRD from the fully autoinhibited state would be sufficient to compromise the structural integrity of the catalytic domain (von Castelmur *et al.,* 2012). With the NL also seen to interact with the CRD in the ATP-binding pocket of TwcK, a model of mechanoactivation has been proposed in which mechanical force leads to the removal of the NL. Subsequently, it was suggested that rather than being mechanically removed, the portion of the CRD that occupies the TwcK active site and folds into an energetically unfavourable 3_{10} helix, could be out-competed by sufficient substrate concentration in the absence of its interactions with the NL, possibly inducing a local unfolding-folding transition of the CRD (von Castelmur *et al.,* 2012; Mayans *et al.,* 2013). In this model, the CRD could remain bound to the C-lobe of the catalytic domain of TwcK, potentially acting as a buffer against mechanically induced damage to the catalytic domain (von Castelmur *et al.,* 2012;

Mayans *et al.*, 2013). The TwcK CRD not requiring mechanical removal was in keeping with the observation that the presence of the CRD alone (Kin-CRD construct) is permissive to significant levels of catalysis (~50% of maximal activity) (von Castelmur *et al.*, 2012).

In the previous chapter of this thesis, the crystal structure of the TwcK catalytic domain in complex with staurosporine in its closed conformation is presented. The construct used for crystallisation comprised the TwcK catalytic domain as well as the αR1 helix of the CRD (Kin- αR1). The absence of the αR1 helix in the crystal structure of TwcK catalytic domain suggests that in previous catalytic assays involving the Kin-CRD construct, the CRD may be misfolded or not bound to the TwcK catalytic domain, accounting for the measured catalysis.

Mechanoactivation provides an attractive model for regulation of the titin-like kinases within the context of mechanical tension in the sarcomere. However, the idea of an active site working under mechanical deformation is in many ways contrary to fundamental enzymatic principles, i.e. an active site formed by a network of specifically interacting residues, the precise spatial arrangement of which is critical for catalysis. In protein kinases, catalysis relies on a number of residues, the function and location of which are conserved throughout protein kinases **(reviewed in chapter 2)** (Knighton *et al.*, 1991; Madhusudan *et al.*, 1994; Jeffery *et al.*, 1995; Matte *et al.*, 1998; Kornev *et al.*, 2006; Steichen *et al.*, 2009).

Furthermore, protein kinases catalyse a bi-substrate reaction, with lobe closure upon ATP binding required for transfer of the γ phosphate from ATP to the peptide substrate (Li *et al.*, 2002; Nolen *et al.*, 2003). The precise network of interactions and conformational motions required for phosphotransfer by protein

kinases therefore make the prospect of a functional protein kinase active site under mechanical deformation a striking one. Indeed, mechanical inhibition of enzymatic activity through the use of DNA springs has been demonstrated for PKA (Choi & Zocchi, 2006) as well as for guanylate kinase (Choi *et al.*, 2005; Joseph *et al.*, 2014) and luciferase (Tseng & Zocchi, 2013).

In order to undergo mechanoactivation, TwcK must therefore be able to efficiently perform catalysis while under stretch-induced mechanical deformation. With the NL of TwcK predicted through MD simulations to be highly mechanically labile (von Castelmur *et al.*, 2012), this study aims to explore the hypothesis that stretch-induced removal of the NL can free the active site of TwcK, leading to a kinase that is catalytically active under mechanical deformation.

4.1.2 Using TwcK:DNA chimeras to study mechanoactivation

The study of molecules under stretch is most commonly carried out using AFM and Optical Tweezers (OT). While powerful techniques in their own right, a number of inherent limitations mean that an alternative technique was required for investigating TwcK catalysis under stretch.

A major limitation of AFM, is its lack of reliability for measurements below approx. 10 pN due to the large size and stiffness of the cantilevers to which the AFM tip is attached, as well as the effect of thermal motions (Neuman & Nagy, 2008). With the possibility of the TwcK NL unfolding at lower forces than those measurable by AFM (von Castelmur *et al.*, 2012), a technique is required that can be reliably used to study the effect of lower forces on TwcK.

The use of optical tweezers, while theoretically capable of measuring forces below the limit of AFM (> 0.1 pN), also has significant drawbacks when applied to the study of proteins. Firstly, high resolution optical trapping requires highly pure samples at a level of homogeneity not always achievable when working with protein samples (Neuman & Nagy, 2008). Secondly, significant heating of the sample can occur due to the intensity of the laser used (Seol *et al.,* 2006), a significant issue when studying enzymatics.

Crucially for the study of mechanical activation of TwcK, none of the aforementioned techniques can be used to apply mechanical force whilst simultaneously monitoring catalysis. Having been previously used to mechanically control enzymatic activity, we identified the use of DNA as a molecular spring through protein:DNA chimeras (Choi *et al.,* 2005; Choi & Zocchi, 2006; Tseng and Zocchi, 2013; Joseph *et al.,* 2014) as a technique to explore the hypothesis of the NL as the primary mechanosensory component of TwcK and the ability of TwcK to maintain catalysis while subject to mechanical force. This represents the first use of protein:DNA chimeras to study activation, rather than inhibition of enzymatic activity under stretch.

Mechanical control of enzymatic activity through the use of protein:DNA chimeras has been developed by Prof. Giovanni Zocchi, UCLA, and relies on the principle that proteins are inherently pliable molecules that can be deformed by mechanical stress (Zocchi, 2009). Importantly, the use of protein:DNA chimeras is a 'batch' technique (rather than single molecule) that is carried out entirely in solution, allowing for the use of physiological buffer conditions and the direct monitoring of catalysis through *in vitro* assays. The ability to quantitate catalysis during stretch is

fundamental to this work and is not possible through any other experimental technique.

In this method, a loop of DNA is attached to the protein of interest and used as a molecular spring to generate mechanical force. The mechanical force is directed to specific sites within the protein and is well established as a method of allosteric control of enzymatic activity (Choi *et al.,* Zocchi, 2005; Choi & Zocchi, 2006; Zocchi, 2009; Tseng & Zocchi, 2013; Joseph *et al.*, 2014).

The construction of protein:DNA chimeras involves attachment of single stranded DNA arms to engineered cysteines within the protein via a hetero-bifunctional chemical cross linker. Individual arms are attached and ligated to form a single stranded spring. Due to the persistence length of ssDNA (approx. 1 nm, or 3 bp) (Tseng *et al.,* 2009), the ssDNA spring remains flexible and does not exert tension on the protein. The ssDNA spring is then hybridised with a complementary DNA strand to form a dsDNA spring. The rigidity of dsDNA is significantly higher than ssDNA with a theoretical persistence length of approx. 50 nm (150 bp) (Borochov *et al.,* 1981; Tseng *et al.,* 2009). A reduction in persistence length and thus rigidity has however been described for dsDNA in increasing NaCl concentrations (Borochov *et al.,* 1981). The rigid dsDNA spring undergoes bending due to its covalent attachment at fixed points on the protein surface, exerting mechanical force on the protein (Zocchi, 2009; Tseng *et al.,* 2009). The concept of a protein:DNA to exert force on a protein is illustrated in **Fig. 4.1.2.1.**

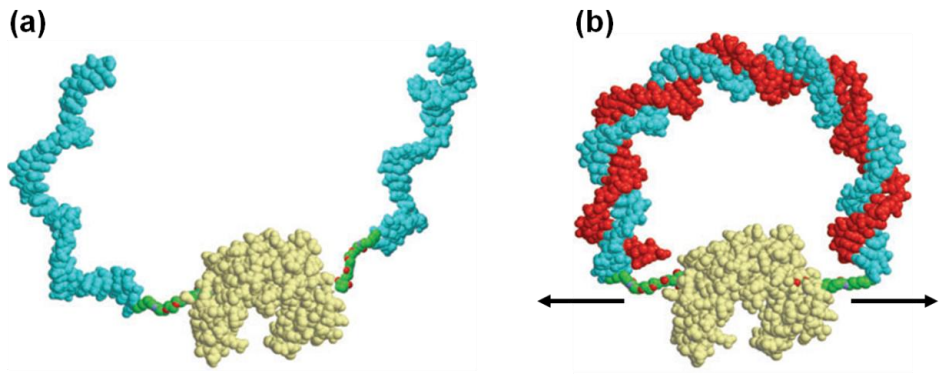


Fig 4.1.2.1 Two-armed protein:DNA chimera. Representation of a guanylate kinase:DNA chimera **(a)** two different DNA arms (light blue) are attached to introduced cysteine residues via a crosslinker (green) **(b)** ligation and hybridisation with a complementary DNA strand (red) to form a double stranded DNA spring. Direction of applied force is indicated with arrows. Figure adapted from Zocchi, 2009.

The use of a double stranded DNA spring has been calculated to operate at levels of force that are lower than those reliably measurable by AFM experiments, with approximately 7.4 pN for a 40 bp dsDNA spring and a maximum achievable force of approx. 10 pN (Tseng *et al.*, 2009). Lower forces are also achievable by titration with shorter complementary DNA strands. This allows us to study the effect of lower forces on TwcK that would be indistinguishable from experimental noise in previous studies.

Another advantage of the use of protein:DNA chimeras is that the majority of the elastic energy of the protein:DNA chimera system is stored in the DNA spring, rather than the protein. This means that the protein is not likely to be deformed on a global basis and that force is highly directed along the molecular axis joining the sites of DNA attachment (Tseng *et al.*, 2009), allowing us to target specific regions of TwcK for stretch.

Here, we explore the principles of mechanoactivation in TwcK by using protein:DNA chimeras to stretch the NL of an autoinhibited TwcK construct and

monitor catalysis in the mechanically deformed state, the first example of protein:DNA chimeras to study stretch-activation of enzymatic activity.

4.1.3 Experimental approach

In this work, a TwcK:DNA chimera was constructed to direct mechanical force to the NL. The NL of TwcK was suggested to be mechanically labile by MD simulations, requiring low force for mechanical removal (von Castelmur *et al.,* 2012). The NL was therefore identified as a suitable target for exploring mechanoactivation, and was chosen to be the site of attachment for a DNA spring. Under tension, the activity of the TwcK:DNA chimera can then be measured by phosphotransfer assay, revealing the effect of mechanical deformation on TwcK catalysis.

For construction of the TwcK:DNA chimera, the autoinhibited NLKinCRD construct (as described in **section 1.4** and von Castelmur *et al.,* 2012 was used as a starting point. To stretch the NL, a construct was designed in which cysteine residues were introduced to the NL at residue N131 (N131C) and in the N-lobe of the TwcK catalytic domain at residue A177 (A177C). The position of introduced cysteine residues for attachment of DNA is illustrated in **Fig 4.1.3.1.** along with a schematic representation of the designed TwcK protein:DNA.

From a visual inspection of the TwcKR crystal structure, both N131 and A177 were judged to be surface exposed and not participating in interactions with other residues. N131 is located near to the region of the NL that interacts with the CRD in the ATP binding site. It was therefore thought that mechanical force directed

to N131C would be effective in releasing the autoinhibition of the NL. A177 is located at the rear of the TwcK N-lobe, distant from the catalytic residues. It was considered that attachment of a DNA spring at A177C would be unlikely to have deleterious effects on the TwcK active site.

For TwcK:DNA chimera production, native surface exposed cysteine residues at C266 and C368 were replaced by serine, generating an 'exposed cysteine null' construct named NLKinCRD^{ECN}. Removal of native cysteines is required to ensure attachment of DNA only to the introduced cysteines at N131C and A177C. The final construct used for the TwcK:DNA chimera was NLKinCRD$^{ECN/N131C/A177C}$, designed to direct the mechanical force of the DNA spring to the NL in order to assess the ability of TwcK to undergo mechanoactivation.

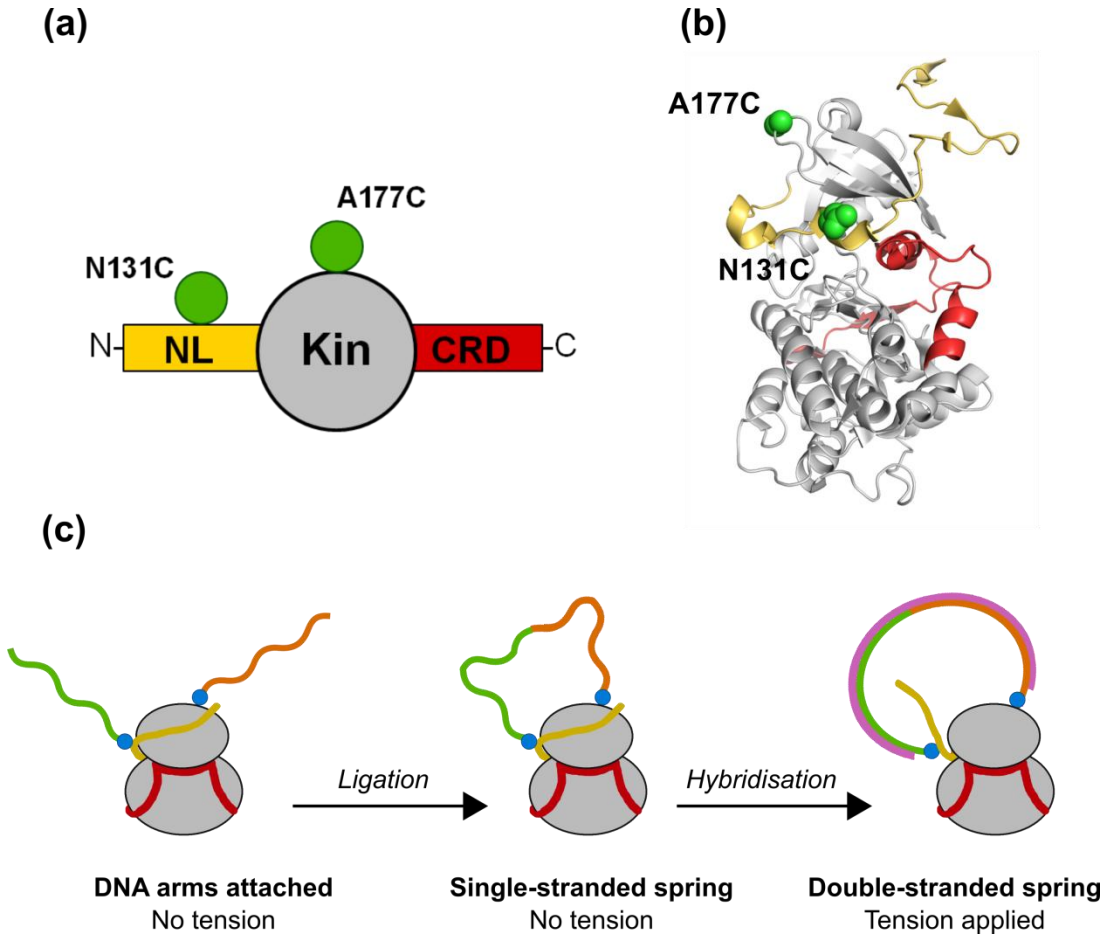


Fig 4.1.3.1 TwcK:DNA chimera design. Cysteine residues were introduced to TwcK NLKinCRD at positions N131 and A177C for attachment of a DNA spring to the NL and kinase N-lobe. **(a)** The location of introduced cysteines at N131 and A177 are represented in cartoon form. **(b)** The location of N131 and A177 represented by spheres within the TwcKR crystal structure. **(c)** Schematic representation of the proposed TwcK protein:DNA chimera. DNA arms are attached to the protein at the NL (N131C) and N-lobe (A177C) to form a DNA spring. Tension will be applied in an attempt to mechanically displace the NL.

4.2 Methods

4.2.1 Introduction of cysteines at N131 and A177 of TwcK by site-directed mutagenesis for attachment of a DNA spring

To stretch the NL, an autoinhibited NLKinCRD construct was designed in which cysteine residues were introduced to the NL at residue N131 and in the N-lobe of the TwcK catalytic domain at residue A177. For introduction of mutations to NLKinCRD, Quikchange site-directed mutagenesis was carried out based on standard protocols provided by Agilent. Pfu DNA polymerase and DpnI were both purchased from Thermo Fisher.

In generating a suitable construct for attachment of DNA arms, it is required to remove surface exposed native cysteine residues to ensure DNA attachment only at the desired sites. A visual inspection of the TwcKR crystal structure (PDB: 3UTO) was carried out to identify all surface exposed cysteine residues within the NLKinCRD region of TwcKR. Two sites, C266 and C368 were identified and subsequently mutated to serine, creating a passivated NLKinCRD construct, here named NLKinCRD^{ECN} (exposed cysteine null). Substitutions to cysteine were subsequently made at N131 (within the NL) and A177 (N-terminal kinase lobe), generating an NLKinCRD$^{ECN/N131C/A177C}$ variant, differing from wild-type NLKinCRD by 4 point mutations. A list of the mutagenic primers used is given in **Table 4.2.1.1**.

Table 4.2.1.1 Quikchange mutagenic primers for NLKinCRD$^{ECN/N131C/A177C}$ production. The site of mutagenesis is underlined.

Primer Name		Sequence (5'-3')
C266S	sense	caagtctgtaagggactc**a**gtcatatgcatgagaaca
	antisense	tgttctcatgcatatgac**t**gagtcccttacagacttg
C368S	sense	ttgaggaatgttaagagc**a**gtgactggaatatggacg
	antisense	cgtccatattccagtcac**t**gctcttaacattcctcaa
N131C	sense	Ggaaccgtttcttctaattatgac**tg**ctacgttttgatatctggaagca
	antisense	Tgcttccagatatcaaaaacgtag**ca**gtcataattagaagaaacggttcc
A177C	sense	Gcatttggagttgtgcacagagtaactgaaaga**tgc**actggaaacaactttgc
	antisense	Gcaaagttgtttccagtgcatctttcagttact**ctg**tgcacaactccaaatgc

4.2.2 Recombinant protein expression and purification of TwcK constructs

Recombinant expression of all TwcK constructs was carried out as described in section **2.2.3.** Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 2 mM β-ME, pH 7.9) containing 20 µg/mL DNase I (Sigma Aldrich) and 1 complete EDTA-free protease inhibitor tablet (Roche) per litre of cell culture. Cell lysis was by sonication on ice followed by clarification of lysate by centrifugation at 39000 *g* for 45 minutes.

Cell lysate was syringe filtered (0.22μm) and applied to a 5 mL HisTrap HP column connected to an Åkta FPLC system (GE Healthcare) equilibrated in lysis buffer containing 20 mM imidazole. Elution of His_6-tagged protein was by continuous imidazole gradient. Following Ni^{2+} purification, eluted samples were buffer exchanged into lysis buffer. His_6-Tag removal was not performed due to the requirement for an affinity tagged protein sample for protein:DNA chimera production.

Size exclusion chromatography (SEC) was performed on a Superdex 200 16/60 column (GE Healthcare) in 50 mM Tris-HCl, 50 mM NaCl, 2 mM β-ME, pH 7.9 (SEC buffer). The resulting fractions were analysed by SDS-PAGE (12% acrylamide) and stored in SEC buffer at 4^oC for later experimentation. Protein concentrations were measured by absorbance at 280 nm using a nanodrop instrument (Thermo Fisher).

4.2.3 Preparation of single stranded DNA arms

The production of TwcK protein:DNA chimeras was performed by myself in the laboratory of Prof. Giovanni Zocchi, University of California, Los Angeles. Methods for protein:DNA chimera production were based on those presented in Choi *et al.*, Zocchi, 2005; Tseng & Zocchi, 2013; Joseph *et al.*, 2014, as described in the following:

The spring length used for construction of the TwcK:DNA chimera was 40 bp, requiring individual 20mer arms to be ligated to form the DNA spring. Reaction of individual 20mer DNA arms with NHS-$[PEG]_2$-Maleimide hetero-bifunctional crosslinker (Sigma Aldrich) was performed using a 100:1 molar ratio of crosslinker

to DNA in 100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.0. Starting amounts of DNA were 65 nmols of DNA arm A and 100 nmols DNA Arm B.

Anion exchange chromatography was performed on the DNA-crosslinker reactions using a UNO Q1R anion exchange column (Biorad) to remove excess free crosslinker. Anion exchange chromatography was carried out in 100 mM sodium phosphate pH 7.0, 150 mM NaCl, 1 mM EDTA with elution by increasing concentrations of NaCl. Sequences of DNA arms are given in **Table 4.2.6.1**, along with details of DNA modifications and other DNA oligonucleotides used.

4.2.4 Attachment of single stranded DNA arms to NLKinCRD$^{ECN/N131C/A177C}$

For attachment of the DNA-crosslinker complexes to TwcK NLKinCRD$^{ECN/N131C/A177C}$, 500 nmols of His_6-tagged NLKinCRD$^{ECN/N131C/A177C}$ was buffer exchanged into 100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.0 using PD-10 desalting columns (GE Healthcare) and incubated with 65 nmols of crosslinked DNA arm A for 1.5 hours at room temperature. Anion exchange chromatography was then performed as described in **section 4.2.3** to purify protein-crosslinker-DNA arm A complexes. Purified NLKinCRD$^{ECN/N131C/A177C}$ with attached DNA arm A was then incubated with 100 nmols of DNA arm B in 100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.0 overnight at room temperature.

Following overnight incubation with DNA arm B, NLKinCRD$^{ECN/N131C/A177C}$ protein-DNA mixture was buffer exchanged to affinity buffer (50 mM Tris, 500 mM

NaCl, 0.1 mM DTT, pH 7.9) using PD-10 desalting columns (GE Healthcare). The sample was then mixed with 1-2 ML Ni-NTA resin (Qiagen) pre-equilibrated in affinity buffer and incubated for 1 hour at 4^{o}C. The slurry was then loaded onto an empty gravity flow column (GE Healthcare) and washed with 10 mL affinity buffer containing 20 mM Imidazole. Elution of NLKinCRD$^{ECN/N131C/A177C}$ double armed protein-crosslinker-DNA complex was with 5 mL affinity buffer containing 400 mM Imidazole. The successful attachment of DNA arms NLKinCRD$^{ECN/N131C/A177C}$ was assessed by SDS-PAGE, coomassie stained for protein and SyBr Gold (Invitrogen) stained for DNA. Co-localisation of stains indicated successful DNA attachment.

4.2.5 Ligation of DNA arms to form a single stranded DNA spring

Double-armed NLKinCRD$^{ECN/N131C/A177C}$ was buffer exchanged into ligation buffer (50 mM Tris, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5) using PD-10 desalting columns (GE Healthcare) in preparation for ligation of the DNA arms to form a single stranded spring. Ligation was performed overnight at 16^{o}C using 2 U/µL T4 DNA Ligase (NEB). The ligation reaction also contained an 18mer DNA splint complementary to short regions of each DNA arm to facilitate ligation by bringing the two DNA arms into close proximity. The nucleotide sequence of this DNA splint is given in **Table 4.2.6.1.** Ligated product was further purified by Ni^{2+} affinity using Ni-NTA resin as described in **section 4.2.4** to remove T4 DNA ligase from the sample. Buffer exchange into hybridisation buffer (50 mM Tris, 50 mM NaCl, pH 7.4) was performed using PD-10 desalting columns (GE Healthcare).

Ligated product was analysed by SDS-PAGE as described in **section 4.2.4** and protein concentration was measured by Bradford assay (Bradford, 1976).

4.2.6 Hybridisation of single stranded spring to form a double stranded DNA spring

Hybridisation of the ligated ssDNA to form a dsDNA spring was performed using a complementary ssDNA strand 40 bases in length (sequence in **Table 4.2.6.1**) in a 5:1 complement: single stranded chimera molar ratio. Hybridisation was in hybridisation buffer overnight at room temperature. The nucleotide sequence of the Concentration of protein:DNA chimeras was measured by Bradford assay (Bradford, 1976).

Table 4.2.6.1 DNA oligonucleotides used for protein:DNA chimera production. DNA arms are modified for reactivity with the NHS-$[PEG]_2$-Maleimide hetero-bifunctional crosslinker; AmC6 = C6 amine, AmC3 = C3 amine, Phos = phosphorylation. DNA splint is used to facilitate ligation of single DNA arms and complement is used for hybridisation to form a double stranded DNA spring.

DNA Arm A 20mer	5'-/AmC6/GAGTGTGGAGCCTAGACCGT-3'
DNA Arm B 20mer	5'-/Phos/CGGTACCATCCAAGCAGCTG/AmC3/-3'
DNA splint 18mer	5'-TGGTACCGACGGTCTA-3'
40mer complement	5'-CAGCTGCTTGGATGGTACCGACGGTCTAGGCTCCACACTC-3'

4.2.7 Measurement of TwcK:DNA chimera catalysis

Phosphotransfer assays were performed as described in section **2.2.4**, with chicken gizzard myosin light chain derived model peptide substrate with sequence KKRARAATSNVFS as previously reported by Heierhorst *et al.,* 1996b and von Castelmur *et al.,* 2012.

4.3 Results

4.3.1 Production of recombinantly expressed NLKinCRD$^{ECN/N131C/A177C}$

For TwcK:DNA chimera production, the NLKinCRD$^{ECN/N131C/A177C}$ mutated construct was produced by site-directed mutagenesis. This had its surface exposed native cysteines mutated to serine and introduced cysteines at N131 and A177 for DNA attachment. NLKinCRD$^{ECN/N131C/A177C}$ was recombinantly expressed in *E. coli* and purified chromatographically. Purified NLKinCRD$^{ECN/N131C/A177C}$ was approx. 90% pure as judged by SDS-PAGE, with a contaminant protein of low apparent molecular mass remaining following SEC **(Fig 4.3.1.1)**.

Expression of NLKinCRD$^{ECN/N131C/A177C}$ gave a yield of approx. 20 mg soluble, purified protein per litre of cell culture. SEC chromatograms show a highly similar elution profile for NLKinCRD$^{ECN/N131C/A177C}$ and wild-type NLKinCRD **(Fig 4.3.1.1).**

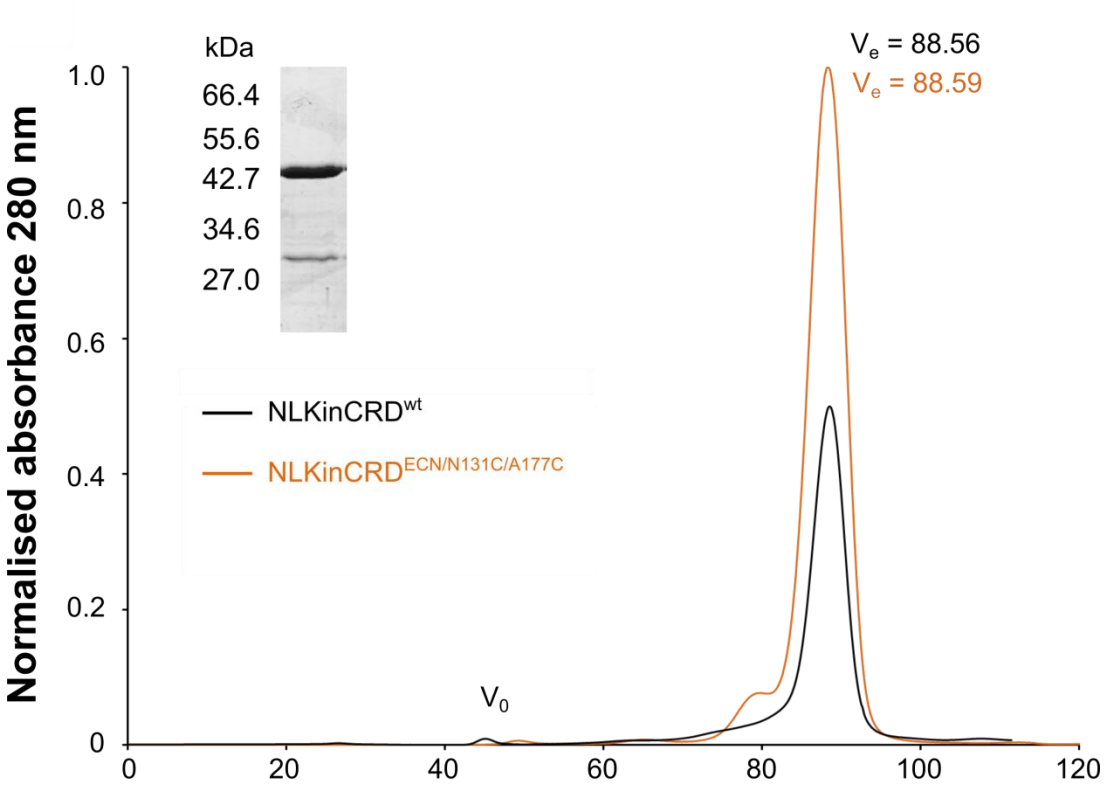


Fig 4.3.1.1 Purification of NLKinCRD$^{ECN/N131C/A177C}$. SEC chromatogram for recombinantly expressed NLKinCRD^{wt} and NLKinCRD$^{ECN/N131C/A177C}$ from a Superdex 200 16/60 column. The column void volume (V_0) and exclusion volume (V_e) are indicated. SDS-PAGE (inset) for NLKinCRD$^{ECN/N131C/A177C}$ from SEC. Molecular weight marker (Broad Range 2-212 kDA, NEB) is labelled adjacent.

4.3.2 TwcK is activated by mutation in the NL at residue N131

NLKinCRD$^{ECN/N131C/A177C}$ was assayed for catalytic activity to assess the suitability of the construct for assessing changes in catalysis under stretch in the TwcK:DNA chimera. Phosphotransfer assays included the Kin construct (free of autoinhibitory tails) as a positive control as well as the wild-type NLKinCRD to represent the autoinhibited state of TwcK **(constructs described in section 2.2.1)**.

Unexpectedly, phosphotransfer assays revealed that the activity of the NLKinCRD$^{ECN/N131C/A177C}$ construct was significantly above that of the autoinhibited wild-type NLKinCRD, displaying 24% of the activity of the Kin positive control **(Fig 4.3.2.1)**.

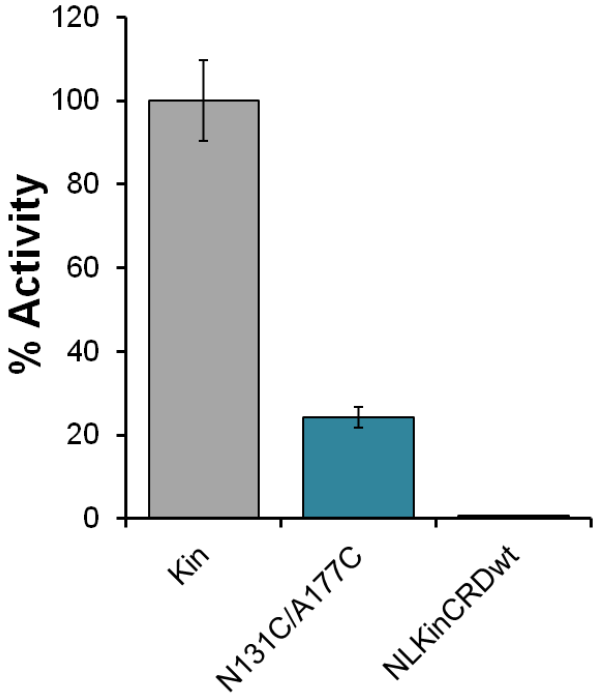


Fig 4.3.2.1 Catalytic activity of NLKinCRD$^{ECN/N131C/A177C}$. Phosphotransfer assay included the TwcK catalytic domain (Kin) as a positive control and designated as 100% activity, with the activity of all other samples expressed as a % of the positive control. NLKinCRD$^{ECN/N131C/A177C}$ was seen to have 24% activity compared to the positive control, with wild-type NLKinCRD having negligible activity (0.5%). Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

The results of the phosphotransfer assay on NLKinCRD$^{ECN/N131C/A177C}$ led us to conclude that the mutation of N131 within the NL of TwcK had likely caused a disruption of the NL, causing it to partially unfold and thereby activating TwcK catalysis. However, the unexpectedly high levels of catalysis seen for NLKinCRD$^{ECN/N131C/A177C}$ represent an advantage for monitoring catalysis of subsequent TwcK:DNA chimeras.

The use of a construct with intermediate levels of catalysis (in this case 24%) provide a readable signal output for assessing whether levels of catalysis increase or decrease under stretch. When starting from a fully autoinhibited construct, no increase activity under tension could be explained by *a)* failure to release the autoinhibitory NL or *b)* deformation of the TwcK active site. The intermediate activity of NLKinCRD$^{ECN/N131C/A177C}$ therefore allows us to address one of the fundamental questions posed at the beginning of this chapter; can TwcK perform catalysis under mechanical deformation?

Samples were also assayed for activity in which the TwcK catalytic domain (Kin) was mixed with increasing concentrations of free DNA with sequence corresponding to the full 40 mer complementary strand used to hybridise the 40 bp DNA spring. This showed that the presence of DNA does not influence TwcK catalysis and that observed changes in catalysis in protein:DNA chimeras can be attributed to the stretch generated by the DNA spring, rather than an interaction with the TwcK active site.

4.3.3 Construction of TwcK:DNA chimera to apply mechanical stress to NLKinCRD$^{ECN/N131C/A177C}$

Construction of protein:DNA chimeras involves the covalent attachment of DNA arms to engineered cysteines within the protein. Firstly, 20mer DNA arms were prepared individually for attachment to TwcK NLKinCRD$^{ECN/N131C/A177C}$. DNA arms were incubated with the NHS-$[PEG]_2$-Maleimide hetero-bifunctional crosslinker. Anion exchange chromatography was performed for each DNA arm-crosslinker reaction to remove excess crosslinker. The elution profile for each DNA

arm-crosslinker complex showed two peaks; one at low NaCl concentration corresponding to free crosslinker and a peak eluted at 400-600 mM NaCl corresponding to purified DNA arm-crosslinker complexes **(Fig 4.3.3.1).**

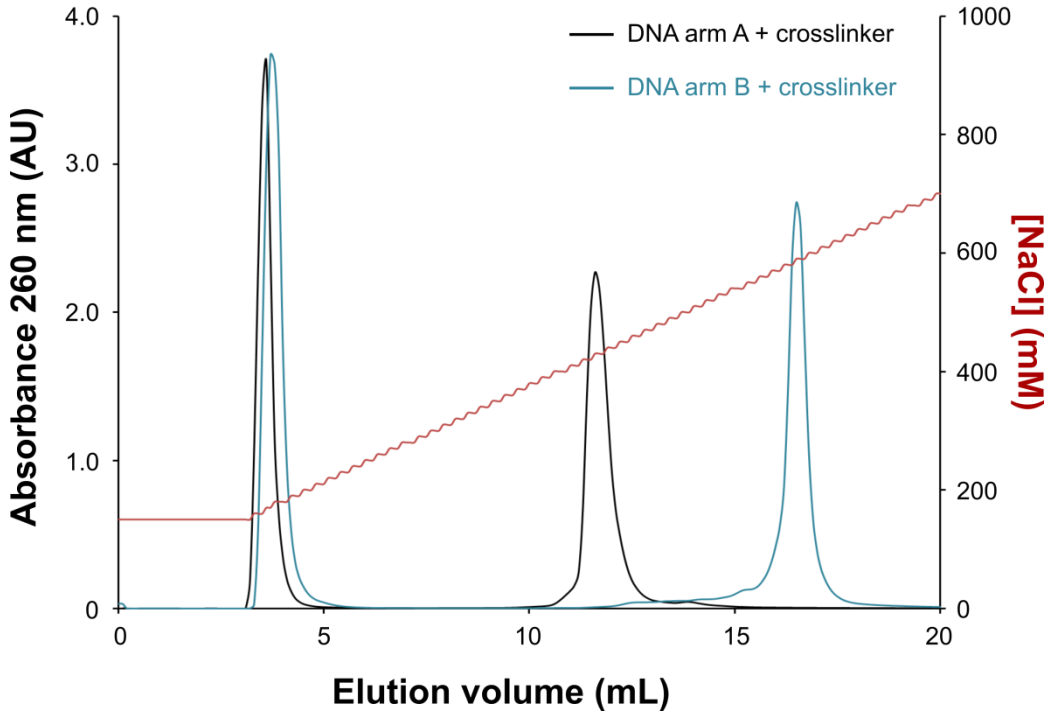


Fig 4.3.3.1 Purification of DNA arms. Anion exchange chromatograms for purification of DNA arm A (black) and DNA arm B (blue) following reaction with crosslinker. Elution was with NaCl gradient, shown in red. Peaks at low salt concentration correspond to excess crosslinker, with peaks at 400-600 mM NaCl corresponding to purified DNA arms with covalently attached crosslinker.

To construct the single stranded DNA spring, individual DNA arms were attached independently to the protein. Following purification of DNA-crosslinker complexes, DNA arm A-crosslinker was incubated with purified NLKinCRD$^{ECN/N131C/A177C}$ protein. To produce a sample in which the majority of the protein has only a single DNA arm attached at this stage, the incubation of NLKinCRD$^{ECN/N131C/A177C}$ with DNA arm A was performed with a large excess of protein. Single-armed protein was then purified by AEC **(Fig 4.3.3.2).**

Separation of free protein from single-armed protein by AEC relies on the binding of only DNA to the positively charged column media, allowing retention of protein only if DNA has been successfully attached. For this reason, AEC was carried out at pH 7.0; NLKinCRD$^{ECN/N131C/A177C}$ should carry only a weak overall negative charge at pH 7.0 due to a theoretical isoelectric point (pI) of 5.86. A starting concentration of 150 mM NaCl was used to further inhibit retention of unmodified protein. The elution profile for AEC purification of single-armed NLKinCRD$^{ECN/N131C/A177C}$ shows a major peak corresponding to the desired single-armed NLKinCRD$^{ECN/N131C/A177C}$ as well as a smaller peak eluted at a higher NaCl concentration **(Fig 4.3.3.2)**. This smaller peak corresponds to NLKinCRD$^{ECN/N131C/A177C}$ with two identical DNA arms attached at both cysteine residues.

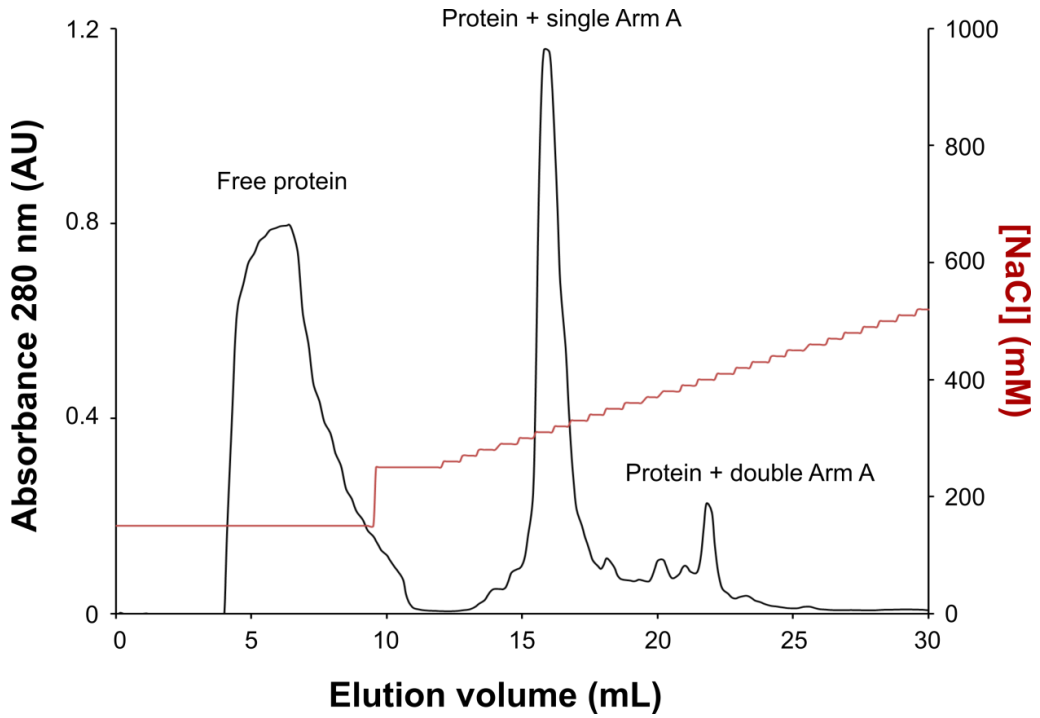


Fig 4.3.3.2 Anion exchange purification of single-armed NLKinCRD$^{ECN/N131C/A177C}$. Elution was with NaCl gradient, shown in red. Peaks are labelled to signify the corresponding eluted species.

Following incubation of single-armed NLKinCRD$^{ECN/N131C/A177C}$ with DNA arm B, the two DNA arms were ligated to form a single stranded spring. The resulting product was purified by Ni^{2+} affinity via the His_6-tag of NLKinCRD$^{ECN/N131C/A177C}$. Hybridisation to form a double stranded spring was performed using a 40 mer complementary DNA strand. The TwcK:DNA chimera construction process was confirmed by SDS-PAGE. Co-localisation of Coomassie (protein) and SyBr Gold (DNA) stains, as well as an associated shift in electrophoretic mobility showed the successful attachment of DNA to NLKinCRD$^{ECN/N131C/A177C}$ (**Fig 4.3.3.3).** The final yield of ligated, double-armed NLKinCRD$^{ECN/N131C/A177C}$ was approximately 11 nmols as judged by Bradford assay.

SDS-PAGE (**Fig 4.3.3.3)** revealed a number of additional species, which were observed with significantly lower electrophoretic mobility. These species correspond to ligation events in which compatible DNA arms from two or more different protein:DNA chimeras become ligated. The resultant complexes consist of multiple protein:DNA chimeras joined together. SDS-PAGE results were consistent with previous observations of a final yield of correctly assembled protein:DNA chimera of 50-70% of the total species (Tseng & Zocchi, 2013).

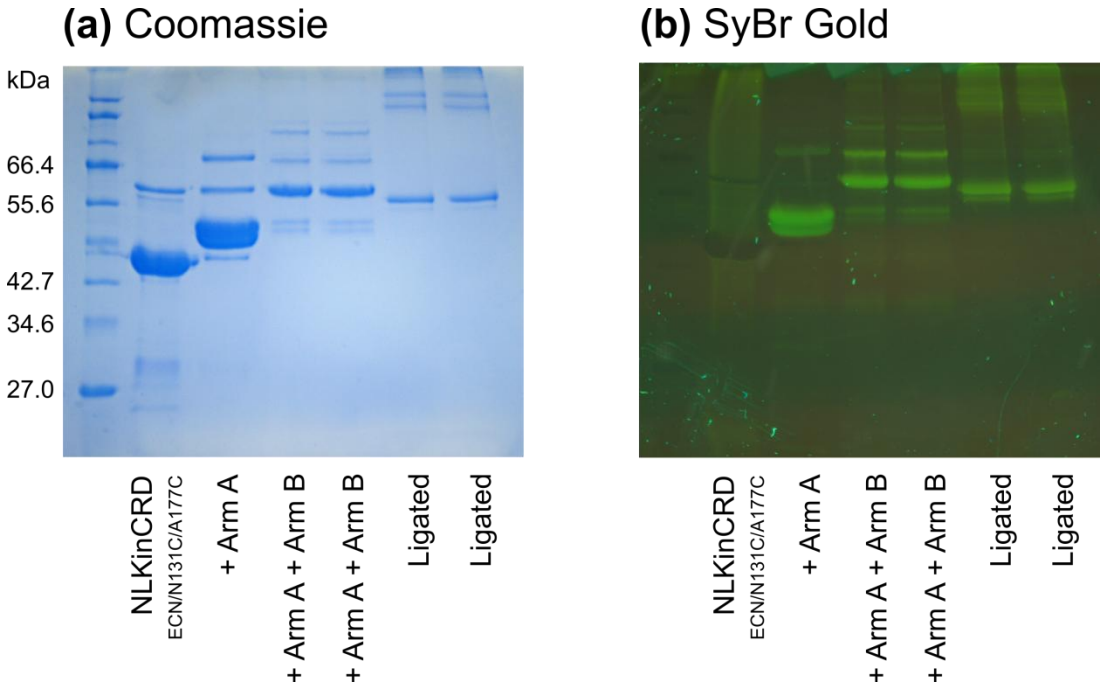


Fig 4.3.3.3 Monitoring of TwcK:DNA chimera construction by SDS-PAGE. Co-localisation of staining by **(a)** Coomassie for protein and **(b)** SyBr Gold for DNA indicates successful attachment of DNA.

4.3.4 Tension applied to the TwcK:DNA chimera results in a decrease in catalytic activity

To assess the effect of mechanical tension directed to the NL of TwcK, the catalytic activity of the NLKinCRD$^{ECN/N131C/A177C}$ protein:DNA chimera was tested by phosphotransfer assay. The assay was performed to test the activity of the TwcK:DNA chimera hybridised with a 40mer DNA complement, producing a fully double stranded DNA spring of 40 bp in length. The 40 bp spring represents the greatest amount of tension that can be generated from the spring constructed from the ligation of two 20mer DNA arms (Zocchi, 2009). The assay included the Kin construct (free of autoinhibitory tails) as a positive control as well as the wild-type

NLKinCRD to represent the autoinhibited state of TwcK **(constructs described in section 2.2.1)**.

The NLKinCRD$^{ECN/N131C/A177C}$:DNA chimera with a 40 bp double stranded spring was seen to show a reduced activity compared to the pre-tension state, displaying 11% activity of the Kin positive control, corresponding to an approx. 50% reduction in activity upon stretch **(Fig 4.3.4.1)**. The reduction in catalysis seen under tension in the TwcK:DNA chimera suggests that the force applied is distorting the TwcK catalytic domain. This therefore indicates the TwcK catalytic domain is sensitive to mechanical deformation and is not itself intrinsically resistant to stretch.

The observed reduction in the catalysis for the NLKinCRD$^{ECN/N131C/A177C}$:DNA chimera stretched by a 40 bp dsDNA spring represents an encouraging preliminary result in exploring TwcK mechanoactivation. Although mechanoactivation was not observed when applying stretching force to sites at N131C and A177C, the observed change in catalysis is supportive of the use of a DNA spring as a suitable technique for TwcK stretching experiments. These results show that a TwcK:DNA chimera can be successfully constructed to apply force to TwcK and that the applied force is sufficient to alter TwcK catalysis.

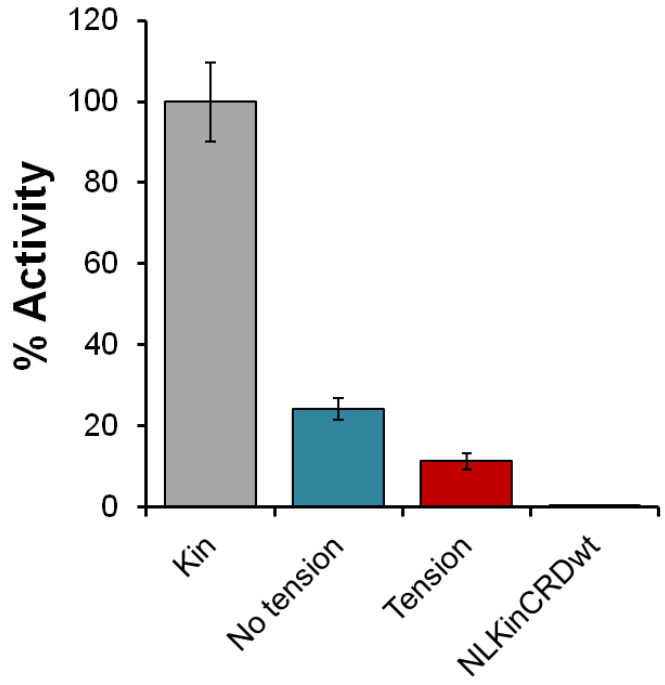


Fig 4.3.4.1 Catalytic activity of NLKinCRD$^{ECN/N131C/A177C}$:DNA chimera. For phosphotransfer assays, the TwcK catalytic domain (Kin) was used as a positive control and designated as 100% activity, with the activity of all other samples expressed as a % of the positive control. 'No tension' represents NLKinCRD$^{ECN/N131C/A177C}$ with ssDNA spring and 'Tension' represents NLKinCRD$^{ECN/N131C/A177C}$ with a 40 bp dsDNA spring. A reduction in activity of ~50% was observed upon induction of tension with the dsDNA spring. Error bars were calculated by subtracting the standard deviation of background measurements from the standard deviation of sample measurements. Assays were carried out in duplicate.

4.4 Discussion

In this work we present important preliminary work towards establishing the use of protein:DNA chimeras to investigate the mechanoactivation of TwcK. The application of force directed to the NL of the NLKinCRD$^{ECN/N131C/A177C}$ construct via a 40 bp dsDNA spring showed a decrease in activity compared to the pre-tension state. This shows that the mechanical force generated by the 40 bp dsDNA spring is sufficiently high to impact TwcK catalysis, validating the use of TwcK:DNA chimeras (with further optimisation) as a suitable technique to explore the effect of mechanical deformation on TwcK activity.

The constructed TwcK:DNA chimera was designed to stretch TwcK via the NL. The catalytic data presented here for NLKinCRD$^{ECN/N131C/A177C}$ show intermediate levels of catalysis, in stark contrast to the autoinhibited wild-type NLKinCRD that shows negligible catalysis. These data suggest that a small local disruption to the NL of TwcK, in this case through mutation of N131 to cysteine, is sufficient to cause partial activation of TwcK. If small distortions of the NL are capable of activating TwcK, the approach of directing low mechanical force to the NL to activate catalysis would appear to be justified. Additionally, If a perturbation of the NL has caused activation whilst the CRD is present, this would imply that said disruption has also lead to displacement of the CRD in order to free the active site for substrate binding. This is in keeping with the proposed hypothesis of the CRD 'looping out' or falling off the kinase domain, rather than requiring mechanical displacement (von Castelmur *et al.*, 2012).

N131, The site of DNA attachment in the NL, is located in a region in which interactions between the NL, CRD and the kinase domain are seen to occur. The solvent exposed N131 within the NL that was substituted for a cysteine for DNA spring attachment is within a helical region of the NL, close to the ATP binding pocket of TwcK. This helical region is conserved across TwcKs (von Castelmur *et al.,* 2012) and likely makes an important contribution to autoinhibition by the NL.

In close proximity to N131, Y129 is highly conserved in TwcKs and is seen to interact via hydrophobic interaction with P447 in the αR2 helix of the CRD at the ATP binding pocket in the crystal structure of TwcKR. Y129 also hydrogen bonds with E160 from the kinase N-lobe, helping to anchor the NL to the catalytic domain of TwcK **(Fig 4.4.1)** (von Castelmur *et al.,* 2012). A local disruption in this region could therefore disturb crucial interactions with both the kinase domain and the CRD that contribute to the autoinhibited state of TwcK. This therefore indicates that the helical region of the NL in which N131 is located is important for maintaining an autoinhibited state.

Construction of a TwcK:DNA chimera using the NLKinCRD$^{ECN/N131C/A177C}$ construct showed that mechanical force can be applied to TwcK using a DNA spring. Catalytic assays showed that under stretch, the activity of the NLKinCRD$^{ECN/N131C/A177C}$ TwcK:DNA chimera was reduced by approx. 50%. The results of these assays therefore indicate that mechanical deformation of TwcK can inhibit catalysis. With the TwcK susceptible to mechanical inhibition, it appears that the catalytic domain of TwcK is not itself intrinsically resistant to mechanical tension. With the second DNA attachment point located at A177C in the N-lobe of the TwcK catalytic domain, it is likely that the stretching force generated by the DNA spring is directly impinging on the catalytic domain of TwcK. Glycine-rich

loop closure was seen to be the only significant conformational change required for TwcK to enter the active conformation **(chapter 3)** and it is likely that the force directed through DNA attachment at A177C is restricting glycine-rich loop closure.

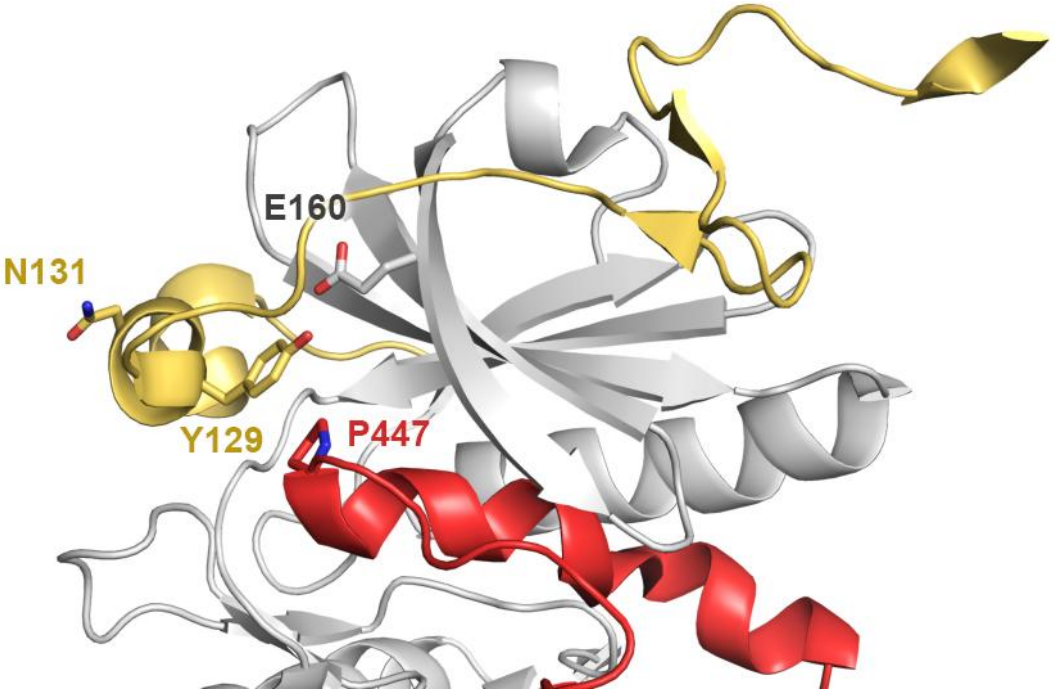


Fig 4.4.1 Mutation of N131 is likely to disrupt interactions between the NL and kinase domain of TwcK. Residue N131 is located within a helical region of the NL (gold). Nearby Y129 hydrogen bonds with E160 in the kinase domain (grey) and has a hydrophobic interaction with P447 of the CRD (red). Mutation of N131 to cysteine for DNA attachment may have caused disruption of these interactions and led to the partial activation of catalysis seen in NLKinCRD$^{ECN/N131C/A177C.}$.

It has been hypothesised that the CRD could remain partially bound to the catalytic domain of TwcK under stretch, acting as a mechanical buffer to protect against mechanically induced damage (von Castelmur *et al.,* 2012). The idea of a mechanical buffer is compatible with the data presented here showing that TwcK catalysis is sensitive to mechanical force, however whether the CRD can fulfil this

role is uncertain. The portion of the CRD that blocks the active site (αR2 and αR3 helices) must be displaced in order for catalysis to occur as it blocks substrate binding. However, the structure of the TwcK catalytic domain presented in **chapter 3**, with a disordered αR1 helix suggests that the αR1 helix may not remain bound to the TwcK catalytic domain as once thought. Steered MD simulations also suggested that the βR4 strand of the CRD unfolds at low force (von Castelmur *et al.,* 2012). While TwcK catalysis has been shown here to be sensitive to mechanical deformation, the CRD may not act as a mechanical buffer.

4.5 Future perspectives

Protein:DNA chimeras are a highly versatile technique for studying the mechanical properties of proteins. A major feature of the protein:DNA chimera technique is that the magnitude of force applied by the DNA spring can be easily modulated by varying the properties of the DNA spring used. Firstly, the length of the complementary strand used for hybridisation of the ssDNA spring can be reduced, lowering the double stranded length and therefore stiffness of the DNA spring (Choi *et al.*, 2005). By using a range of DNA complement lengths, it has been demonstrated that reducing the hybridised length of the dsDNA spring leads to a lower degree of inhibition in mechanically deformed proteins (Choi *et al.*, 2005).

The total length of the DNA spring can also be varied in order to modulate the force applied to the protein. A longer dsDNA spring of 60 bp has been used in much of the published work involving protein:DNA chimeras to control enzymatic activity (Choi *et al.*, 2005; Choi & Zocchi, 2006; Tseng & Zocchi, 2013; Joseph *et al.*, 2014). With longer spring lengths, the amount of bending induced in the DNA spring upon hybridisation would lowered and therefore would be expected to exert lower force. With TwcK catalysis shown here to be inhibited by mechanical deformation, future experiments with TwcK:DNA chimeras could explore the levels of force required to inhibit catalysis by making use of the customisability of the DNA spring.

Another important feature of the use of DNA as a molecular spring is the reversibility of the applied tension. This is of particular importance to the study of TwcK mechanoactivation. TwcK and the other members of the titin-like family are located within the muscle sarcomere. In playing a role in mechanosensing in the sarcomere, TwcK would be required to return to a pre-tension state when muscle relaxation occurs. If TwcK were to undergo mechanoactivation through mechanical deformation, the effect of the deformation must therefore be reversible. In studying mechanical inhibition of enzymatic activity, it has been demonstrated that the mechanical tension of the DNA spring can be abolished by digestion with DNase (Choi *et al.*,, 2005). Further TwcK:DNA experiments can therefore be conducted on the ability of TwcK to return to a pre-tension state following mechanical deformation.

For further exploration into the effect of mechanical force on the NL of TwcK, a new construct for TwcK:DNA chimera construction has already been designed. In this construct, attachment sites will be introduced at residues T124 and Q143, both of which are located within the NL **(Fig 4.5.1)**. The data presented in this chapter suggest substitution of N131 causes partial activation of TwcK catalysis due to its close proximity to a number of interacting residues that contribute to autoinhibition. It was also concluded that the attachment point at A177C was likely leading to mechanical force directly impinging on the TwcK catalytic domain. We have used these findings to inform the design of this new TwcK:DNA chimera. Both T124 and Q143 are located in loop regions of the NL that appear to be uninvolved in interactions that are likely to contribute significantly to autoinhibition. It is hoped that with the introduction of two attachment points within the NL, mechanical force directed to the catalytic domain can be avoided. Additionally, it is hoped that

stretching of either end of the NL can induce a significant unfolding of the NL in order to study the effect of full NL removal on TwcK catalysis.

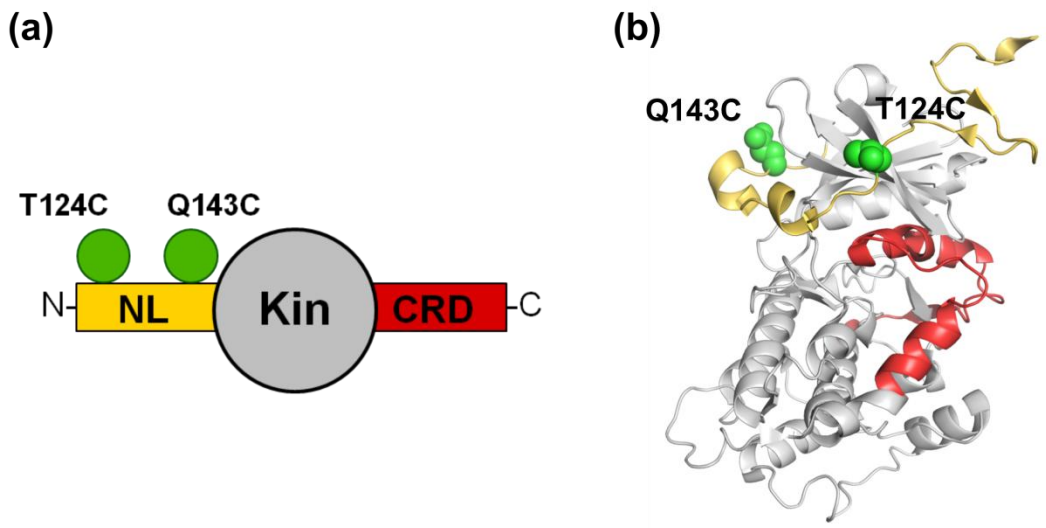


Fig 4.5.1 A new NLKinCRD construct design for TwcK:DNA construction. Cysteine residues are to be introduced at T124C and Q143C for attachment of a DNA spring to the NL. **(a)** The location of introduced cysteines at T124 and Q143 are represented in cartoon form. **(b)** The location of T124 and Q143 represented by spheres within the TwcKR crystal structure.

Finally, as the technique is carried out entirely in solution, the use of protein:DNA chimeras presents the possibility to perform biophysical characterisation of TwcK under stretch. A good candidate technique for characterising the conformational changes undergone in TwcK is electron paramagnetic resonance (EPR) spectroscopy. EPR can be used to measure distances on the nanometre scale between labelled sites (Jeschke, 2002) and is used to study the dynamic properties of proteins, such as conformational changes upon substrate binding (Freeman *et al.,* 2011). Proteins are commonly labelled via introduced cysteine residues with nitrous oxide (NO) spin labels to provide unpaired electrons

as required for EPR measurements. In the case of TwcK:DNA chimeras, cysteine residues are used for DNA attachment and can therefore not be used for introduction of a spin label.

Recently, a genetically encoded artificial nitroxide amino acid has been used to measure distances and conformational flexibilities by EPR in thioredoxin without the need for covalent attachment of a spin label (Schmidt *et al.,* 2015). The application of genetically encoded spin labels to study TwcK conformational changes under stretch from a DNA spring would be a highly exciting prospect. For example, distance measurements could be used to determine whether the NL of TwcK becomes dislodged from the catalytic domain under tension. Such measurements could also be used to explore the hypothesis of the CRD looping out or falling off if the NL is released, perhaps by labelling of the CRD while the NL is under tension from a DNA spring.

In conclusion, the use of TwcK:DNA chimeras as established in this work represents a highly promising first step towards a more detailed understanding of the response of TwcK and other titin-like kinases to mechanical deformation in the sarcomere.

Chapter 5
General Discussion

In this thesis, it has been shown that recombinantly expressed TwcK undergoes autophosphorylation at a number of key mechanistic sites (T212 in the hinge region, T301 adjacent to the DFG motif and T316 in the P+1 loop). These findings indicate a third potential mode of regulating TwcK activity, which may act in concert with the intrasteric inhibition by the NL and CRD tails. TwcK autophosphorylation events are inhibitory or neutral towards phosphotransfer activity *in vitro*, although it is as yet unclear what the *in vivo* function of TwcK autophosphorylation may be. Although it first must be established that TwcK autophosphorylation occurs *in vivo*, taking into account current knowledge of TwcK mechanistic principles, one can speculate on a number of possible scenarios in which TwcK autophosphorylation may play a role in TwcK regulation and muscle function.

Firstly, TwcK autophosphorylation may be a means by which to exert more precise temporal control over TwcK catalysis. The current working hypothesis of TwcK activation involves release of the NL due to build up of mechanical tension in the sarcomere, mechanically priming TwcK for catalysis (von Castelmur *et al.*, 2012). This model for TwcK activation implies that TwcK will be catalytically inactive in the absence of tension and constitutively active under mechanical stress. In keeping with this, we have shown via the use of TwcK:DNA chimeras, that TwcK is likely to be able to remain catalytically active under mechanical stress. However, it is possible that TwcK phosphotransfer activity is required only in short bursts, while periods of sustained tension within the sarcomere could prolong the activation

of TwcK catalysis for longer than necessary. In this scenario, negative regulation of catalysis by autophosphorylation could be employed to inhibit TwcK catalysis while remaining under mechanical tension, allowing for modulation of activity while not intrasterically inhibited by the NL and/or CRD tails. As illustrated in Fig 5.1.1, such a mechanism would require the action of a phosphatase in order to return TwcK to a ground state from which it can be re-activated mechanically. The kinase domains of the *C. elegans* titin-like protein UNC-89 have been previously shown to interact with the phosphatase SCPL-1 (Qadota *et al.*, 2008), showing that at least some members of the titin-like kinase family have phosphatase binding partners. Thus it is possible that TwcK autophosphorylation is part of a cyclical regulatory mechanism consisting of activation by mechanical tension, followed by desensitisation by inhibitory phosphorylation and resensitisation by dephosphorylation, before being able to once again be activated by mechanical tension.

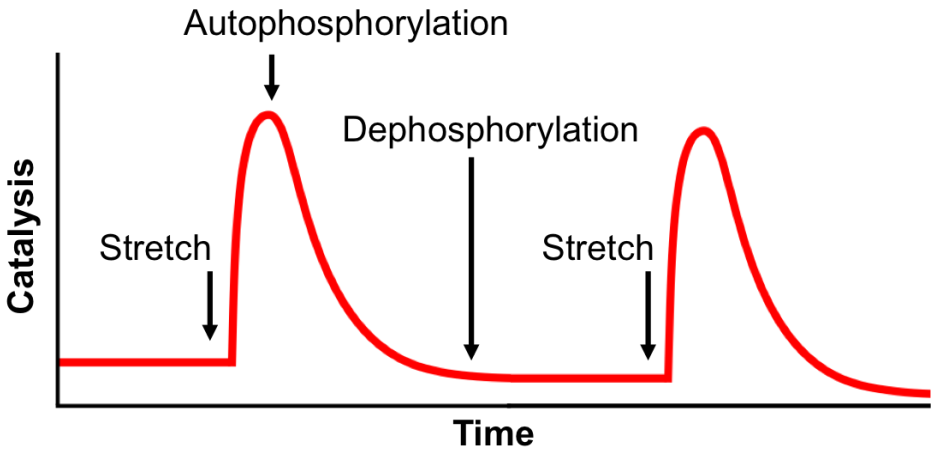


Fig 5.1.1 Hypothetical model of TwcK regulation involving mechanoactivation followed by inhibitory autophosphorylation. Autophosphorylation may serve to inhibit TwcK catalysis while under mechanical tension, requiring resensitisation by dephosphorylation before being able to undergo further stretch activation.

Another possibility for the interplay between autophosphorylation and mechanoactivation is in regulating binding of the NL and CRD autoinhibitory tails of TwcK. Of the phosphorylation sites identified in this work, three are located within regions of the catalytic domain that if modified, could affect binding of both the NL and CRD tails. T212 is located at the rear of the TwcK catalytic domain in a region that facilitates interdomain movement by acting as a hinge point between the N- and C-terminal kinase lobes (Kannan & Neuwald, 2005; Kannan *et al.,* 2008). The NL of TwcK is expected to contribute to autonhibition by occupying the hinge region of TwcK and limiting the conformational motions required to perform phosphotransfer (von Castelmur *et al.,* 2012). If phosphorylation of T212 was to affect interdomain movement, it could be that modification at this site could restrict TwcK to a conformation in which the NL is no longer able to bind.

Phosphorylation of both T301 and T316 are likely to affect binding of the CRD. T301 is located adjacent to the DFG motif, which sits within the active site cleft of TwcK. When bound to TwcK, the αR3 helix of the CRD occupies the active site in close proximity to the DFG motif and T301. Therefore, phosphorylation of T301 may cause sufficient local rearrangement to restrict binding of the CRD within the active site cleft. Similarly, in the crystal structure of autoinhibited TwcK, T316 interacts with K436 from the αR1 helix of the CRD. Phosphorylation of T316 may cause rearrangement of the P+1 loop sufficient to disrupt this interaction between the CRD and catalytic domain, perhaps limiting the capability of the CRD to re-bind following expulsion from the active site during the catalytic cycle.

Titin-like kinases act as signalling platforms in the sarcomere (Lange *et al.,* 2005; Lange *et al.,* 2012; Wilson *et al.,* 2012; Matsunaga *et al.,* 2015) and it has been proposed that mechanical release of the autoinhibitory tails could expose the

kinase surface for protein-protein interactions with components of signalling cascades (Mayans *et al.,* 2013). Autophosphorylation in TwcK may then function to prevent re-binding of autoinhibitory tails, preserving a conformation permissive to this scaffolding role in the absence of prolonged mechanical tension. In such a scenario, dephosphorylation could be required to fully disassemble the signalling complex formed by TwcK and return to an autoinhibited state.

If the main role of TwcK autophosphorylation were to keep TwcK in a productive state for acting as a signalling scaffold, despite its inhibitory effect on catalysis, it could be that the main role of TwcK catalysis is in fact autophosphorylation, rather than targeting another peptidic substrate for phosphorylation. The kinase domain of IRE1 (inositol-requiring enzyme 1) is one such kinase that exclusively performs autophosphorylation. IRE1 is an endoplasmic reticulum (ER) associated protein involved in the unfolded protein response and consists of an N-terminal stress-sensing domain in the ER lumen connected by a trans-membrane helix to a cytoplasmic Ser/Thr kinase and RNase domain (Cox *et al.,* 1993). ER stress drives dimerisation of IRE1 lumenal domains, leading to dimerisation of the cytoplasmic domains (Shamu & Walter, 1996). Autophosphorylation of IRE1 is then required for activation of the RNase domain, which acts to induce translation of a transcription factor involved in transcription of unfolded protein response target genes (Shamu & Walter, 1996).

Upon dimerisation, the IRE1 kinase domain autophosphorylates in an autoinhibitory region comprising the activation loop and adjacent region. In the unphosphorylated form, the autoinhibitory region inhibits IRE1 function by occupying the kinase and/or RNase active site, with autophosphorylation of the activation loop residues S841 and T844 required to adopt a kinase-active

conformation, leading to RNase activation (Lee *et al., 2008;* Korennykh *et al.,* 2009; Joshi *et al.,* 2015). An IRE1 construct in which the autoinhibitory region was removed was still able to activate the RNase activity and stress response, with the kinase-dead mutant (D797A, catalytic base) lacking autoinhibitory region also shown to be functional (Mannan *et al.,* 2013). An IRE1 chimera in which the activation loop was replaced with the activation loop from another kinase, containing a phosphomimetic was also able to activate the stress response (Mannan *et al.,* 2013). The function of IRE1 protein kinase activity is therefore to induce activating conformational changes via autophosphorylation, rather than to phosphorylate a downstream target.

In addition to catalytically active kinases such as TwcK, the titin-like family of kinases contains a number of pseudokinases, such as titin kinase, which is able to act as a signaling platform whilst lacking phosphotransfer activity. Despite being catalytically active and undergoing autophosphorylation, to-date no physiological peptidic substrate has been identified for TwcK. As explored in this work, mlc-4 from C. elegans was identified as a promising candidate substrate, but TwcK phosphotransfer activity towards an mlc-4 derived peptide was poor, despite significant sequence similarity to the current model peptide substrate. As previously discussed, peptide arrays should be employed to identify further potential substrates for TwcK. However, the possibility remains that, similar to IRE1, TwcK may not have a substrate outside of itself, and that its phosphotransfer activity is limited to autophosphorylation *in vivo*. Such autophosphorylation may facilitate its function as a signaling scaffold in the sarcomere, while the pseudokinases of the titin-like family may not require phosphorylation-mediated conformational rearrangement to fulfill such a role.

The unusually open position of the αC-helix seen in the crystal structure of TwcK in complex with staurosporine presented in this work may also give some hints as to whether the primary role for TwcK catalysis is autophosphorylation. Such an open αC-helix conformation may need to be maintained in order to accommodate a folded protein substrate, rather than an extended peptide substrate. The majority of the autophosphorylation sites identified in this work are not in particularly accessible, extended loop regions, and thus would be difficult to accommodate in to the active site of a canonical kinase active conformation. Perhaps then, the αC-helix of TwcK, which remains in a more open position while still maintaining the essential salt-bridge between K185 and E201, occupies such a conformation in order to accommodate a larger substrate, i.e regions of its own catalytic domain.

In the recent work studying the effect of loss of TwcK activity in *C. elegans* (Matsunaga *et al.*, unpublished – **section 7.1.2**), it appears that the role of TwcK activity is to regulate the contraction/relaxation cycle of the sarcomere, with worms carrying the K185A kinase-dead mutation in TwcK showing an increased rate of contraction/relaxation. In light of the work presented in this thesis, it is possible that the phenotypic effects seen *in vivo* are due to a loss of TwcK autophosphorylation. As speculated above, it is possible that autophosphorylation is used by TwcK to preserve the mechanically primed state of TwcK for its role as a signalling scaffold following the release of mechanical tension. In worms lacking TwcK activity, TwcK may therefore be quicker to relax into its pre-tension state and thus any interactions formed with binding partners in the mechanically deformed state cannot be preserved for any longer than the mechanical tension lasts. While considerably more work is required to study the function of TwcK, it appears likely that TwcK is tightly

regulated by a combination of NL and CRD tails acting in response to mechanical stimuli as well as biochemical regulation via autophosphorylation.

Chapter 6

References

Adams, J.A. (2001) "Kinetic and catalytic mechanisms of protein kinases." *Chemical Reviews* **101**(8): 2271-2290

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D. (2012) "Towards automated crystallographic structure refinement with phenix.refine." *Acta Crystallographica Section D Structural Biology* **68**: 352-367

Allen, B.G., Walsh, M.P. (1994) "The biochemical basis of the regulation of smooth-muscle contraction." *Trends in Biochemical Sciences* **19**(9): 362-268

Bailey, F.P., Byrne, D.P., Oruganty, K., Eyers, C.E., Novotny, C.J., Shokat, K.M., Kannan, N., Eyers, P.A. (2015) "The Tribbles 2 (TRB2) pseudokinase binds to ATP and autophosphorylates in a metal-independent manner." *Biochemical Journal* **467**(1): 47-62

Barden, J.A., Sehgal, P., Kemp, B.E. (1996) "Structure of the pseudosubstrate recognition site of chicken smooth muscle myosin light chain kinase." *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1292**(1): 106-112

Bayliss, R., Haq, T., Yeoh, S. (2015) "The Ys and wherefores of protein kinase autoinhibition." *Biochimica et Biophysica Acta - Proteins and Proteomics* **1854**(10): 1586-1594

Beltrami, A., Chaikuad, A., Daga, N., Elkins, J.M., Mahajan, P., Savitsky, P., Vollmar, M., Krojer, T., Muniz, J.R.C., Fedorov, O., Allerston, C.K., Yue, W.W., Gileadi, O., von Delft, F., Liu, Q., Lee, W.H., Brian, B.D., Arrowsmith, C.H., Edwards, A.M., Weigelt, J., Bountra, C., Knapp, S., Bullock, A. (2011) "LIMK1 protein kinase in complex with staurosporine." *Annotation to PDB entry 3S95*. Accessible at : http://www.thesgc.org/sites/default/files/active/Isee/LIMK1A_3s95_v1_372c/LIMK1A_3s95_v1_372c_index.html

Bem, E.M., Bem, H., Reimchüssel, W. (1980) "Determination of phosphorus-32 and calcium-45 in biological samples by Cerenkov and liquid scintillation counting." *The International Journal of Applied Radiation and Isotopes* **31**(6): 371-374

Ben-Levy, R., Leighton, I.A., Doza, Y.N., Attwood, P., Morrice, N., Marshall, C.J., Cohen, P. "Identification of novel phosphorylation sites required for activation of MAPKAP kinase-2." *The EMBO Journal* **14**: 5920–5930

Ben-Shimon, A., Niv, M.Y. (2011) "Deciphering the arginine-binding preferences at the substrate-binding groove of Ser/Thr kinases by computational surface mapping." PLoS Computational Biology 7(11):e1002288

Benian, G.M., Kiff., J.E., Neckelmann, N., Moerman, D.G., Waterston, R.H. (1989) "Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*." *Nature* **342**(6245): 45-50

Benian GM, Tang X, Tinley TL (1996) "Twitchin and related giant Ig superfamily members of C elegans and other invertebrates." *Advances in Biophysics* **33**:183–197.

Benian, G.M., Epstein, H.F. (2011) "*Caenorhabditis elegans* muscle: a genetic and molecular model for protein interactions in the heart." *Circulation Research* **109**(9): 1082-1095

Berkemeier, F., Bertz, M., Xiao, S., Pinotsis, N., Wilmanns, N., Gräter, F., Rief, M. (2011) "Fast-folding alpha-helices as reversible strain absorbers in the muscle protein myomesin." *Proceedings of the National Academy of Sciences of the United States of America* **108**(34): 14139-14144

Blume-Jensen, P., Hunter, T. (2001) "Oncogenic kinase signalling." *Nature* **411**: 355-365

Böhmer, F.D., Uecker, A. (2009) "A substrate peptide for the FLT3 receptor tyrosine kinase" *British Journal of Haematology* **144**(1): 127-130

Bogomolovas, J., Gasch, A., Simkovic, F., Rigden, D.J., Labeit, S., Mayans, O. (2014) "Titin kinase is an inactive pseudokinase scaffold that supports MuRF1 recruitment to the sarcomeric M-line." *Open Biology* **4**(5): 140041

Borochov, N., Eisenberg, H., Kam, Z. (1981) "Dependence of DNA conformation on the concentration of salt." *Biopolymers* **20**(1): 231-235

Bossemeyer, D., Engh, R.A., Kinzel, V., Ponstingl, G., Huber, R. (1993) "Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24)." *The EMBO Journal* **12**: 849-850

Bradford, M.M. (1976) "Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Analytical Biochemistry* **72**: 248–254.

Bullard, B., Linke, W., Leonard, K. (2002) "Varieties of elastic protein in invertebrate muscles." *Journal of Muscle Research and Cell Motility* **23**(5-6): 435-447

Butler, T.M., Siegman, T.J. (2011) "A force-activated kinase in a catch smooth muscle." *Journal of Muscle Research and Cell Motility* **31**(5-6): 349-358

Carrera, A.C., Alexandrov, K., Roberts, T.M. (1993) "The conserved lysine of the catalytic domain of protein kinase is actively involved in the phosphotransfer reaction and not required for anchoring ATP." *Proceedings of the National Academy of Sciences of the United States of America* **90**(2): 442-446

Centner, T., Yano, J., Kimura, E., McElhinny, A.S., Pelin, K., Witt, C.C., Bang, M., Trombitas, K., Granzier, H., Gregorio, C.C., Sorimachi, H., Labeit, S. (2001) "Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain." *Journal of Molecular Biology* **306**: 717–726

Chen, V.B., Arendall III, B., Headd, J.J., Keedy, F.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. (2010) "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallographica* **D66**: 12-21

Chang, A.N., Mahajan, P., Knapp, S., Barton, H., Sweeney, L., Kamm, K.E., Stull, J.T. (2016) "Cardiac myosin light chain is phosphorylated by Ca^{2+}/calmodulin-dependent and -independent kinase activities." *Proceedings of the National Academy of Sciences of the United States of America* **113**(27): E3824-3833

Choi, B., Zocchi, G., Wu, Y., Chan, S., Jeanne Perry, L. (2005) "Allosteric control through mechanical tension." *Physical Review Letters* **95**(7): 078102

Choi, B., Zocchi, G. (2006) "Mimicking cAMP-dependent allosteric control of protein kinase A through mechanical tension." *Journal of the American Chemical Society* **128**(6): 8541-8548

Clark, K.A, McElhinny, A.S, Beckerle, M.C, Gregorio, C.C. (2002) "Striated muscle cytoarchitecture: an intricate web of form and function." *Annual Review of Cell and Developmental Biology* **18**: 637-706

Cohen, P. (2001) "The role of protein phosphorylation in human health and disease." *The FEBS Journal* **268**(19): 5001-5010

Cook, A., Lowe, E.D., Chrysina, E.D., Skamnaki, V.T., Oikonomakos, N.G., Johnson, L.N. (2002) "Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: implications for the protein kinase mechanism." *Biochemistry* **41**: 7301-7311

Cox, A., Radzio,-Andzelm, E., Taylor, S.S. (1994) "Domain movements in protein kinases." *Current Opinion in Structural Biology* **4**(6): 893-901

Cox, J.S., Shamu, C.E., Walter, P. (1993) "Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase" *Cell* **73**: 1197-1206

Davis, R.J. (1995) "Transcriptional regulation by MAP kinases." *Molecular Reproduction and Development* **42**(4): 459-467

Dibb, N.J., Brown, D.M., Karn, J., Moerman, D.G., Bolten, S.L., Waterston, R.H. (1985) "Sequence analysis of muations that affect the synthesis, assembly and enzymatic activity of the *un-54* myosin heavy chain of *Caenorhabditis elegans*." *Journal of Molecular Biology* **183**(4): 543-551

Dissmeyer, N., Schnittger, A. (2011) "Use of phospho-site substitutions to analyze the biological revlevance of phosphorylation events in regulatory networks." *Methods in Molecular Biology* **779**: 93-138

Dorey, K., Engen, J.R., Kretzschmar, J., Wilm, M., Neubauer, G., Schindler, T., Superti-Furga, G. (2001) "Phosphorylation and structture-based functional studies reveal a positive and a negative role for the activation loop of the c-Abl tyrosine kinase." *Oncogene* **20**(56): 8075-8084

Dougherty, M.K., Müller, J., Ritt, D.A., Zhou, M., Zhou, X.Z., Copeland, T.D., Conrads, T.P., Veenstra, T.D., Lu, K.P., Morrison, D.K. (2005) "Regulation of Raf-1 by direct feedback phosphorylation." *Molecular Cell* **17**(2): 215-224

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010) "Features and development of Coot." *Acta Crystallographica Section D Structural Biology* **66**: 486-501

Endicott, J.A., Noble, M.E.M., Johnson, L.N. (2012) "The structural basis for control of eukaryotic protein kinases." *Annual Review of Biochemistry* **81**: 587-613

Fisher, R.P., Morgan, D.O (1994) "A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase." *Cell* **78**: 713-724

Forbes, J.G., Flaherty, D.B., Ma, K., Qadota, H., Benian, G.M., Wang, K. (2010) "Extensive and modular intrinsically disordered segments in *C. elegans* TTN-1 and implications in filament binding, elasticity and oblique striation." *Journal of Molecular Biology* **398**(5): 672-689

Freeman, A.D.J., Ward, R., El Mkami, H., Lilley, D.M.J., Norman, D.G. (2011) "Analysis of conformational changes in the DNA junction-resolving enzyme T7 endonuclease I on binding a four-way junction, using EPR." *Biochemistry* **50**(46): 9963-9972

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiacovo, M.P., Church, G.M., Calarco, J.A. (2013) "Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system." *Nature Methods* **10**: 741-743

Funabara, D., Hamamoto, C., Yamamoto, C., Yamamoto, K., Inoue, A., Ueda, M., Osawa, R., Kanoh, R., Hartshorne, D.J., Suzuki, S., Watabe, S. (2007) "Unphosphorylated twitchin forms a complex with actin and myosin that may contribute to tension maintenance in catch." *Journal of Experimental Biology* **210**(24): 4399-4410

Gallagher, P.J., Herring, B.P., Trafny, A., Sowadski, J., Stull, J.T. (1993) "A molecular mechanism for autoinhibition of myosin light chain kinases." *Journal of Biological Chemistry* **268**: 26578-26582

Gally, C., Wissler, F., Zahreddine, H., Quintin, S., Landmann, F., Labouesse, M. (2009) "Myosin II regulation during *C. elegans* embryonic elongation: LET-502/ROCK, MRCK-1 and PAK-1, three kinases with different roles." *Development* **136**: 3109-3119

Gautel, M., Castiglione Morelli, M.A., Pfuhl, M., Motta, A., Pastore, A. (1995) "A calmodulin-binding sequence in the C-terminus of human cardiac titin kinase" *European Journal of Biochemistry* **230**(2): 752-759

Gerlits, O., Waltman, M.J., Taylor, S., Langan, P., Kovalevsky, A. (2013) "Insights into the phosphoryl transfer catalyzed by cAMP-dependent protein kinase: an X-ray crystallographic study of complexes with various metals and peptide substrate SP20." *Biochemistry* **52**(21): 3721-3727

Getz, T.M., Dangelmaier, C.A., Jin, J., Daniel, J.L., Kunapuli, S.P. (2011) "Differential phosphorylation of myosin light chain (Thr)18 and (Ser)19 and functional implications in platelets." *Journal of Thrombosis and Haemostasis* **8**(10): 2283-2293

Gibbs, C.S., Zoller, M.J. (1991) "Rational scanning mutagenesis of a protein kinase identifies function regions involved in catalysis and substrate recognition." *Journal of Biological Chemistry* **266**(14): 8923-8931

Gräter, F., Shen, J., Jiang, H., Gautel, M., Grubmüller, H. (2005) "Mechanically induced titin-kinase activation studied by force-probe molecular dynamics simulations." *Biophysical Journal* **88**(2): 790-804

Graves, P.R., Winkfield, K.M., Haystead, T.A. (2005) "Regulation of zipper-interacting protein kinase activity in vitro and in vivo by multisite phosphorylation." *Journal of Biological Chemistry* **280**(10): 9363-9374

Greene, D.N., Garcia, T., Sutton, R.B., Gernert, K.M., Benian, G.M., Oberhauser, A.F. (2008) "Single-molecule force spectroscopy reveals a stepwise unfolding *Caenorhabditis elegans* giant protein kinase domains." *Biophysical Journal* **95**(3): 1360-1370

Hanson, P.I., Schulman, H. (1992) "Inhibitory autophosphorylation of multifunctional Ca2+/calmodulin-dependent protein kinase analyzed by site-directed mutagenesis." *Journal of Biological Chemistry* **267**(24): 17216-172214

Heierhorst, J., Probst, W.C., Vilim, F.S., Buku, A., Weiss, K.R. (1994) "Autophosphorylation of molluscan twitchin and interaction of its kinase domain with calcium/calmodulin." *Journal of Biological Chemistry* **269**(33): 21086-21093

Heierhorst, J., Probst, W.C., Kohanski, R.A., Buku, A., Weiss, K.R. (1995) "Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase." *European Journal of Biochemistry* **233**(2): 426-31

Heierhorst, J., Kobe, B., Feil, S.C., Parker, M.W., Benian, G.M., Weiss, K.R., Kemp, B.E. (1996a) "Cas2+/S100 regulation of giant protein kinases." *Nature* **380**(6575): 636-639

Heierhorst, J., Tang, X., Lei, J., Probst, W.C., Weiss, K.R., Kemp, B.E., Benian, G.M. (1996b) "Substrate specificity and inhibitor sensitivity of Ca2+/S100-dependent twitchin kinases." *European Journal of Biochemistry* **242**(3): 454-459

Hemmer, W., McGlone, M., Tsigelny, I., Taylor, S.S. (1997) "Role of the glycine triad in the ATP-binding site of cAMP-dependent protein kinase." *Journal of Biochemistry* **272**(2): 16946-16954

Ho, B.K., Dill, K.A. (2006) "Folding very short peptides using molecular dynamics." *PLoS Computational Biology* **2**(4): e27

Hu, S.H., Parker, M.W., Lei, J.Y., Wilce, M.C., Benian, G.M., Kemp, B.E. (1994) "Insights into autoregulation from the crystal structure of twitchin kinase." *Nature* **369**(6481): 581-585

Hu, L.Y., Kontrogianni-Konstantopoulos, A. (2013) "The kinase domains of obscurin interact with intercellular adhesion proteins." *The FASEB Journal* **27**: 2001-2012

Huang, H., Zhao, R., Dickson, B.M., Skeel, R.D., Post, C.B. (2012) "αC helix as a switch in the conformational transition of Src/CDK-like kinase domains." *The Journal of Physical Chemistry B* **116**(15): 4465-4475

Huse, M., Kuriyan, J. (2002) "The conformational plasticity of protein kinases." *Cell* **109**: 275-282

Ikebe, M. "Phosphorylation of a second site for myosin light chain kinase on platelet myosin" *Biochemistry* **28**(10): 2283-2293

Ikebe, M., Maruta, S., Reardon, S. (1989) "Location of the inhibitory region of smooth muscle myosin light chain kinase." *Journal of Biological Chemistry* 264: 6967-6971

Ikuta, M., Kornienko, M., Byrne, N., Reid, J.C., Mizuarai, S., Kotani, H., Munshi, S.K. (2007) "Crystal structures of the N-terminal kinase domain of human RSK1 bound to three different ligands: Implications for the design of RSK1 specific inhibitors." *Protein Science* **16**(12): 2626-2635

Ito, M., Guerriero V. Jr., Chen, X. M., Hartshorne, D. J. (1991) "Definition of the

inhibitory domain of smooth muscle myosin light chain kinase by site-directed mutagenesis." *Biochemistry.* **30**: 3498-3503

Itoh, K., Hara, T., Shibata, N. (1992) "Diphosphorylation of platelet myosin by myosin light chain kinase." *Biochimica et Biophyisica Acta - Molecular Cell Research* **1133**(3): 286-292

Iyver, G.H., Garrod, S., Woods, V.L., Taylor, S.S. (1991) "Catalytic independent functions of a protein kinase as revealed by a kinase-dead mutant: study of the Lys72His mutant of cAMP-dependent kinase." *Journal of Molecular Biology* **352**(5): 1110-1122

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., Pavletich, N.P. (1995) "Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex." *Nature* **376**: 323-320

Jeschke, G. (2002) "Distance measurements in the nanometer range by pulse EPR." *ChemPhysChem* **3**(11): 927-932

Johnson, L.N., Noble, M.E., Owen, D.J. (1996) "Active and inactive protein kinases: structural basis for regulation." *Cell* **85**(2): 149-158

Johnson, L.N., Lewis, R.J. (2001) "Structural basis for control by phosphorylation." *Chemical Reviews* **101**: 2209-2242

Johnson, D.A., Akamine, P., Radzio-Andzelm, E., Madhusudan, M., Taylor, S.S. (2001) "Dynamics of cAMP-dependent protein kinase" *Chemical Reviews* **101**: 2243-2270

Joseph, C., Tsenf, C.Y., Zocchi, G., Tlusty, T. (2014) "Asymmetric effect of mechanical stress on the forward and reverse reaction catalyzed by an enzyme." *PLoS One* **9**(7): e101442

Joshi, A., Newbatt, Y., McAndrew, P.C., Stubbs, M., Burke, R., Richards, M.W., Bhatia, C., Caldwell, J.J., McHardy, T., Collins, I., Bayliss, R. (2015) “Molecular mechanisms of human IRE1 activation through dimerization and ligand binding” *Oncotarget* **6**(15): 13019-13035

Kabsch, W. (201) "XDS." *Acta Crystallographica Section D Structural Biology* **66**: 125-132

Kamm, K.E., Stull, J.T. (2001) "Dedicated myosin light chain kinase with diverse cellular functions" *Journal of Biological Chemistry* **276**(7): 4527-4230

Kamm, K., Stull, J. (2011) "Signaling to myosin regulatory light chain in sarcomeres." *Journal of Biological Chemistry* **286**(12): 9941-9947

Kannan, N., Neuwald, A.F. (2005) "Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component?" *Journal of Molecular Biology* **351**(5): 956-972

Kannan, N., Neuwald, A.F., Taylor, S.S. (2009) "Analogous regulatory sites within the αC-β4 loop regions of ZAP-70 tyrosine kinase and AGC kinases." *Biochimica et Biophysica Acta - Proteins and Proteomics* **1784**(1): 27-32

Katzemich, A., West, R.J.H., Fukuzawa, A., Sweeney, S.T., Gautel, M., Sparrow, J., Bullard, B. (2015) "Binding partners of the kinase domains in *Drosophila* obscurin and their effect on the structure of the flight muscle." *Journal of Cell Science* **128**(18): 3386-3397

Kemp, B.E., Pearson, R.B. (1985) "Spatial requirements for location of basic residues in peptide substrates for smooth muscle myosin light chain kinase." *Journal of Biological Chemistry* **260**: 3355-3359

Kemp, B.E., Pearson, R.B., Guerriero Jr, V., Bagchi, I.C., Means, A.R. (1987) "The calmodulin binding domain of chicken smooth muscle myosin light chain kinase contains a pseudosubstrate sequence." *Journal of Biological Chemistry* **262**(6): 2542-2548

Kenyon, C.P., Roth, R.L., van der Westhuyzen, C.W., Parkinson, C.J. (2012) "Conserved phosphoryl transfer mechanisms within kinase families and the role of the C8 proton of ATP in the activation of phosphoryl transfer." *BMC Research Notes* **5**: 131

Knighton, D. R., Zheng, J. H., Eyck, Ten, L. F., Xuong, N. H., Taylor, S. S., Sowadski, J. M. (1991) "Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase." *Science* **253**: 414–420.

Knighton, D.R., Pearson, R.B., Sowadski, J.M., Means, A.R., Ten Eyck., L.F., Taylor, S.S., Kemp, B.E. (1992) "Structural basis of the intrasteric regulations of myosin light chain kinases." *Science* **268**(5079): 130-135

Kobe, B., Heierhorst, J., Feil, S.C., Parker, M.W., Benian, G.M., Weiss, K.R., Kemp, B.E. (1996) "Giant protein kinases: domain interactions and structural basis of autoregulation." *The EMBO Journal* **15**(24): 6810-6821

Kobe, B., Kampmann, T., Forwood, J.K., Listwan, P., Brinkworth, R.I. (2005) "Substrate specificity of protein kinases and computation prediction of substrates." *Biochimica et Biophysica Acta - Proteins and Proteomics* **1754**(1-2): 200-209

Koch, C.A., Anderson, D., Moran, M.F., Ellis, C., Pawson, T. (1991) "SH2 and SH3 domains: elements that control interactions of cytoplasmic signalling proteins." *Science* **252**(5006): 668-674

Kolodney, M. S., Elson, E. L. (1993) "Correlation of myosin light chain phosphorylation with isometric contraction of fibroblasts." *Journal of Biological Chemistry* **268:** 23850-23855

Komatsu, S., Hosoya, H. (1996) “Phosphorylation by MAPKAP kinase 2 activates Mg(2+)-ATPase activity of myosin II.” *Biochemical and Biophysical Research Communications* **223**(3): 741-745

Kontrogianni-Konstantopoulos, A., Ackerman, M.A., Bowman, A.L., Yap, S.V., Bloch, R.J. (2009) "Muscle giants: Molecular scaffolds in sarcomerogenesis" *Physiological Reviews* **89**: 1217-1267

Korennykh, A.V., Egea, P.D., Korostelev, A.A., Finer-Moore, J., Zhang, C., Shokat, K.M., Stroud, R.M., Walter, P. (2009) “The unfolded protein response signals through high-order assembly of Ire1” *Nature* **457**: 687-693

Kornev, A.P., Haste, N.M., Taylor, S.S., Eyck, L.F. (2006) "Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism." *Proceedings of the National Academy of Sciences of the United States of America* **103**(47): 17783-17788

Krueger, J.K., Padre, R.C., Stull, J.T. (1995) "Intrasteric regulation of myosin light chain kinase." *Journal of Biological Chemistry* **270**: 16848-16853

Krueger, J.K., Olah, G.A., Rokop, S.E., Zhi, G., Stull, J.T., Trewhella, J. (1997) "Structures of calmodulin and a functional myosin light chain kinase in the activated complex: a neutron scatterin study." *Biochemistry* **36**(20): 6017-6023

Krueger, J.K., Zhi, G., Stull, J.T., Trewhella, J. (1998) "Neutron-scattering studies reveal further details of the Ca2+calmodulin-dependent activation mechanism of myosin light chain kinase." *Biochemistry* **37**(40)L 13997-14004

Krüger, M., Linke, W.A. (2011) "The giant protein titin: a regulatory node that integrates myocyte signaling pathways" *The Journal of Biological Chemistry* **286**(12): 9905-9912

Lange, S., Xiang, F., Yakovenko, A., Vihola, A., Hackman, P., Rostkova, E., Kristensen, J., Brandmeier, B., Franzen, G., Hedberg, B., Gunnarsson, L., Hughes, S., Marchand, S., Sejersen, T., Richard, I., Edstrom, L., Ehler, E., Udd, B., Gautel, M. (2005) "The kinase domain of titin controls muscle gene expression and protein turnover." *Science* **308**: 1599-1603

Lange, S., Perera, S., The, P., Chen, J. (2012) "Obscurin and KCTD6 regulate cullin-dependent small ankyrin-1 (sAnK1.5) protein turnover." *Molecular Biology of the Cell* **23**: 2490-2504

Lehman, W., Craig, R., Vibertt, P. (1994) "Ca2+-induced tropomyosin movement in *Limulus* thin filaments revealed by three-dimensional reconstruction." *Nature* **368**: 65–67

Lee, K.P., Dey, M., Neculai, D., Cao, C., Dever, T.E., Sicheri, F. (2008) “Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing” *Cell* **132**: 89-100

Lei, J., Tang, X., Chambers, T.C., Pohl, J., Benian, G.M. (1994) "Protein kinase domain of twitchin has protein kinase activity and an autoinhibitory region" *Journal of Biological Chemistry* **269**(33): 21078-21085

Leung, T., Chen, X.Q., Manser, E., Lim, L. (1996) "The p160 RhoA-binding kinase ROK alpha is a member of a kinase family and is involved in the reorganization of the cytoskeleton." *Molecular Cell Biol*ogy **16**: 5313–5327.

Li, F., Gangal, M., Juliano, C., Gorfain, E., Taylor, S.S., Johnson, D.A. (2002) "Evidence for an internal entropy contribution to phosphoryl transfer: a study of domain closure, backbone flexibility, and the catalytic cycle of cAMP-dependent protein kinase." *Journal of Molecular Biology* **315**: 459-469

Lin, X., Qadota, H., Moerman, D.G., Williams, B.D. (2003) "*C. elegans* PAT-6/actopaxin plays a critical role in the assembly of integrin adhesion complexes *in vivo*." *Current Biology* **13**: 922-932

Lin, Z., Jia, L., Tomchick, D.R., Luo, Z., Yu, H. (2014) "Substrate-specific activation of the mitotic kinase Bub1 through intramolecular autophosphorylation and kinetochore targeting." *Structure* **22**(11): 1616-1627

Linke, W.A., Kulke, M., Li, H., Fujita-Becker, S., Neagoe, C., Manstein, D.J., Gautel, M., Fernandez, J.M. (2002) "PEVK domain of titin: an entropic spring with actin-binding properties." *Journal of Structural Biology* **137**(1-2): 194-205

Linke, W.A., Krüger, M. (2010) "The giant protein titin as an integrator of myocyte signalling pathways" *Physiology (Bethesda)* **25**(3): 186-1898

Lochhead, P.A. (2009) "Protein kinase activation loop autophosphorylation in cis: overcoming a catch-22 situation." *Science Signalling* **2**(54): pe4

Lowe, E.D., Noble, M.E., Skamnaki, V.T., Oikonomakos, N.G., Owen, D.J., Johnson, L.N. (1997) "The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition." *The EMBO Journal* **16**: 6646-6658

Luther, P. (2009) "The vertebrate muscle Z-disc: Sarcomere anchor for structure and signalling." *Journal of Muscle Research and Cell Motility* **30**: 171-185

Mackenzie, J.M., Epstein, H.F. (1980) "Paramyosin is necessary for determination of nematode thick filament length in vivo." *Cell* **22**:747–755

Madhusudan, M., Trafny, E.A., Xuong, N., Adams, J.A., Ten Eyck, L.F., Taylor, S.S., Sowadski, J.M. (1994) "cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer" *Protein Science* **3**: 176-187

Mannan, M.A., Shadrick, W.R., Biener, G., Shin, B., Anshu, A., Raicu, V., Frich, D.N., Dey, M. (2013) “An Ire1-Phk1 chimera reveals a dispensable role of autokinase activity in endoplasmic reticulum stress response” *Journal of Molecular Biology* **425**: 2083-2099

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. "The protein kinase complement of the human genome." *Science* **298**(5600): 1912-1934

Matsunaga, Y., Qadota, H., Furukawa, M., Choe, H.H., Benian, G.M. (2015) "Twitchin kinase interacts with MAPKAP kinase 2 in Caenorhabditis elegans striated muscle." *Molecular Biology of the Cell* **26**(11): 2096-2111

Matte, A., Tari, L.W., Delbaere, L.T. (1998) "How do kinases transfer phosphoryl groups?" *Structure* **6**: 413-419

Mayans, O., van der Ven, P.F., Wilm, M., Mues, A., Young, P., Fürst, D.O., Wilmanns, M., Gautel, M. (1998) "Structural basis for activation of the titin kinase domain during myofibrillogenesis" *Nature* **395**(6705): 863-869

Mayans, O., Benian, G.M., Simkovic, F., Rigden, D.J. (2013) "Mechanistic and functional diversity in the mechanosensory kinases of the titin-like family." *Biochemical Society Transactions* **41**(4): 1066-1071

McClendon, C.L., Kornev, A.P., Gilson, M.K., Taylor, S.S. (2014) "Dynamic architecture of a protein kinase." *Proceedings of the National Academy of Sciences of the United States of America* **111**(43): E4623-4631

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007) "Phaser crystallographic software." *Journal of Applied Crystallography* **40**: 658-674

McNamara, L.K., Watterson, D.M., Brunzelle, J.S. (2009) "Structural insight into nucleotide recognition by human death-associated protein kinase." *Acta Crystallographica Section D Structural Biology* **65**(3): 241-248

Meggio, F., Donella Deana, A., Ruzzene, M., Brunati, A.M., Cesaro, L., Guerra, B., Meyer, T., Mett, H., Fabbro, D., Furet, P. (1995) "Different susceptibility of protein kinases to staurosporine inhibition. Kinetic studies and molecular basis for the resistance of protein kinase CK2." *Europan Journal of Biochemistry* **234**(1): 317-322

Michnoff, C. H., Kemp, B. E., Stull, J. T. (1986) "Phosphorylation of synthetic peptides by skeletal muscle myosin light chain kinases." *Journal of Biological Chem*istry **261**: 8320-8326.

Moerman, D.G., Plurad, S., Waterston, R.H. (1982) "Mutations in the *unc-54* myosin heavy chain gene of Caenorhabditis elegans that alter contractility but not muscle structure." *Cell* **29**(3): 773-781

Moerman, D.G., Benian, G.M., Barstead, R.J., Schriefer, L.A., Watserston, R.H. (1988) "Identification and intracellular localization of the unc-22 gene product of Caenorhabditis elegans." *Genes & Development* **2**(1): 93-105

Mrosek, M., Labeit, D., Witt, S., Heerklotz, H., von Castelmur, E., Labeit, S., Mayans, O. (2007) "Molecular determinants for the recruitments of the ubiquitin-ligase MuRF-1 onto M-line titin." *The FASEB Journal* **21**: 1383-1392

Müller, S., Chaikuad, A., Gray, N.S., Knapp, S. (2015) "The ins and outs of selective kinase inhibitor development." *Nature Chemical Biology* **11**(11): 818-821

Murphy, S.M., Bergman, M., Morgan, D.O. (1993) "Suppression of c-Src activity by C-terminal Src kinases involves the SH2 and SH3 domains: analysis with Saccharomyces cervisiae." *Molecular Cell Biology* **13**(9): 5290-5300.

Neuman, K.C., Nagy, A. (2008) "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy." *Nature Methods* **5**(6): 491-505

Ngok, S.P., Geyer, R., Kourtidis, A., Storz, P., Anastasiadis, P.Z. (2013) "Phosphorylation-mediated 14-3-3 protein binding regulates the function of the rho-specific guanine nucleotide exchange facter (RhoGEF) Syx." *Journal of Biological Chemistry* **288**(9): 6640-6650

Nolen, B., Ngo, J., Chakrabarti, S., Vu, D., Adams, J.A., Ghosh, G. (2003) "Nucleotide-induced conformational changes in the Saccharomyces cerevisiae SR protein kinase, Sky1p, revealed by X-ray crystallography" *Biochemistry* **42**: 9575-9585

Olson, N. J., Pearson, R. B., Needleman, D. S., Hurwitz, M. Y., Kemp, B. E., Means, A. R. (1990) "Regulatory and structural motifs of chicken gizzard myosin light chain kinase." *Proceedings of the National Academy of Sciences of the United States of America* **87**: 2284-2288

Pavletich, N.P. (1999) "Mechanisms of cyclin-dpendent kinase regulation: structures of cdks, their cyclin activators and cip and INK4 inhibitors." *Journal of Molecular Biology* **287**(5): 821-828

Pearce, L.R., Komander, D., Alessi, D.R. (2010) "The nuts and bolts of AGC protein kinases." *Nature Reviews Molecular Cell Biology* **11**: 9-22

Pearson, R. B., Wettenhall, R. E., Means, A. R., Hartshorne, D. J., Kemp, B. E. (1988) "Autoregulation of enzymes by pseudosubstrate prototopes: myosin light chain kinase." *Science* **241:** 970-973

Pike, A., Rellos, P., Niesen, F.H., Turnbull., A., Oliver, A.W., Parker, S.A., Turk, B.E., Pearl, L.H., Knapp, S. (2008) "Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites." *The EMBO Journal* **27**: 704-714

Pines, J. (1994) "Protein kinases and cell cycle control." *Seminars in Cell Biology* **5**(6): 399-408

Pinna, L.A., Ruzzene, M. (1996) "How do protein kinases recognize their substrates?" *Biochimica et Biophysica Acta - Molecular Cell Research* **1314**(3): 191-225

Pintard, L., Willis, J.H., Willems, A., Johnson, J.L., Srayko, M., Kurz, T., Glaser, S., Mains, P.E., Tyers, M., Bowerman, B., Peter, M. (2003) "The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase." *Nature* **425**: 311-316

Prade, L., Engh, R.A., Girod, A., Kinzel, V., Huber, R., Bossemeyer, D. (1997) "Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential." *Structure* **5**(12): 1627-1637

Puchner, E.M., Alexandrovich, A., Kho, A.L., Hensen, U., Schäfer, L.V., Brandmeier, B., Gräter, F., Grubmüller, H., Gaub, H.E., Gautel, M. (2008) "Mechanoenzymatics of titin kinase." *Proceedings of the National Acacdemy of Sciences of the United States of America* **105**(36): 13385-13390

Puchner, E.M., Gaub, H.E. (2010) "Exploring the conformation-regulated function of titin kinase by mechanical pump and probe experiments with single molecules." *Angewandte Chemie Internation Edition in English* **49**(6): 1147-1150

Qadota, H., McGaha, L.A., Mercer, K.B., Stark, T.J., Ferrara, T.M., Benian, G.M. (2008) "A novel protein phosphatase is a binding partner for the protein kinase domains of UNC-89 (obscurin) in *Caenorhabditis elegans*." *Molecular Biology of the Cell* **19**:2424-2432

Roskoski Jr., R. (2004) "Src kinase regulation by phosphorylation and dephosphorylation." *Biochemical and Biophysical Research Communications* **331**(1): 1-14

Schmidt, M.J., Fedoseev, A., Bücker, D., Borbas, J., Peter, C., Drescher, M., Summerer, D. (2015) "EPR distance measurement in native proteins with genetically encoded spin labels." *ACS Chemical Biology* **10**(12): 2764-2771

Schwaiger, I., Sattler, C., Hostetter, D.R., Rief, M. (2002) "The myosin coiled-coil is a truly elastic protein structure." *Nature Materials* **1**(4): 232-235

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L. (2005) "The FoldX web server: an online force field." *Nucleic Acids Research* **33**: W382-W388

Seol, Y., Carpenter, A.E., Perkins, T.T. (2006) "Gold nanoparticles: enhanced optical trapping and sensitivity coupled with significant heating." *Optics Letters* **31**(16): 2429-2431

Shamu, C.E & Walter, P. (1996) "Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus" *The EMBO Journal* **15**: 3028-3039

Shigaki, S., Yamaji, T., Han, X., Yamanouchi, G., Sonoda, T., Okitsu, O., Mori, T., Niidome, T., Katayama, Y. (2007) "A peptide microarray for the detection of protein kinase activity in cell lysate" *Analytical Science* **23**(3): 271-275

Sicheri, F., Kuriyan, J. (1997) "Structures of Src-family tyrosine kinases." *Current Opinion in Structural Biology* **7**: 777-785

Siegman, M.J., Funabara, D., Kinoshita, S., Watabe, S., Hartshorne, D.J., Butler, T.M. (1998) "Phosphorylation of a twitchin-related protein control catch and calcium sensitivity of force production in invertebrate smooth muscle." *Proceedings of the National Academy of Sciences of the United States of America* **95**(9): 5383-5388

Sjostrom, M., Squire, J.M. (1977) "Cryo-ultramicrotomy and myofibrillar fine structure: a review." *Journal of Microscopy* **111**(3): 239-278.

Small, T.M., Gernert, K.M., Flaherty, D.B., Mercer, K.B., Borodovsky, M., Benian, G.M. (2004) "Three new isoforms of *Caenorhabditis elegans* UNC-89 containing MLCK-like protein kinase domains." *Journal of Molecular Biology* **342**: 91–108

Stahl, S.W., Puchner, E.M., Alexandrovich, A., Gautel, M., Gaub, H.E. (2011) "A conditional gating mechanism assures the integrity of the molecular force-sensor titin kinase." *Biophysical Journal* **101**(8): 1978-1986

Steichen, J.M., Iyver, G.H., Li, S., Saldanha, S.A., Deal, M.S., Woods Jr, V.L., Taylor, S.S. (2009) "Global consequences of activation loop phosphorylation of protein kinase A." *Journal of Biological Chemistry* **285**(6): 3825-3832

Steichen, J.M., Iyver, G.H., Li, S., Saldanha, S.A., Deal, M.S., Woods Jr., V.L., Taylor, S.S. (2010) "Global consequences of activation loop phosphorylation on protein kinase A." *Journal of Biological Chemistry* **285**(6): 3825-3832

Steichen, J.M., Kuchinskas, M., Keshwani, M.M., Yang, J., Adams, J.A., Taylor, S.S. (2012) "Structural basis for the regulation of protein kinase A by activation loop phosphorylation." *Journal of Biological Chemistry* **287**(18): 14672-14680

Sweeney, H.L., Bowman, B.F., Stull, J.T. (1993) "Myosin light chain phosphorylation in vertebrate striated muscle: regulation and function." *American Journal of Physiology* **264**(5): 1085-1095

Taylor, S.S., Kornev, A.P. (2011) "Protein kinases: evolution of dynamic regulatory proteins." *Trends in Biochemical Sciences* **36**(2): 65-77

The *C. elegans* sequencing consortium. (1998) "Genome sequence of the nematode *C. elegans*: A platform for investigating Biology." *Science* **282**: 2012-2018

Thiele, A., Stangl, G.I., Schutkowski, M. (2011) "Deciphering enzyme function using peptide arrays" *Molecular Biotechnology* **49**(3): 283-305

Thomas, J.E., Soriano, P., Brugge, J.S. (1991) "Phosphorylation of c-Src on tyrosine 527 by another protein tyrosine kinase." *Science* **254**(5031): 568-571

Timm, T., Li, X.Y., Biernat, J., Jiao, J., Mandelkow, E., Vandekerckhove, J., Mandelkow, E.M. (2003) "MARKK, a Ste20-like kinase, activates the polarity-inducing kinase MARK-PAR-1" *The EMBO Journal* **22**(19): 5090-5101

Treiber, D.K., Shah, N.P. (2013) "Ins and outs of kinase DFG motifs." *Chemistry & Biology* **20**(6): 745-746

Tseng, C., Wang, A., Zocchi, G., Rolih, B., Levine, A.J. (2009) "Elastic energy of protein-DNA chimeras." *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* **80**(6 Pt 1): 061912

Tseng, C.Y., Zocchi, G. (2013) "Mechanical control of Renilla luciferase." *Journal of the American Chemical Society* **135**(32): 11879-11886

Vibert, P., Edelstein, **S.** M., Castellani, L. & Elliot Jr., B. W. (1993) "Mini-titins in striated and smooth molluscan muscles: structure, location and immunological crossreactivity." *Journal of Muscle Research and Cell Motility* **14**: 598-607.

von Castelmur, E., Strümpfer, J., Franke, B., Bogomolovas, J., Barbieri, S., Qadota, H., Konarev, P.V., Svergun, D.I., Labeit, S., Benian, G.M., Schulten, K., Mayans, O. (2012) "Identification of an N-terminal inhibitory extension as the primary mechanosensory regulator of twitchin kinase" *Proceedings of the National Academy of Sciences of the United States of America* **109**(34): 13608-13613

Vulpetti, A., Bosotti, R. (2004) "Sequence and structural analysis of kinase ATP pocket residues." *Il Farmaco* **59**(10):759-765

Wallace, A.C., Laskowski, R.A., Thornton, J.M. (1994) "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions." *Protein Engineering* **8**(2): 127-134

Wang, C., Shang, Y., Yu, J., Zhang, M. (2012) "Substrate recognition mechanism of atypical protein kinase Cs revealed by the structure of PKCι in complex with a substrate peptide from Par-3." *Structure* **20**: 791-801

Warner, A., Xiong, G., Qadota, H., Rogalski, T., Vogl, A.W., Moerman, D.G., Benian, G.M. (2013) "CPNA-1, a copine domain protein, is located at integrin adhesion sites and is required for myofilament stability in *Caenorhabditis elegans*." *Molecular Biology of the Cell* **24**: 601-616

Waterston, R.H., Thomson, J.N., Brenner, S. (1980) "Mutants with altered muscle structure of Caenorhabditis elegans." *Developmental Biology* **77**(2): 271-302

Wilson, K.J., Qadota, H., Mains, P.E., Benian, G.M. (2012) "UNC-89 (obscurin) binds to MEL-26, a BTB-domain protein, and affects the function of MEI-1 (katanin) in striated muscle of *Caenorhabditis elegans*." *Molecular Biology of the Cell* **23**: 2623-2634

Wolny, M., Batchelor, M., Knight, P.J., Paci, E., Dougan, L., Peckam, M. (2014) "Stable single α-helices are constant force springs in proteins." *Journal of Biological Chemistry* **289**(4): 27825-27835

Xiong, G., Qadota, H., Mercer, K.B., McGaha, L.A., Oberhauser, A.F., Benian, G.M. (2009) "A LIM-9 (FHL)/SCPL-1 (SCP) complex interacts with the C-terminal protein kinase regions of UNC-89 (obscurin) in *Caenorhabditis elegans*." *Journal of Molecular Biology* **386**: 976-988

Xu, Z.B., Chaudhary, D., Olland, S., Wolfrom, S., Czerwinski, R., Malakian, K., Lin, L., Stahl., M.L., Joseph-McCarthy, D., Benander, C., Fitz, L., Greco, R., Somers, W.S., Mosyak, L. (2004) "Catalytic domain crystal structure of protein kinase C-theta (PKCtheta)." *Journal of Biological Chemistry* **279**(48): 50401-50409

Yamada, A., Yoshio, M., Kojima, H., Oiwa, K. (2001) "An in vitro assay reveals essential protein components for the "catch" state of invertebrate smooth muscle" *Proceedings of the National Academy of Sciences of the United States of America* **98**(12): 6635-6640

Yoshioka, K., Foletta, V., Bernard, O., Itoh, K. (2003) "A role for LIM kinase in cancer invasion." *Proceedings of the National Academy of Sciences of the United States of America* **100**(12): 7247-7252

Zhang, W., Liu, H.T. (2002) "MAPK signal pathways in the regulation of cell proliferation in mammalian cells." *Cell Research* **12**: 9-18

Zheng, J., Trafny, E.A., Knighton, D.R., Zuong, N., Taylor, S.S., Ten Eyck, L.F., Sowadski, J.M, (1993) "2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor" *Acta Crystallographica* **D49**: 362-365

Zhu, Y.L., Ai, Y., Gilchrist, A., Maulik, N., Watras, J., Sha'afi, R.J., Das, D.K., Huang, C.K. (1997) "High expression and cativation of MAP kinase-activated protein kinase 2 in cardiac muscle cells." *Journal of Molecular and Cellular Cardiology* **29**(8): 2159-2168

Zhu, G., Fuji, K., Liu, Y., Codrea, V., Herrero, J., Shaw, S. (2005) "A single pair of acidic residues in the kinase major groove mediates strong substrate preference for P-2 or P-5 arginine in the AGC, CAMK and STE families." *Journal of Biological Chemistry* **280**(43): 36372-36379

Zocchi, G. (2009) "Controlling proteins through molecular springs." *Annual Review of Biophysics* **38**: 75-88

Chapter 7

Appendices

7.1 Appendix I - Published papers and manuscript contributions

7.1.1 "Exploration of pathomechanisms triggered by a single-nucleotide polymorphism in titin's I-band: the cardiomyopathy-linked mutation T2580I"

Bogomolovas, J., Fleming, J., Anderson, B., Williams, R., Lange, S., Simon, B., Khan, M.M., Rudolf, R., Franke, B., Bullard, B., Rigden, D.J., Granzier, H., Labeit, S., Mayans, O.

Open Biology (2016) **6**: 160114

Personal contribution to this manuscript includes: crystallisation of the I9-I11 tandem from the I-band region of titin, consisting of three Ig domains. X-ray data collection at Diamond Light Source and structure elucidation of the titin I9-11 fragment in two different space groups: P1 and $P2_12_12_1$. Contribution to refinement of the I10 single Ig domain from the I-band region of titin. Crystal structures are deposited in the Protein Data Bank under the following entries: **5JDD**, Crystal structure of I9-I11 tandem from titin ($P2_12_12_1$); **5JDE**, Crystal structure of I9-I11 tandem from titin (P1); **5JDJ**, Crystal structure of domain I10 from titin in space group $P2_12_12_1$.

Contributions can be seen in the following sections: Methods; the crystal structure of I10 reveals a high-energy conformation for residue T2850; the crystal structure of I9-I11 shows that T2850 mediates interdomain contacts.

OPEN BIOLOGY

rsob.royalsocietypublishing.org

Research

Cite this article: Bogomolovas J *et al.* 2016 Exploration of pathomechanisms triggered by a single-nucleotide polymorphism in titin's I-band: the cardiomyopathy-linked mutation T2580I. *Open Biol.* **6**: 160114. http://dx.doi.org/10.1098/rsob.160114

Received: 18 April 2016
Accepted: 1 September 2016

Subject Area:
structural biology/cellular biology/genetics

Keywords:
cardiomyopathy, missense single-nucleotide polymorphism, titin protein structure, transgenic muscle, transgenic mouse model

Authors for correspondence:
Siegfried Labeit
e-mail: labeit@medma.de
Olga Mayans
e-mail: olga.mayans@uni-konstanz.de

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsob.160114.

THE ROYAL SOCIETY PUBLISHING

Exploration of pathomechanisms triggered by a single-nucleotide polymorphism in titin's I-band: the cardiomyopathy-linked mutation T2580I

Q1

Julius Bogomolovas[1,2], Jennifer Fleming[2,9], Brian R. Anderson[3], Rhys Williams[2,9], Stephan Lange[4], Bernd Simon[5], Muzamil M. Khan[6,7], Rüdiger Rudolf[6,7], Barbara Franke[9], Belinda Bullard[8], Daniel J. Rigden[2], Henk Granzier[3], Siegfried Labeit[1] and Olga Mayans[2,9]

[1]Department of Integrative Pathophysiology, Medical Faculty Mannheim, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany
[2]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK
[3]Department of Cellular and Molecular Medicine and Sarver Molecular Cardiovascular Research Program, University of Arizona, Tucson, AZ 85724, USA
[4]School of Medicine, University of California San Diego, 9500 Gilman Drive, MC-0613C, La Jolla, CA 92093, USA
[5]European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[6]Institute of Molecular and Cell Biology, Mannheim University of Applied Sciences, Paul-Wittsackstraße 110, 68163 Mannheim, Germany
[7]Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[8]Department of Biology, University of York, York YO10 5DD, UK
[9]Department of Biology, University of Konstanz, 78457 Konstanz, Germany

OM, 0000-0001-6876-8532

Missense single-nucleotide polymorphisms (mSNP) in titin are emerging as a main causative factor of heart failure. However, distinguishing between benign and disease-causing mSNPs is a substantial challenge. Here, we research the question of whether a single mSNP in a generic domain of titin can affect heart function as a whole and, if so, how. For this, we studied the mSNP, T2850I, seemingly linked to arrhythmogenic right ventricular cardiomyopathy (ARVC). We used structural biology, computational simulations and transgenic muscle *in vivo* methods to track the effect of the mutation from the molecular to the organismal level. The data show that the T2850I exchange is compatible with the domain three-dimensional fold, but that it strongly destabilizes it. Further, it induces a change in the conformational dynamics of the titin chain that alters its reactivity, causing the formation of aberrant interactions in the sarcomere. Echocardiography of knock-in mice indicated a mild diastolic dysfunction arising from increased myocardial stiffness. In conclusion, our data provide evidence that single mSNPs in titin's I-band can alter overall muscle behaviour. Our suggested mechanisms of disease are the development of non-native sarcomeric interactions and titin instability leading to a reduced I-band compliance. However, understanding the T2850I-induced ARVC pathology mechanistically remains a complex problem and will require a deeper understanding of the sarcomeric context of the titin region affected.

1. Introduction

Cardiovascular disease is the major cause of death worldwide. Evidence accumulated over the past two decades has identified titin as a main coordinator of

RSOB160114—12/9/16—20:29–Copy Edited by: Not Mentioned

cardiac muscle homeostasis, and its dysfunction through genetic mutation as an important factor in cardiomyopathy. Titin mutations have now been linked to familial dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC) (recently reviewed in [1]). Pathogenicity often results from mutations that lead to the truncation of the titin chain, as identified in 25% of familial cases of idiopathic DCM and 18% of sporadic cases [2]. Several missense single-nucleotide polymorphisms (mSNP) in titin have also been associated with heart disease. These comprise the ARVC-linked T2850I in the I-band [3]; the DCM-associated V54M, A743 V in the Z-disc [4] and W930R in the Z-disc/I-band transition zone [5]; R740 L [6] and S3799Y [7] that increase binding to α-actinin and FHL2, respectively, causing HCM; and Y7621C [8] located in the A/I-band junction linked to restrictive cardiomyopathies. Consequently, genetic screening of the titin gene (*TTN*) is clinically relevant in most cardiomyopathy cases [1,9]. However, while the robust discrimination between benign and disease-causing titin truncations can be achieved by coupling *TTN* exon inclusion data and the position of the mutation within the gene [10], predicting the pathogenicity of mSNPs in titin is considerably more complex. In contrast to truncations, mSNPs might result in individual molecular phenotypes and appear to trigger distinct pathomechanisms [11]. As a result, the differentiation between benign and pathogenic mSNP variants is currently very demanding, requiring the integration of functional assays, robust bioinformatics, large control cohorts and expert clinical evaluation [9].

In this work, we set to answer the question of whether single mSNPs in titin can generically and without the contribution of additional genetic factors lead to cardiac disease. For this, we examined the unique mSNP T2850I in the I-band domain of titin I10, as paradigm of an alteration in a general component of the titin chain. Domain I10 does not support specific interactions or have any known specialized roles, so that it does not constitute an *a priori* sensitive locus of the chain. The T2850I exchange has been associated with ARVC based on linkage studies, being completely segregated with the ARVC phenotype in nine patients from a large family, including two fifth-degree relatives, and was absent in 300 cardiomyopathy and 400 control chromosomes [3]. ARVC is characterized by life-threatening arrhythmias, being the main cause of sudden death in the population below 25 years of age. Mechanistically, ARVC is thought to result from a perturbed desmosomal force transmission [12]. Titin, however, is an intrasarcomeric protein not known to relate to cell adhesion. Thus, we asked whether mSNPs in titin can induce an ARVC phenotype. In previous studies, we have showed that the T2580I mSNP destabilizes the affected immunoglobulin domain in titin, I10, and speculated that this instability increases the vulnerability of titin to proteolysis *in situ* potentially leading to myocardial damage [3,13]. Here, we study the dysfunction caused by this mutation by implementing an integrative approach from the protein domain to the whole organism. Results indicate that, in addition to fold instability, the mSNP alters the chemical reactivity of titin leading to the formation of non-native interactions in the sarcomere. Even though the molecular and cellular effects of the exchange are mild, they appear to lead to a detectable alteration of the diastolic behaviour of the heart in KI mice. Thereby, our results establish that a single SNP in a common Ig component of titin can disturb the overall performance of the heart in the absence of other genetic factors. The result points to a high potential of titin mSNPs for causing cardiac disease.

2. Material and methods

T2850 is encoded by the ACC triplet on chromosome 2 : 178769890–178769893 referenced to GRCh38/hg38 human genome assembly. This is mutation T2896I in [3] and mutation T16I in [13].

For *in vitro* studies, I9–I11 (residues 2749-3009) and I10 (2835-2895) were expressed recombinantly in *E. coli* and purified to homogeneity by chromatography. For NMR studies, I10 was produced in M9 medium with $^{15}NH_4Cl$ as source of nitrogen and ^{13}C-glucose as carbon source.

I10 crystals were grown in 0.2 M $CaCl_2$, 0.1 M Tris pH 7.5 with either (a) 30% [w/v] PEG 3350, 3% [v/v] Isopropanol or (b) 25% [w/v] PEG 8000 as precipitants. Crystal (a) yielded a crystallographic model to 2.00 Å resolution with an $R_{factor}/R_{free} = 17.53/22.99\%$. Crystal (b) produced a model to 1.8 Å and $R_{factor}/R_{free} = 17.45/20.49\%$. I19–I11 crystals were produced in (c) 0.1 M Tris HCl pH 8.5, 30% [w/v] PEG 4000, 0.2 M $MgCl_2$; (d) 0.1 M Bis-Tris Propane pH 8.5, 20% [w/v] PEG 3350, 0.2 M sodium acetate. A three-dimensional model from (c) was to 1.9 Å with $R_{factor}/R_{free} = 19.28/22.83\%$; and from (d) to 1.53 Å with an $R_{factor}/R_{free} = 16.15/19.65\%$.

For molecular dynamic simulations, the crystal structure of I10–I11 was used as wild-type and the T2850I modelled *in silico*. A 'protein in a box of water' simulation of 50 ns followed standard protocols in GROMACS 5.0. Principal component analysis was used to analyse the trajectories for differences.

All animal experiments were approved by local ethics committees. T2850I mutation carrying mice were generated using gene-targeting. Genotype was confirmed by direct sequencing. The mice were on a mixed background, 50% 129S6 and 50% C57BL/6. Echocardiography was performed under isoflurane anaesthesia using a Vevo 2100 High-Resolution Imaging System and accompanying software.

GFP-tagged titin fragments were introduced into *tibialis anterior* muscle using electroporation under general anaesthesia. Mice were sacrificed 10 days later and transfected muscle sections visualized by laser scanning confocal microscopy for localization of titin fragments or differential centrifugation of muscle extracts was also performed. Tissue was fractionated into cytosolic, microsomal and particulate fractions and titin fragments were detected by Western blotting using anti-GFP antibody.

Neonatal mouse cardiomyocytes were isolated from 1 to 3-day-old pups. Cells were transfected with GFP-tagged titin fragments and fixed 48 h later. Cells were counterstained with antibodies against α-actinin, filamentous actin and DNA and imaged by confocal microscopy.

A comprehensive description of material and methods is provided in the electronic supplementary material.

3. Results

3.1. The T2850I mutation strongly destabilizes domain I10

To estimate the effect of the T2850I exchange on the fold stability of I10, the thermal denaturation of wild-type and

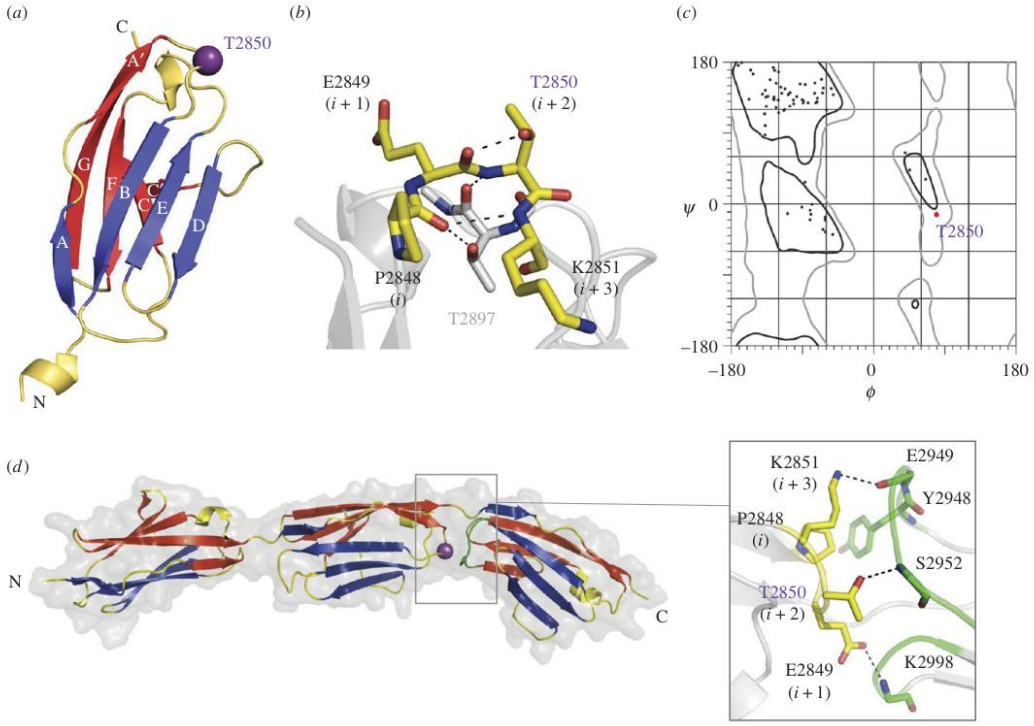


Figure 1. Crystal structure of I10 in isolation and in the context of the I9–I11 tandem. (*a*) Representation of $I10^{WT}$. The component β-sheets [A′FCC′G and ABED] are coloured in red and blue, respectively. Residue T2850 in the A′B β-turn is indicated with a purple sphere; (*b*) structural features of the A′B β-turn of $I10^{WT}$. Hydrogen bonds are indicated by dashed lines; (*c*) Ramachandran plot showing that the main chain conformation of T2850 is located in the generously allowed left-handed α-helical region (αL) (the electron density map of T2850 in electronic supplementary material, figure S2 shows that the conformation of this residue is well-defined experimentally) and (*d*) crystal structure of I9–I11 and detail of the I10(yellow)–I11(green) domain interface (inset).

mutated samples ($I10^{wt}$ and $I10^{T2850I}$, respectively) was monitored using differential scanning fluorimetry (DSF) (electronic supplementary material, Section S1). The melting temperature (T_m) of $I10^{wt}$ was derived from a single transition melting profile as 62.3 ± 1.3°C, in good agreement with its T_m value previously calculated from CD data (60.1 ± 0.2°C) [14]. This value indicates that I10 is an intrinsically stable domain. The T_m value of $I10^{T2850I}$ was 51.4 ± 1.8°C, approx. 11°C below that of $I10^{wt}$. Comparably acute ΔT_m decreases in titin Ig domains have only been observed when drastic truncations or modifications of the fold occurred. Accordingly, our measurement of a truncated version of I10 missing the N-terminal β-strand A showed a $\Delta T_m \approx -14$°C (electronic supplementary material, figure S1) and the deletion of just four residues in other Igs of titin decreased their T_m values by up to 15°C [15,16]. A similar ΔT_m decrease was observed in domain Z1 when the CD loop was replaced in its totality for an exogenous sequence that eliminated β-strand C′ and, thereby, removed the native capping of the Ig β-sandwich fold [17]. These data indicate that the impact of the single-point mutation T2850I on I10 is equivalent to that of a large insult on the fold, such as removal of β-strand A or the CD loop. This significant destabilizing effect suggests that T2850 plays an important structural role in I10.

3.2. The crystal structure of I10 reveals a high-energy conformation for residue T2850

To gain a molecular insight into the structural role of residue T2850, we elucidated the atomic structure of domain I10 at 1.74 Å resolution using X-ray crystallography (diffraction data statistics and model refinement parameters are given in the electronic supplementary material, table S2). We calculated a total of 17 molecular copies of I10 originating from two different space groups and diverse non-crystallographic symmetries. The structural features of all 17 molecular copies of I10 were in strict agreement (global RMSD = 0.62 Å for 82 matched Cα atoms across all models, calculated using MUSTANG [18]). The structures showed that I10 displays a classical Ig I-type fold (figure 1*a*) and that residue T2850 is located at the C-terminal pole of the fold, in position $i + 2$ of the β-turn connecting β-strands A′ and B (experimental electron density shown in the electronic supplementary material, figure S2). This is a β-turn type II, where residue $i + 2$ adopts a left-handed helix conformation (αL Ramachandran region) that is sterically restricted [19]. T2850 is stabilized in this conformation by two hydrogen bonds (figure 1*b*): (i) a short, strong bond (2.4 Å) between its side chain hydroxyl and the carbonyl of the previous residue, E2849 and (ii) a second contact between its main chain amide group and the main chain

carbonyl of T2897 in the neighbouring EF loop. The side chain hydrogen bond turns the otherwise atomic clash into a productive interaction and permits T2850 to adopt a high-energy conformation within the generously allowed αL Ramachandran region (figure 1*c*).

A side chain with a branched or cyclic Cβ atom (such as that of residues Ile, Val and Pro) in position $i+2$ could result in clashes with the carbonyl group of the preceding $i+1$ residue and, thereby, be sterically unsatisfactory. To investigate this deduction, we carried out a study of 62 641 unique β-turns type II naturally occurring in proteins using PDBeMotif [20] (electronic supplementary material, table S3). We observed that position $i+1$ is permissive to all residue types, but that position $i+2$ strongly favours glycine (approx. 78% occurrence) with other amino acid types present at lower, but significant frequencies. In agreement with our expectations, amino acids with a branched Cβ-atom are very rare in position $i+2$: isoleucine is present only in 2 out of 62 641 cases (3×10^{-5} occurrence), and valine in 22/62 641 (3.5×10^{-4}). Proline, with a cyclic Cβ-atom, is the most uncommon residue at 1/62 641 cases (1.5×10^{-5}). The few cases where proline or isoleucine residues were found in position $i+2$ were examined manually and found to correspond to exceptionally rare turns in internal core positions, stabilized by rich hydrophobic contacts. Deductions from natural residue occurrence were in full agreement with calculations of the differential free energy (ΔΔG) of residue tolerance in I10 using FoldX [21], which also proposed Ile, Pro and Val as poorly tolerated residues at this locus of the I10-fold (electronic supplementary material, figure S3). Our conclusion is further supported by previous studies that evaluated amino acid energetics and compositional potential in type-II β-turns using small sample populations [19,22]. We further observed that threonine residues at this position invariably establish a tight hydrogen bond with the preceding main chain carbonyl group. This hydrogen bond is not present in β-turns II containing the Thr-resembling residues Ser and Cys, indicating that this is a mechanism to specifically stabilize the branched Cβ-atom of Thr in that position.

3.3. T2850i does not induce structural changes in I10 but increases its internal flexibility

The effect of the T2850I exchange on I10 was studied using NMR. H^1-N^{15} HSQC spectra of both wild-type and mutated samples showed sharp, well-dispersed peaks characteristic of folded proteins, revealing that the overall fold is preserved in $I10^{T2850I}$ (figure 2*a*). However, a fraction of amide resonances was perturbed by the mutation. Chemical shift perturbations were quantified using a weighted average difference [23] and regarded as 'moderate' ($0.03 < \Delta\delta_{AV} < 0.15$ ppm) or 'large' ($\Delta\delta_{AV} > 0.15$ ppm). They showed that in the 91 residue-long $I10^{T2850I}$, 6 residues underwent large and 10 residues moderate chemical shift perturbations. When mapped onto the crystal structure of I10, the perturbations largely clustered around the mutation site, in the A′B β-turn and the neighbouring EF loop that are interconnected through hydrogen bonds mediated by T2850 (figure 2*a*). This shows that the mutation only introduces modest local distortions in the I10-fold. The characteristic backbone conformation of the A′B turn with a positive main chain dihedral ϕ angle in position $i+2$ is best quantified by the measurement of the indirect coupling constant $3J(C'_{k-1}H_k\alpha)$. In this experiment, we observed similar couplings throughout the backbone of wild-type and mutant samples and especially the large J-values characteristic for the positive ϕ angles of both Thr and Ile residues (figure 2*b*). This indicated that the protein backbone conformations of both wild-type and mutant are highly similar and that the exchange does not notably affect the protein structure.

Next, we examined whether the T2850I mutation alters the internal dynamics of the protein by measurement of ^{15}N relaxation data. The ratio of transverse (R2) and longitudinal (R1) relaxation rates gives a direct indication of the internal backbone dynamics of an amino acid. The average value of R2/R1 is determined by the overall rotational tumbling of the protein in solution and unaltered for most regions in the wild-type and mutant proteins (figure 2*c*). The outstanding exception are residues in the A′B β-turn and the preceding β-strand A: here we observed a significant increase in the R2/R1 ratios, which indicates that in the mutated protein these residues are involved in a slow (microsecond to milliseconds timescale) structural exchange process between a wild-type like ground state and one (or several) altered conformations. Derivation of quantitative dynamical parameters using the S2-order parameters, which is an indicator of the fast timescale (picosecond to nanosecond), reveals that backbone dynamics are very similar for wild-type and mutant Ig10. Thus, it can be concluded that the T2850I exchange increases internal domain flexibility not only locally but that it has a knock-on effect on the preceding secondary structure, loosening the N-terminal fraction of the I10-fold, shown above to be important for the stability of I10.

3.4. The crystal structure of I9–I11 shows that T2850 mediates interdomain contacts

Residue T2850 is located at the C-terminal loop region of the Ig fold. To study whether this region is involved in interactions with the next Ig domain packed serially along the chain and, thus, whether the T2850I exchange influences the Ig-tandem architecture of titin, we determined the crystal structure of I9–I11 (comprising I10 in its poly-domain context) to 1.53 Å resolution. Three molecular copies were obtained in two space groups that agreed in showing I9–I11 in an extended conformation (figure 1*d*). Such extended arrangements are common in titin Ig-tandems [24,25]. In I9–I11, the component domains follow a regular arrangement, where domains display relative torsion angles of 45°–68° and are connected by short, 2-residue sequences (TL and PI, respectively). In each Ig-pair, the domain interface is consistently formed by the A′B β-turn of the N-terminal domain slotting between the BC and FG loops of the C-terminal domain. Here, domains interact directly through a limited number of contacts (listed in the electronic supplementary material, Section S2). Both I9–I10 and I10–I11 interfaces consist of a central hydrophobic residue contributed by the linker sequence (TL; PI), flanked by two polar interactions between loop residues from the neighbouring domains (figure 1*d*, inset). In addition, electrostatic potential maps of single domains in the I9–I11 tandem revealed that the N- and C-terminal poles of all domains are, respectively, positively and negatively charged. This suggests that an electrostatic component further assists the organization of these Ig domains within the titin chain.

Residue T2850 is buried within the I10–I11 interface (burial fraction of side chain is 0.79 as calculated with FoldX [21]). Here, T2850 preserves its intradomain hydrogen bonds (to T2897 and

RSOB160114—12/9/16—20:30–Copy Edited by: Not Mentioned

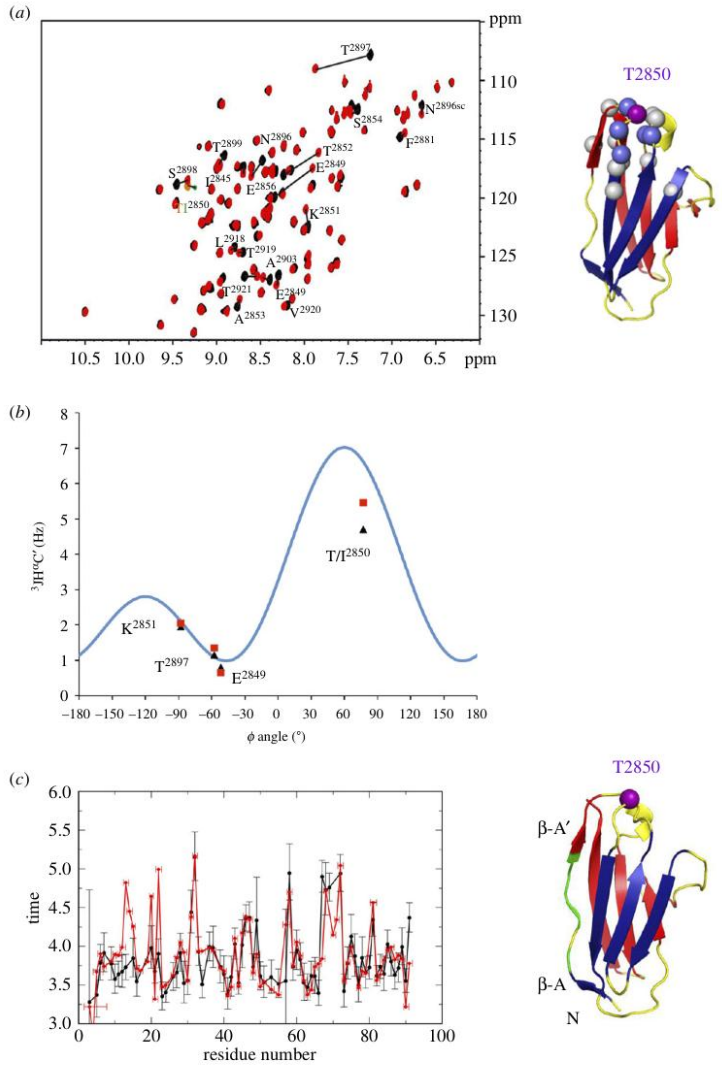


Figure 2. NMR analysis of changes in I10 induced by the T2850I exchange. (*a*) H^1-N^{15} HSQC spectra (left). Residues affected by notable differences are indicated and mapped on the crystal structure of I10 (right). It can be seen that significant changes only occur in the immediate vicinity of the T2850I exchange; (*b*) Measured 3 $J(C'_{k-1}H\alpha_k)$ values plotted against the ϕ_k dihedral angle observed in the crystal structure of $I10^{WT}$. The blue curve shows the dependency of the coupling constant upon the torsion angle ϕ. Both wild-type and mutant show large J-couplings for residue 2850 characteristic for positive ϕ angles; (*c*) Ratio of transverse (R2) and longitudinal (R1) ^{15}N relaxation rates (left). Values above the average can be mapped to the A′B β-turn and β-strand A (right; green). Here and throughout this figure, values corresponding to $I10^{WT}$ samples are in black and $I10^{T2850I}$ in red.

E2849) observed in the isolated I10 structure and, in addition, hydrogen bonds with S2952 from the BC loop of I11. This suggests that the exchange of the T2850 residue will cause a local alteration of the conformational dynamics of the titin chain.

3.5. The T2850I exchange alters the relative orientation of domains in the chain

As residue T2850 is involved in both the intra- and interdomain organization of titin, we studied next the effect of the T2850I exchange on the conformational dynamics of the I10–I11 tandem using molecular dynamics simulations (MDS) (electronic supplementary material, Section S4). In total, 50 ns simulations were performed on the wild-type and on two T2850I models that corresponded to two different rotamers of the isoleucine residue. In addition, one wild-type and one T2850I variant were repeated with a different water model. The simulations were continued until no new conformational space was sampled as reflected by principal component analysis (electronic supplementary material, figure S4*b–d*). As a control that the time of the simulation was not limiting possible

interdomain motions of larger amplitude, we performed a simulation on the Ig pair I65–I66 (extracted from PDB entry 3B4B [26]). In I65–I66, Ig domains are linked by a hydrophilic linker, three residues-long that allows unrestricted motions. This simulation confirmed that 50 ns simulation allows the sampling of multiple extreme conformations (electronic supplementary material, figure S4*a*).

Calculated trajectories for all I10–I11 models were converted to backbone models and concatenated, then subjected to covariance analysis. This showed that interdomain motions could be largely described by the first three principal component eigenvectors (PC1, PC2 and PC3), with PC1 and PC2 describing the largest movements. PC1 describes the rotational movement of I10 with respect to I11 (which totals a range of 68°), while PC2 describes the bending opening between I10 and I11 (spanning a range of 65°). When PC1 and PC2 values are plotted for all the trajectories, a difference in the conformational space sampled by wild-type and T2850I mutant is revealed—in particular, for the dominant component PC1 (figure 3*a*,*b*). Wild-type trajectories have more positive eigenvalues; in other words, inter-domain positions are aligned straight along an imaginary axis connecting the two domains centers of mass, with both domains facing a similar starting direction. Conversely, the trajectories for T2850I mutants show negative eigenvalues that describe a greater degree of twisting, where the I10 and I11 modules face different directions. In combination with other components of movement, such as PC2's bending, this results in the T2850I mutants favouring a more bent and twisted conformation, that contrasts with the more extended arrangement preferred by the wild-type (figure 3*b*). This result suggests that wild-type and mutant favour somewhat different areas of conformational space, which might lead to titin acquiring different chemical surface properties at the I10–I11 locus upon mutation.

3.6. The T2850I exchange affects the myocellular reactivity of the titin Ig-tandem *in vivo*

To test whether the altered conformation of the T2850I-containing titin chain leads to an altered function *in vivo*, we assayed a GFP-tagged I7–I13 titin fragment in its wild-type and mutated versions in skeletal muscle of living mice. The fragment was introduced into the *tibialis anterior* muscle by electroporation and its expression monitored by Western Blot of muscle extracts using an anti-GFP antibody (electronic supplementary material, Section S5). Single sharp protein bands corresponding to the expected molecular weight of I7–I13 were revealed this way. Signs of differential degradation of wild-type and mutated forms were not evident. This did not agree with a previous *in vitro* study, where recombinant I7–I13 and I7–I13^{T2850I} samples incubated with heart extracts indicated a notably reduced half-life of the mutated variant, probably as a result of proteolysis [13]. The findings in the current study suggest that the proteases that acted on the recombinant samples when extracts were used might not natively access cytoplasmic samples in the myofibril (e.g. due to compartmentalization) or, alternatively, that the proteolytic components of skeletal muscle are not representative of those from cardiac tissue. Either way, the results did not bring further support to *in situ* titin proteolysis as a mechanism of disease.

Imaging of the transfected muscles using *in vivo* confocal light microscopy revealed a different myocellular localization of wild-type and mutant I7–I13 (figure 3*c*). Wild-type fragments remained mostly soluble in the cytoplasm, associating only weakly with the sarcomere as suggested by the weak striated pattern. Fixed sections of the GFP-I7–I13 transgenic muscle that had been further stained with a fluorescent phalloidin conjugate (which binds F-actin forming the thin filaments), revealed that wild-type I7–I13 associated with the sarcomere in areas lacking F-actin, i.e. the H-zone of the A-band. As the native location of I7–I13 within the titin chain is the I-band, we concluded that the weak A-band patterning resulted from unspecific interactions due to the elevated expression levels. This result is not unexpected as previous studies with recombinant titin fragments did not detect binding of titin Ig-tandems to actin or other sarcomeric components [28,29]. By contrast, I7–I13^{T2850I} formed a well-defined striated pattern. Phalloidin staining showed that I7–I13^{T2850I} bands colocalized with F-actin, at the I-band. However, subcellular fractionation and actin co-sedimentation experiments showed that the binding was not to actin itself or other primary sarcomeric components (electronic supplementary material, Section S5). Unfortunately, efforts to identify the interactor using pull-downs in muscle extracts and yeast two-hybrid screens were not successful. Contrary to the transgenic mouse muscle experiments, the transfection of mouse neonatal cardiomyocytes with I7–I13 and I7–I13^{T2850I} did not reveal localization differences (figure 4). Both fragments remained diffused in the cytoplasm under basal conditions and even after isoproterenol stimulation (electronic supplementary material, Section S6). Neonatal cardiomyocytes have fully developed sarcomeres but lack T-tubules, which make us speculate that a T-tubule component might be the putative target of the pathological interaction. Nonetheless, as the I7–I13^{T2850I} sarcomeric pattern agrees with the expected native position of I7–I13, this pathologically increased affinity for an I-band component is likely to be physiologically relevant for this mSNP in its natural context within full-length titin.

3.7. Transgenic mice carrying the T2850I exchange present enhanced diastolic stiffness

To determine whether the mutation has functional effects at the organ level, we generated T2850I KI mice and performed echocardiography. No differences were found in left ventricular chamber dimensions during diastole and systole, ejection fraction or stroke volume (table 1). Pulse-wave Doppler echocardiography was used to measure the velocity of diastolic filling at the level of the mitral valve. Filling is known to occur in two waves: early diastolic filling (E-wave) and late diastolic filling due to atrial contraction (A-wave). The E/A ratio is the ratio of the early (E) to late (A) ventricular filling velocities and reflects diastolic function. The E/A ratio is used as an index for diastolic heart failure. Depending on the genetic background of the mouse strain, the E/A ratio in mice is 1.25–1.6. For example, in FVB mice after a two-week long transverse aortic constriction, which is a severe insult on the heart, the E/A ratio increases from 1.26 to 2.5 [30]. In T2850I KI mice, the E/A ratio was increased (from 1.67 in wild-type to 2.19 in the mutant) suggesting diastolic dysfunction. Furthermore, the E-wave deceleration time (DT) varies inversely with LV diastolic stiffness and is another measure for the myocardial stiffening [31]. A significant E-wave DT reduction was found in T2850I KI mice, further supporting an increase in diastolic chamber stiffness. In conclusion, T2850I KI mice have a diastolic dysfunction.

RSOB160114—12/9/16—20:30–Copy Edited by: Not Mentioned

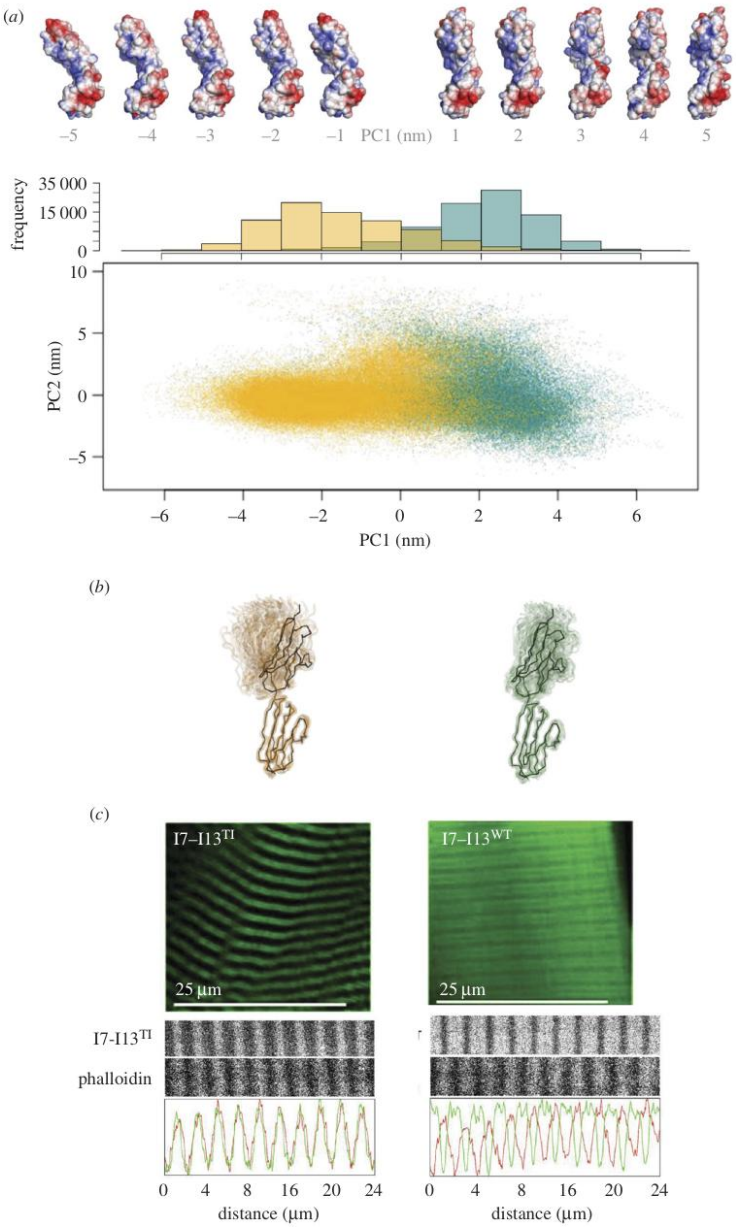


Figure 3. Different dynamic properties of wild-type and T2850I-mutated structures. (*a*) Molecular Dynamics Simulations of I10–I11. Top: electrostatic surface representation of conformational classes (using APBS [27]). Bottom: PC1 values plotted against PC2 values, showing static shift of relative conformations between trajectories with a histogram of PC1 eigenvalue frequency; (*b*) cloud cartoon representation of trajectory snapshots at 0.5 ps intervals. The initial crystal structure is shown as black ribbon. Throughout (*a,b*), T2850I is in gold and wt in green; (*c*) Differential localization *in vivo* of wild-type and mutant GFP-I7–I13 samples. Wild-type samples remained mostly diffused in the cytoplasm, with a weak association with the sarcomeric A-band as indicated by double-labelling with Phalloidin. T2850I samples strongly interacted with the sarcomeric I-band co-localizing with the phalloidin stain.

4. Discussion

The genetic screening of the *TTN* gene, now pursued in large patient populations, has the potential to assist the diagnosis of cardiomyopathies, assess prognosis and guide therapy. However, identifying disease-causing missense alleles in titin is challenging, with the unexpectedly large number of rare genetic variants making their scoring a staggering task [32,33].

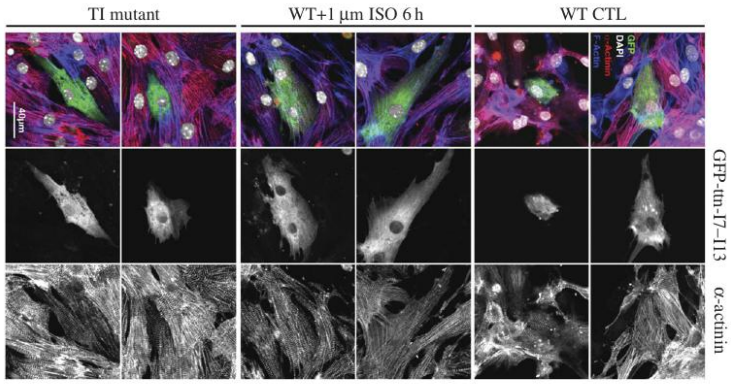


Figure 4. Expression of I7–I13 samples in neonatal cardiomyocytes. Z-discs were stained with α-actinin (right panel, red in the overlay), filamentous actin with fluorescent phalloidin (blue in the overlay), transfected titin fragments with EGFP (middle panel, green in the overlay) and nuclei with DAPI (white in the overlay). Both, wild-type (WT CTL) and mutant fragments of titin (TI) remain mainly diffuse in the cytoplasm of neonatal cardiomyocytes. No striated pattern compatible with sarcomeric targeting was observed, even in cells transfected with the wild-type construct that were challenged with 1 µM isoprotereonol (ISO) for 6 h. Shown are two representative cells for each condition. Scale bar, 40 µm.

Table 1. Cardiac parameters in T2850I transgenic mice. LVIDd: left ventricular internal diastolic diameter; WTd: diastolic wall thickness (average of posterior and anterior walls); LVIDs; left ventricular internal systolic diameter; WTs: systolic wall thickness(average of posterior and anterior walls); LV Vol, d: left ventricular diastolic volume; LV Vol, s: left ventricular systolic volume; LVW: left ventricular weight (mg); EF: ejection fraction; SV: stroke volume; MV E, mitral valve early diastolic peak filling velocity; MV A, mitral valve late diastolic peak filling velocity; MV Decel: deceleration time of E-wave (ms); MV E/A: ratio of MV E : MV A; LA: left atrium. *p*-value: significance value calculated with *t*-test; $*p < 0.05$. Q4

	WT ($n = 7$)	T2850I ($n = 7$)	*p*-value
LVIDd (mm)	4.72 ± 0.14	4.78 ± 0.14	0.79
WTd (mm)	0.82 ± 0.01	0.84 ± 0.02	0.42
LVIDs (mm)	3.41 ± 0.14	3.54 ± 0.16	0.55
WTs (mm)	1.20 ± 0.04	1.21 ± 0.02	0.84
LV Vol;d (µl)	103.1 ± 6.6	106.8 ± 7.2	0.72
LV Vol;s (µl)	48.4 ± 4.5	54.1 ± 5.7	0.44
LVW	158 ± 6	172 ± 9	0.21
EF (%)	53.3 ± 2.9	49.9 ± 2.0	0.35
SV (µl)	54.8 ± 4.2	52.6 ± 2.6	0.67
MV E (mm s^{-1})	581 ± 32	647 ± 53	0.31
MV A (mm s^{-1})	351 ± 22	300 ± 26	0.15
MV Decel	30.3 ± 0.8	**26.5 ± 1.1***	**0.017**
MV E/A	1.67 ± 0.08	**2.19 ± 0.15***	**0.011**
E/E′	35.7 ± 2.9	35.6 ± 3.1	0.99
LA (mm)	3.07 ± 0.18	3.26 ± 0.24	0.54

Rare sarcomeric gene variations associated with cardiomyopathy in small patient families were found in 17% of the NHLBI GO Exome Sequencing Project population [34] indicating that association studies alone are insufficient to evaluate the pathological potential of rare variants. In addition, rare genetic variants of sarcomeric genes often have complicated penetrance patterns. It was shown that the risk of adverse cardiovascular events grows with increasing numbers of rare sarcomeric variants [35]. Thus, methods are now urgently needed that can improve mSNP classification and unleash the wealth of information in *TTN* databases.

To gain an insight into the damage potential of mSNPs in titin—in particular, in the heart that appears to be more susceptible to mutations than skeletal muscle—we have studied the exchange T2850I in domain I10 of titin's I-band, which has been proposed to be linked to ARVC [3]. In this work, we went beyond our initial clinical and biophysical characterization of the mutation [3,13] and explored manifestations of this mSNP at different levels of biological complexity, ranging from molecular to organismal levels. In this work, atomic resolution three-dimensional structures of the affected titin region were generated as well as, to our knowledge, the first mouse model of a cardiomyopathy-associated titin mSNP. Our results show that the T2850I exchange only has a subtle impact on the titin molecule as well as the tissue and animal levels. This was not anticipated, as it is intuitively expected that the severity of the disease promoted by an mSNP will correlate with the extent of the molecular damage that it causes. Structural data (HSQC and 3 $J(C'_{k-1}H_k\alpha)$) demonstrate that the T2850I exchange is tolerated by the I10-fold, not causing any detectable structural aberration. However, DSF data show that the mutated domain is notably less stable than the wild-type. The large difference in measured thermal stability was difficult to understand, as the exchange involves the removal of a single hydrogen bond (established by the Thr side chain with the main chain of the preceding residue; T2850-OH:CO-E2849) in the otherwise-intact fold. However, NMR relaxation data showed that the exchange not only increased the flexibility of the affected loop but also of the preceding β-strand A. This N-terminal segment of the fold is known to be of key importance for the chemical [26], thermal [15,16] and mechanical stability of the fold [36]. Consequently, our own measurement of an I10 variant lacking the N-terminal β-strand A showed a thermal destabilization comparable to

that of the T2850I point mutant (electronic supplementary material, Section S1). That the mutated Ig is also mechanically weaker has been established previously by AFM [13], so that its unfolding is likely to occur at physiological muscle tension. The impact of an unfolded Ig on the overall elasticity of the titin chain seems negligible. However, it was originally proposed [13] that Ig unfolding (resulting from either stretch or spontaneous unfolding due to the intrinsic weakness of the mutated domain) might lead to *in situ* proteolysis of the titin chain, even though depletion of mutated titin from sarcomeres was not tested. In the current study, degradation of I7–I13 tandems expressed in muscle was not apparent, suggesting that I10 unfolding might have other functional consequences. A recent study has shown that induced unfolding of titin Ig segments in cardiomyocytes results in elevated stiffness caused by the aggregation of titin chains [37]. Yet, in diseased human muscle and heart, HSP27 or αβ-crystallin associate with the unfolded modules, preventing aggregation and suppressing the stiffening [37]. Hence, the extent to which I10 unfolding, specifically, contributes to the T2850I pathology is unclear.

The second consequence of the mutation is at the chain level, effected by its location in a domain junction. Our previous characterization of titin segments showed that the properties of domain interfaces are finely tuned along the chain, exhibiting conserved features in different parts of the sarcomere [24,25]. Not surprisingly, MDS suggests that the T2850I exchange influences the conformational dynamics of the titin chain at that locus, possibly altering its surface chemistry. An important deduction from this work is that mutations located at domain interfaces in titin can have an unexpected pathological potential, even when not causing major structural damage to the component domains. This finding is in agreement with our transgenic muscle data that point to the development of pathological binding capacity in the mutated titin's I-band *in situ*. Interestingly, the differential sarcomeric localization of wild-type and mutated samples could not be reproduced in neonatal cardiomyocytes and further efforts to identify the possible interactor of mutated titin failed. Upon differential centrifugation of transgenic muscles extracts, titin fragments were solely found in cytoplasmic and not in particulate myofibrillar or microsomal fractions, implying a fragile nature of the pathological interaction. Speculatively, we consider the possibility that the interaction might involve components of T-tubules as these are absent in neonatal cardiomyocytes.

Interspecies conservation of the affected residue allowed us to generate a knock-in mouse model carrying the titin T2850I mSNP. Echocardiographic assessment revealed diastolic dysfunction in this mouse model, confirming that a single-point mutation in titin can have a measurable effect in whole-heart mechanics in the absence of any other genomic differences. However, no degenerative alterations in the left or right ventricles were observed—thereby, not reproducing the human disease phenotype overall. Yet, it is possible that the fibrofatty formations typically manifested at a clinical level in the second or the third decade of life in ARVC patients are not appropriately mimicked in mice. In fact, this appears to be also the case in murine models of other non-titin-based ARVC types (reviewed in [38]). Thus, there is a need for better biological systems to study ARVC. In this respect, new technologies such as iPSC(induced pluripotent stem cells)-derived cardiomyocytes need to be considered. This system has been applied to the study of cardiomyopathies caused by titin truncating mutations [39] and perhaps might be applicable to the study of titin mSNPs. In ARVC, diastolic abnormalities are an early marker of the disease [40], thus the mice model might mimic initial stages of the human pathology. Speculatively, 'stickiness' of Ig-tandems in titin's I-band [28,29] (which natively makes little or no connections to the thin filament during the cardiac cycle) would interfere with its compliance, necessary for efficient diastolic filling [27,41]. This could explain the observed phenotype. Finally, there is the possibility that the effect of the T2850I mutation in patients is potentiated by another co-segregating genetic factor, jointly leading to the disease phenotype (e.g. an additive effect of other unknown rare sarcomeric variants). In this regard, more complete genomic analyses will be required in the future.

In conclusion, our results reveal that the contribution of a single mSNP in a generic domain of titin can lead to a measurable alteration of heart function, even though the mSNP might only cause subtle changes at the protein level. Given the staggering number of mSNPsin titin, this finding establishes the high potential of this protein to contribute to disease mechanisms in the heart.

Ethics. All procedures relating to KI-mice were performed in compliance with institutional guidelines of the University of Arizona and had IACUC approval. Experimental protocols for transgenic muscle experiments were approved by the national German authorities and all animals were kept and treated according to guidelines of the EU Directive 2010/63/EU.

Authors' contributions. J.B., J.F., B.R.A., R.W., B.F., St.L., B.S., M.M.K., R.R., B.B., D.J.R. and O.M. performed experiments and data analysis. J.B., J.F., B.S., H.G., St.L. and O.M. wrote the manuscript. J.B., H.G., St.L. and O.M. conceived the study.

Competing interests. We declare we have no competing interests.

Funding. We acknowledge the financial support of the British Heart Foundation (PG/13/21/3007 to O.M. and D.J.R.), the Leducq Foundation (TNE-13CVD04 to H.G., Si.L. and O.M.), the Hector Foundation (to Si.L.), the Biotechnology and Biological Sciences Research Council (BB/M00676X/1 to R.W., O.M. and D.J.R.), EU-MCSA-IRSES (SarcoSi to Si.L. and O.M.) and NIH grant nos. HL107744, HL128457, HL062881 and HL118524 (H.G. and St.L.).

References

1. Chauveau C, Rowell J, Ferreiro A. 2014 A rising titan: TTN review and mutation update. *Hum. Mutat.* **35**, 1046–1059. (doi:10.1002/humu.22611)
2. Herman DS *et al.* 2012 Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **366**, 619–628. (doi:10.1056/NEJMoa1110186)
3. Taylor M *et al.* 2011 Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation* **124**, 876–885. (doi:10.1161/CIRCULATIONAHA)
4. Itoh-Satoh M *et al.* 2002 Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochem Biophys. Res. Commun.* **291**, 385–393. (doi:10.1006/bbrc.2002.6448)
5. Gerull B *et al.* 2002 Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat. Genet.* **30**, 201–204. (doi:10.1038/ng815)
6. Satoh M, Takahashi M, Sakamoto T, Hiroe M, Marumo F, Kimura A. 1999 Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochem. Biophys. Res. Commun.* **262**, 411–417. (doi:10.1006/bbrc.1999.1221)

RSOB160114—12/9/16—20:34–Copy Edited by: Not Mentioned

7. Matsumoto Y *et al.* 2006 Functional analysis of titin/connectin N2-B mutations found in cardiomyopathy. *J. Muscle Res. Cell Motil.* **26**, 367–374. (doi:10.1007/s10974-005-9018-5)
8. Peled Y *et al.* 2014 Titin mutation in familial restrictive cardiomyopathy. *Int. J. Cardiol.* **171**, 24–30. (doi:10.1016/j.ijcard.2013.11.037)
9. Mestroni L, Taylor MR. 2013 Genetics and genetic testing of dilated cardiomyopathy: a new perspective. *Discov. Med.* **15**, 43–49.
10. Roberts AM *et al.* 2015 Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* **7**, 270ra6. (doi:10.1126/scitranslmed.3010134)
11. Wu B, Eggert J, Alexov E. 2014 Molecular mechanisms underlying pathogenic missense mutations. In *eLS*. Chichester, UK: John Wiley & Sons Ltd.
12. Sen-Chowdhry S, Syrris P, McKenna WJ. 2007 Role of genetic analysis in the management of patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy. *J. Am. Coll. Cardiol.* **50**, 1813–1821. (doi:10.1016/j.jacc.2007.08.008)
13. Anderson BR, Bogomolovas J, Labeit S, Granzier H. 2013 Single molecule force spectroscopy on titin implicates immunoglobulin domain stability as a cardiac disease mechanism. *J. Biol. Chem.* **288**, 5303–5315. (doi:10.1074/jbc.M112.401372)
14. Politou AS, Thomas DJ, Pastore A. 1995 The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys. J.* **69**, 2601–2610. (doi:10.1016/S0006-3495(95)80131-1)
15. Politou AS, Gautel M, Joseph C, Pastore A. 1994 Immunoglobulin-type domains of titin are stabilized by amino-terminal extension. *FEBS Lett.* **352**, 27–31. (doi:10.1016/0014-5385793(94)00911-2)
16. Pfuhl M, Improta S, Politou AS, Pastore A. 1997 When a module is also a domain: the role of the N terminus in the stability and the dynamics of immunoglobulin domains from titin. *J. Mol. Biol.* **265**, 242–256. (doi:10.1006/jmbi.1996.0725)
17. Bruning M, Barsukov I, Franke B, Barbieri S, Volk M, Leopoldseder S, Ucurum Z, Mayans O. 2012 The intracellular Ig fold: a robust protein scaffold for the engineering of molecular recognition. *Protein Eng. Des. Sel.* **25**, 205–212. (doi:10.1093/protein/gzs007)
18. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. 2006 MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559–574. (doi:10.1002/prot.20921)
19. Hutchinson EG, Thornton JM. 1994 A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**, 2207–2216. (doi:10.1002/pro.5560031206)
20. Golovin A, Henrick K. 2008 MSDmotif: exploring protein sites and motifs. *BMC* Bioinform. **9**, 312. (doi:10.1186/1471-2105-9-312)
21. Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. 2011 A graphical interface for the FoldX forcefield. *Bioinformatics* **27**, 1711–1712. (doi:10.1093/bioinformatics/btr254)
22. Madan B, Seo SY, Lee SG. 2014 Structural and sequence features of two residue turns in beta-hairpins. *Proteins* **82**, 1721–1733. (doi:10.1002/prot.24526)
23. Nordstrand K, Sandström A, Aslund F, Holmgren A, Otting G, Berndt KD. 2000 NMR structure of oxidized glutaredoxin 3 from *Escherichia coli*. *J. Mol. Biol.* **303**, 423–432. (doi:10.1006/jmbi.2000.4145)
24. Marino M, Svergun DI, Kreplak L, Konarev PV, Maco B, Labeit D, Mayans O. 2006 Poly-Ig tandems from I-band titin share extended domain arrangements irrespective of the distinct features of their modular constituents. *J. Muscle Res. Cell Motil.* **26**, 355–365. (doi:10.1007/s10974-005-9017-6)
25. Zacharchenko T, von Castelmur E, Rigden DJ, Mayans O. 2015 Structural advances on titin: towards an atomic understanding of multi-domain functions in myofilament mechanics and scaffolding. *Biochem. Soc. Trans.* **43**, 850–855. (doi:10.1042/BST20150084)
26. von Castelmur E *et al.* 2008 A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. *Proc. Natl Acad. Sci. USA* **105**, 1186–1191. (doi:10.1073/pnas.0707163105)
27. Linke WA, Rudy DE, Centner T, Gautel M, Witt C, Labeit S, Gregorio CC. 1999 I-band titin in cardiac muscle is a three-element molecular spring and is critical for maintaining thin filament structure. *J. Cell Biol.* **146**, 631–644. (doi:10.1083/jcb.146.3.631)
28. Kulke M, Fujita-Becker S, Rostkova E, Neagoe C, Labeit D, Manstein DJ, Gautel M, Linke WA. 2001 Interaction between PEVK-titin and actin filaments: origin of a viscous force component in cardiac myofibrils. *Circ. Res.* **89**, 874–881. (doi:10.1161/hh2201.099453)
29. Yamasaki R *et al.* 2001 Titin-actin interaction in mouse myocardium: passive tension modulation and its regulation by calcium/S100A1. *Biophys. J.* **81**, 2297–2313. (doi:10.1016/S0006-3495(01)75876-6)
30. Gao S, Ho D, Vatner DE, Vatner SF. 2011 Echocardiography in mice. *Curr. Protoc. Mouse Biol.* **1**, 71–83. (doi:10.1002/9780470942390.mo100130)
31. Little WC, Ohno M, Kitzman DW, Thomas JD, Cheng CP. 1995 Determination of left ventricular chamber stiffness from the time for deceleration of early left ventricular filling. *Circulation* **92**, 1933–1939. (doi:10.1161/01.CIR.92.7.1933)
32. Nelson MR *et al.* 2012 An abundance of rare functional variants in 202 drug target genes sequenced in 14 002 people. *Science* **337**, 100–104. (doi:10.1126/science.1217876)
33. Keinan A, Clark AG. 2012 Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740–743. (doi:10.1126/science.1217283)
34. Norton N, Robertson PD, Rieder MJ, Züchner S, Rampersaud E, Martin E, Li D, Nickerson DA, Hershberger RE. 2012 Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ. Cardiovasc. Gene.* **5**, 167–174. (doi:10.1161/CIRCGENETICS.111.961805)
35. Bick AG *et al.* 2012 Burden of rare sarcomere gene variants in the Framingham and Jackson heart study cohorts. *Am. J. Hum. Genet.* **91**, 513–519. (doi:10.1016/j.ajhg.2012.07.017)
36. Israelewitz B, Gao M, Schulten K. 2001 Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224–230. (doi:10.1016/S0959-440X(00)00194-9)
37. Kötter S, Unger A, Hamdani N, Lang P, Vorgerd M, Nagel-Steger L, Linke WA. 2014 Human myocytes are protected from titin aggregation-induced stiffening by small heat shock proteins. *J. Cell Biol.* **204**, 187–202. (doi:10.1083/jcb.201306077)
38. Lodder EM, Rizzo S. 2012 Mouse models in arrhythmogenic right ventricular cardiomyopathy. *Front. Physiol.* **3**, 221. (doi:10.3389/fphys.2012.00221)
39. Hinson JT *et al.* 2015 HEART DISEASE. Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science* **349**, 982–986. (doi:10.1126/science.aaa5458)
40. Tandri H, Bomma C, Calkins H, Bluemke DA. 2004 Magnetic resonance and computer tomography imaging of arrhythmogenic right ventricular dysplasia. *J. Magn. Reson. Imaging* **19**, 848–858. (doi:10.1002/jmri.20078)
41. Granzier HL *et al.* 2014 Deleting titin's I-band/A-band junction reveals critical roles for titin in biomechanical sensing and cardiac function. *Proc. Natl Acad. Sci. USA* **111**, 14 589–14 594. (doi:10.1073/pnas.1411493111)
42. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001 Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10 037–10 041. (doi:10.1073/pnas.181342398)

RSOB160114—12/9/16—20:34–Copy Edited by: Not Mentioned

7.1.2 "Catalytic Activity of the Kinase Domain of Twitchin Inhibits Muscle Activity"

Yohei Matsunaga[1], Hyundoo Hwang[2], Barbara Franke[3], Rhys Williams[3,4], McKenna Penley[5], Hiroshi Qadota[1], Levi T. Morran[4], Hang Lu[2], Olga Mayans[3], and Guy M. Benian[1]

[1]Department of Pathology, Emory University, Atlanta, GA 30322 USA; [2]School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA; [3]Department of Biology, University of Konstanz, 78457 Konstanz, Germany; [4]Institute of Integrative Biology, University of Liverpool, L69 7ZB Liverpool; [5]Department of Biology, Emory University, Atlanta, GA 30322 USA

corresponding author: Guy M. Benian, pathgb@emory.edu

Manuscript in preparation

Personal contribution to this manuscript includes: recombinant expression and chromatographic purification of TwcK Kin^{K185A} mutant, consisting of the TwcK kinase domain containing a lysine to alanine mutation in the conserved AxK motif of the β3 strand. Phosphotransfer assays showing that TwcK Kin^{K185A} is catalytically inactive. Recombinant expression and chromatographic purification of TwcKR and TwcKR^{K185A} consisting of Ig-Ig-FnIII-NL-Kin-CRD-Ig domains and preparation of TwcKR and TwcKR^{K185A} samples for small angle x-ray scattering (SAXS).

Contributions can be seen in the following sections: Results and discussion; Recombinant protein production; *in vitro* phosphorylation assay.

ABSTRACT

Sarcomeres, the fundamental units of muscle contraction, contain giant polypeptides (>700,000 Da) of the titin-like family, composed primarily of multiple copies of immunoglobulin (Ig) and fibronectin type 3 (Fn3) domains and one or two protein kinase domains at their C-termini. Members of this family include titin in mammals, twitchin in nematodes and in molluscs, and projectin in insects. The physiological substrates and functions of the kinase domains in these proteins are largely unknown and whether some of these kinase domains are catalytically active in vivo, in particular vertebrate titin kinase, is controversial. Adding to this, a number of scaffolding interactions have now been reported for titin-like kinases, which has led to question whether it is scaffolding and not catalysis that is of functional relevance in this kinases. To bring light into this debate, we explore for the first time the catalytic significance of a titin-like kinase, twitchin kinase, in the muscle sarcomere in vivo. For this, we converted a highly conserved lysine (K) involved in ATP binding to alanine (A) and showed that this abolished the kinase activity of recombinant twitchin kinase in vitro. SAXS analysis indicated that this residue replacement did not alter the structural integrity of the kinase. CRISPR/Cas9 gene editing was used to create the same mutation in the endogenous unc-22 gene, resulting in expression of kinase-dead twitchin as single source in muscle. This mutant, unc-22(KtoA), undergoes normal development, does not twitch, has regular sarcomeric organisation and preserved vital parameters, including life span and brood size. However, both in swimming and crawling, unc-22(KtoA) moves faster than wild type, and has faster rates of contraction and relaxation. This suggests that the catalytic activity of twitchin kinase inhibits muscle contraction and, therefore, that twitchin kinase cellular pathway ultimately regulates the motor proteins of the

contractile sarcomere. Yet, paradoxically, wild type individuals exhibited greater competitive fitness than unc-22(KtoA) mutants. Thus, we conclude that twitchin kinase's catalysis is in vivo relevant in mature active muscle and that, although not being required for organism viability, is essential for competitive survival.

RESULTS AND DISCUSSION

We wondered what the phenotype would be of a nematode in which twitchin's kinase domain is intact but catalytically inactive. A highly conserved lysine (AxK) in the small lobe of protein kinases coordinates ATP and helps transfer the γ-phosphate. Mutation of this lysine into alanine or several other amino acids inactivates many known kinases (Iyer *et al*. 2005). The catalytic core of the protein kinase domain of twitchin is active in *in vitro* kinase assays against a model peptide substrate (Lei *et al*. 1994; von Castelmur *et al*. 2012). We purified bacterially expressed recombinant twitchin kinase from wild type and a mutant in which this lysine was converted to an alanine (KtoA). As shown in Figure 1A, in an in vitro kinase assay, twitchin kinase (KtoA) is nearly devoid of phosphotransferase activity; it displays ~0.1% of wild type activity.

Conversion of the catalytic lysine to alanine is not expected to influence the overall folding or shape of a kinase domain. Nevertheless, we asked whether the KtoA mutation in twitchin kinase would have an effect on the structure of the kinase domain. The expression, yield and solubility of recombinant twitchin kinase (KtoA) was nearly identical to that of twitchin kinase wild type. We studied the overall molecular features of TwcK in its wild type and KtoA mutated forms using SAXS. The experimental scattering curves from these samples (**Figure 1B**) as well as their

corresponding R_g and D_{max} values (**Figure 1C**) are in agreement within the error of this technique. This suggests that both TwcK forms share a same structural conformation and that the introduction of the KtoA mutation does not affect the structure of TwcK detectably.

We next used CRISPR/Cas9 gene editing to create the same KtoA mutation in the endogenous *unc-22* (twitchin) gene (Figure S1). We outcrossed the mutant three times to wild type to remove most of the possible off-target mutations generated by CRISPR. We also sequenced the *unc-22* gene from this *unc-22(KtoA)* strain and verified that no other mutations were present in the *unc-22* gene. We characterized the *unc-22(KtoA)* phenotype as follows. In contrast to all previously characterized *unc-22* mutant alleles (Moerman and Baillie, 1979; Moerman *et al*., 1988; Matsunaga *et al*. 2015), *unc-22(KtoA)* does not twitch and shows a wild type response to nicotine (Figure S2). Moreover, by western blot, *unc-22(KtoA)* expresses normal levels of various twitchin isoforms of the appropriate size (Figure S3). The locomotion of nematodes arises from the alternating contraction and relaxation of body wall striated muscles on its dorsal and ventral sides. By immunostaining with a battery of antibodies to known sarcomeric proteins, *unc-22(KtoA)* displays normal sarcomere structure in its body wall muscle cells (Figure 2). This assessment includes the localization of twitchin, components of thick filaments (MYO-3, UNC-15), M-lines (UNC-89 and PAT-6), and dense bodies (Z-line analogs; ATN-1 and PAT-6). This is in contrast to the loss of function *unc-22(e66)* and null *unc-22(ct37)* alleles which show disorganized sarcomeres, but is similar to the unusual missense *unc-22(e105)* allele (Figure 2). Thus, *unc-22(KtoA)* expresses full-length twitchin isoforms that are stable and normally incorporated into sarcomeres, and loss of

twitchin kinase activity has no effect on the organization (assembly/maintenance) of sarcomeres.

Despite having no effect on muscle structure, *unc-22(KtoA)*, displays a remarkable effect on muscle function. Nematodes have different locomotion patterns depending on media; in liquid they oscillate back and forth in a C-shaped pattern; on a semi-solid surface like agar they move in a sinusoidal pattern. We measured the ability of adult nematodes to move back and forth in a droplet of buffered (M9) water (Epstein and Thomson, 1974). As shown in Figure 3a, when normalized to wild type movement, the loss of function allele *unc-22(e66)* swam less well than wild type. However, *unc-22(KtoA)* mutant worms, unexpectedly, moved ~30% faster than wild type; a similar increase in motility (up to 50% increase) was observed for *unc-22(e105)*. We also measured the ability of adult worms to crawl on an agar surface. As indicated in Figure 3b, these results show that, similar to the results obtained with the swimming assay, *unc-22(e66)* crawls slower than wild type, but both *unc-22(e105)* and especially *unc-22(KtoA)*, crawl faster than wild type. Recently, we have adapted the use of optogenetics to obtain some measures of body wall muscle kinetics in *C. elegans* (Hwang *et al*. 2016). Channelrhodopsin-2 is expressed in motor neurons, and upon exposure to blue light, contraction of body wall muscles is induced. As a proxy of this contraction, we measure the relative body area of the worm, as shown in Figure 3c, and by fitting curves, can derive rate constants for contraction and for relaxation. Figure 3d shows results of this assay on multiple worms from wild type and three *unc-22* mutant strains. As reported previously (Hwang *et al*. 2016), *unc-22(e66)*(red line in Figure 3d) contracts less well and cannot fully maintain a maximally contracted state. In contrast, both *unc-22(e105)*, and especially *unc-22(KtoA)* contract more deeply than wild type. Another

way to look at the graphical representation of the optogenetic experiment in Figure 3d, is to plot the "relative body area at steady state" (Figure 3g). In this way, it is clear that both *unc-22(e105)* and *unc-22(KtoA)* contract more, and *unc-22(e66)* contracts less well than wild type. The reduced contraction ability of *unc-22(e66)* is likely due to the disorganization of its myofilament lattice (Figure 2). We are not sure how to interpret the patterns of *unc-22(e105)* and *unc-22(KtoA)*; perhaps these mutants are induced to contract more because they begin from a more relaxed state as compared with wild type. In addition, all the *unc-22* mutant strains show faster rates of contraction and relaxation (Figure 3 e and f).

In contrast to all other *unc-22* mutant strains investigated (approximately 100), *unc-22(e105)* and *unc-22(KtoA)* are unusual in that they have normal sarcomeric structure and move faster than wild type animals. *unc-22(e105)* has a missense mutation in Ig7, changing a highly conserved glycine to an arginine (Matsunaga *et al*. 2015). Recently we determined the mutation sites in 10 randomly selected *unc-22* mutant strains; all 10 strains contained either nonsense mutations or insertions/deletions resulting in frameshifts and consequent stop codons (unpub. data). Thus, *unc-22(e105)* and *unc-22(KtoA)* mutants express twitchin molecules with missense mutations near the N- and C-termini of twitchin. The similarity in phenotypes perhaps suggests an interaction between the N- and C-termini of twitchin. One possibility is that twitchin is an anti-parallel dimer in which Ig7 interacts with the kinase domain. Additional experiments are required to test this speculation.

WormBase predicts the existence of 9 *unc-22* isoforms (A through I) based on use of alternative promoters and alternative splicing, and this prediction has been verified by partial cDNA analysis. Interestingly, all the predicted twitchin isoforms

contain the kinase domain, a further indication of the importance of this domain for the function of twitchin. To determine where these isoforms are expressed, we created transgenic animals that express GFP from genomic segments upstream of the putative starts for these isoforms. As shown in Figure S4, isoforms A,B, F, G, H and I are expressed in body wall and vulval muscles; isoform D is expressed only in pharyngeal muscle, and both isoforms C and E are expressed in body wall and pharyngeal muscle. To assess the possible function of twitchin kinase in pharyngeal muscle, we assayed pharyngeal muscle pumping rate and pharyngeal muscle structure in *unc-22(KtoA)*. As shown in Figure S5, *unc-22(KtoA)* shows a slightly reduced but statistically significant pharyngeal pumping rate compared to wild type, although *unc-22(KtoA)* shows normal pharyngeal muscle structure by polarized light microscopy.

The fact that loss of twitchin kinase activity results in increased velocity of swimming and crawling, and greater contraction and faster rates of contraction and relaxation, suggests that the function of normal twitchin kinase activity is to inhibit the contraction/relaxation cycle. Among invertebrates (*C. elegans, Drosophila, Aplysia*) twitchin and the similar protein projectin have active protein kinase domains (Lei *et al*., 1994; van Castelmur *et al*. 2012; Ayme-Southgate *et al*. 1995; Heierhorst *et al*. 1995;Heierhorst *et al*. 1996). Twitchin kinase activity is also likely to be conserved among nematodes; our survey of the sequences of twitchin orthologs from 27 nematode species indicates that the catalytic lysine of twitchin kinase is conserved (Table S1). Therefore, it seems that among invertebrates, there has been selection to maintain twitchin kinase activity. To test this hypothesis, we set up a "competition assay". First, we outcrossed *unc-22(KtoA)* three more times to wild type (total of 6 outcrosses), and obtained both *unc-22(KtoA)* and "wild type" from

this sixth outcross. We then cultured 50 GFP tagged wild type animals with 50 tester animals (*unc-22(KtoA) 6X oc*; wild type 6x oc; our usual wild type strain), and measured the frequency of the tester strain after four generations. If two strains were equally competitive, we would expect the 50% frequency to be maintained. However, as shown in Figure 4, *unc-22(KtoA)* fared poorly. Indeed, after four generations this strain decreased in frequency to ~30% of the population, which was a significantly lower proportion relative to both the wild type 6X oc and the usual wild type strain (Figure 4; $\chi^2_2 = 18.6$, $p < 0.0001$). Thus, at least in a laboratory setting, there is selective pressure to maintain the kinase activity of twitchin, despite this reducing the overall locomotion of the animals.

We next wondered what might explain the evolutionary maintenance of kinase activity. Since lack of kinase activity results in faster moving worms, and presumably more metabolic activity and generation of more reactive oxygen species, we speculated that *unc-22(KtoA)* might have a shorter lifespan. However, as shown in Figure S6, this is not the case; in fact, the mean lifespan of wild type is 14 days, whereas the mean lifespan of *unc-22(KtoA)* is 17 days. Given expression of twitchin in egg-laying muscles, we wondered if *unc-22(KtoA)* has a reduced brood size as compared with the GFP tagged wild type strain, and thus, over several generations, this might be a major contributor to the competition failure. Thus, we measured brood sizes of wild type (strain N2), *unc-22(KtoA)* 6X oc, "wild type" from this sixth outcross, and the GFP tagged wild type strain (Figure S7). To our surprise, the strain showing the smallest brood size was the GFP tagged wild type. Therefore, a reduction in brood size cannot explain the mechanism that leads to reduced fitness of animals lacking twitchin kinase activity.

The molecular mechanism by which twitchin kinase activity inhibits the contraction/relaxation cycle is also a matter of speculation. Studies of molluscan muscle suggest that twitchin may inhibit the rate of relaxation (Probst *et al*. 1994; Siegman *et al*. 1998). However, this property is likely due to the function of twitchin in establishing and maintaining the catch state of these muscles, in which twitchin physically attaches thick filaments to thin filaments (Yamada *et al*. 2001; Funabara *et al*. 2007). There is no evidence that nematode striated muscle has a catch state. The twitching of typical *unc-22* loss of function mutants had suggested that twitchin is involved in regulating muscle contraction (Moerman *et al*. 1988), and finding that twitchin contains a protein kinase domain homologous to the well-known regulator, smooth muscle myosin light chain kinase (Benian *et al*. 1989), had suggested a possible molecular mechanism. Until recently, all the *unc-22* mutants were reported to also show disorganization of sarcomeres, so that regions of twitchin responsible for these two possible functions (muscle regulation vs. sarcomere organization) could not be segregated. In Matsunaga *et al*. (2015), we showed that despite its twitching, *unc-22(e105)* has normal muscle structure. Similarly, we show here that *unc-22(KtoA)* has normal muscle structure, but in addition, fails to twitch. What is even more surprising is that both *e105* and KtoA mutants move faster than wild type, indicating that twitchin kinase activity normally inhibits the contraction/relaxation cycle. A clue to the molecular mechanism may arise from finally determining the substrate(s) for twitchin kinase, which we may be able to approach efficiently by utilizing *unc-22(KtoA)* mutant animals.

EXPERIMENTAL PROCEDURES

Plasmids for expression of wild type and K6290A mutant twitchin kinase.

The plasmid for expression of His-tagged wild type twitchin kinase catalytic core (Kin) was described previously (von Castelmur *et al*., 2012). The same catalytic core (Kin) with the K6290A mutation was amplified by PCR using primers TwiK-3 and TwiK-4 from a Fn-Kin-Ig K6290A template described in Matsunaga *et al*., 2015; the DNA sequence of this fragment was confirmed and then cloned into the pETM11 vector for expression as a His-tagged protein.

Recombinant protein production

All TwcK constructs were expressed in *E.coli* Rosetta (DE3) (Merck Millipore) and grown in LB media containing 25 µg/mL kanamycin and 34 µg/mL chloramphenicol. Growth was at 37^{o}C to an OD_{600} of 0.6-0.8 followed by induction of protein expression with 0.5 mM isopropyl β-D-1-thiogalactopyranoside and further growth at 18^{o}C for approximately 18 hours. Cell pellets were harvested by centrifugation and resuspended in 50 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, pH 7.9 (lysis buffer) supplemented with 20 µg/mL DNase I (Sigma Aldrich) and a complete EDTA-free protease inhibitor (Roche). Cell lysis was by sonication on ice, followed by clarification of the lysate by centrifugation.

In preparation for phosphotransfer assays, cell lysates were applied to Ni^{2+}-NTA resin (Qiagen) equilibrated in lysis buffer containing 20 mM imidazole. Elution was with 200 mM imidazole and samples were buffer exchanged into lysis buffer using PD-10 desalting columns (GE Healthcare). His_6-tag removal was by

TEV protease cleavage overnight at 4^{o}C, followed by subtractive Ni^{2+}-NTA purification. Size exclusion chromatography was carried out using a Superdex 200 16/60 column (GE Healthcare) in 50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 7.9. In preparation for SAXS, cell lyastes were applied to a 5 mL HisTrap HP column (GE Healthcare) equilibrated in lysis buffer containing 20 mM imidazole. Elution of His_6-tagged protein was by continuous imidazole gradient, followed by buffer exchange into 50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 7.9 (anion exchange buffer), His_6-tag cleavage by TEV protease and subtractive Ni^{2+} purification. Anion exchange was performed using a 5 mL QFF column equilibrated in anion exchange buffer (GE healthcare) with elution by continuous NaCl gradient. Size exclusion chromatography was carried out using a Superdex 200 16/60 column (GE Healthcare) in 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM TCEP, pH 7.9.

In vitro phosphorylation assay

In vitro phosphotransfer assays were set up in 40 µL reactions containing 30 ng of TwcK and 20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0.2 mg/mL BSA, 0.2 mg/mL peptide substrate. Reactions were initiated by addition of 200 µM [γ^{32}P]-ATP (2 µCi per reaction). The peptide substrate has the sequence KKRARAATSNVFS and is a model peptide substrate derived from chicken myosin light chain (kMLC11-23) as described in Heierhorst *et al*., 1996 and von Castelmur *et al*., 2012. Control reactions were included in all assays containing peptide substrate in the absence of TwcK and with TwcK in the absence of peptide substrate. Incubation of samples was at 25^{o}C for 20 minutes before being blotted on P81 phosphocellulose paper and sugmerged in 100 mM phosphoric acid to terminate reactions. Blotted samples were washed 4 times in 100 mM phosphoric for 5 minutes

per wash before being transferred to acetone and air-dried. Quantification was by Cherenkov counting using a Beckman Coulter scintillation counter.

Small angle X-ray scattering (SAXS)

SAXS data were collected at the B21 beamline of the Diamond Light Source synchrotron (Didcot, UK) using the integrated SEC-SAXS set-up including the HPLC device Agilent 1260C. Approximately 1-1.2mg of sample (TwcK-wt, TwcK-KA) was loaded in 45μL of buffer (50mM Tris pH 7.9, 50mM NaCl, 0.5mM TCEP) onto a pre-equilibrated Shodex KW403 column, at a flow rate of 0.16mL/min. Frames were collected for the entire eluate using a exposure time of 3s per frame and a sample cell thermostated to 20^{o}C. X-ray scattering was recorded on a Pilatus 2M detector (Dectris) at a sample to detector distance of 3.9m and a □=1Å. Data processing used ScÅtter (Diamond, *unpublished*). Scattering curves were analysed with PRIMUS (Konarev *et al*. 2003) and the ATSAS software suite (Petoukhov *et al*. 2012) to determine the radius of gyration (Rg), the maximum dimension (D_{max}), and the pair distribution function P(r). The experimental molecular mass (MM) was calculated according to Rambo & Tainer, 2013 (Rambo and Tainer, 2013) using ScÅtter.

Nematode strains

N2 (Bristol) is the primary wild-type strain, and standard growth conditions were used (Brenner, 1974). The following alleles were used in this study: *eat-2(ad465)*II, *unc-22*(*e66*)IV, *unc-22*(*e105*)IV, *unc-22*(*ct37*)IV, *zxIs6[unc-17p::ChR2(H134R):::YFP; lin15*$^{+}$*]*V (Liewald *et al*. 2008). The double mutant animal, *unc-22(K*6290*A); zxIs6,* was generated by crossing the respective single mutant animals. To check for mutations of mutant animals, we carried out single-worm PCR. After purification of PCR products, we examined a restriction enzyme digestion (*Pst*I) and DNA sequence analysis to verify the mutation.

Creation of *unc-22(K*6290*A)*

*unc-22(K*6290A*)* was generated using a CRISPR technique. For targeting to *unc-22*, a sequence, 5'-actccacatgaatctgaca-3', was cloned into a pDD162 vector using a Q5 site-directed mutagenesis kit (New England BioLabs, USA). To modify the genome using Cas-9 triggered homologous recombination, the following DNA mixtures were injected: 50 ng/µl pDD162 with targeting sequences, 30 ng/µl single-strand oligo as repair template, 30 ng/µl pTG96 as co-injection marker. For screening of the mutants, we picked F1 animals to a single PCR tube, and performed single-worm PCR. After purification of PCR products, we examined a restriction enzyme digestion (*Pst*I) and DNA sequence analysis to verify the mutation. The homozygous mutant was outcrossed 3X to wild type, before further analysis.

Whole genome sequencing

Genomic DNA from *unc-22(K*6290A*)* was prepared and provided to the Emory Integrated Genomics Core(EIGC) for Illumina sequencing. The Genomic Services Laboratory at the HudsonAlpha Institute for Biotechnology (Huntsville, AL) performed quality control and constructed standard Illumina sequencing libraries following the manufacturer's instructions. Sequencing was performed on an Illumina Hiseq version 3 with 100 basepair paired-end reads. Raw data were returned to the EIGC for bioinformatics analysis. Raw reads were mapped against the *C. elegans* reference genome (WS190/ce6) with the PEMapper software package. PEMapper maps short reads to a reference genome by first decomposing those reads in k-mers (16 mers in this case), and then performing a hash-based "rough mapping" of each read. After roughly mapping the read, the fine-scale position of the read is determined by a local Smith-Watterman alignment. Genotypes were determined using PECaller. Intuitively, we envision the 6 channels of data (number of A, C, G, T, deletion, and insertion reads) as being multinomially sampled, with some probability of drawing a read from each of the channels, but the probability varies from experiment to experiment and is itself drawn from a Dirichlet distribution. PECaller combines data across samples in order to identify the genotype with the highest likelihood at each sequenced base. Both PEMapper and PECaller are part of a custom software package developed at Emory University for mapping and identifying variant sites from Illumina raw sequencing data (D. J. Cutler and M. E. Zwick, personal communication). Analysis then focused on the ~60 kb *unc-22* gene. PECaller identified a single homozygous mutation specifying the desired KtoA change, and no other mutations in the *unc-22* gene.

Immunolocalization

Adult nematodes were fixed using a method described previously (Nonet *et al*., 1993; Wilson *et al*. 2012). Primary antibodies were used at the following dilutions: anti-ATN-1 (□-actinin; MH35; Francis and Waterston, 1991) at 1:200, anti-myosin heavy chain A (MHC A; MYO3; 5-6; Miller *et al*., 1983) at 1:200, anti-paramyosin (UNC-15; mouse monoclonal 5-23; Miller *et al*. 1983) at 1:200, anti-PAT-6 (Warner *et al*., 2013) at 1:200, anti-twitchin (I I II; Benian *et al*., 1996) at 1:200, anti-UNC-89 (rabbit polyclonal EU30; Benian *et al*. 1996) at 1:200. For anti-twitchin and anti-UNC-89, the secondary antibody was anti-rabbit Alexa 488 (Invitrogen). For anti-ATN-1, anti-MHC A, and anti-paramyosin, the secondary antibody was anti-mouse Alexa 594 (Invitrogen). For anti-PAT-6, the secondary antibody was anti-rat Alexa 594 (Invitrogen). Each secondary antibody used at 1:200 dilution. Images were captured at room temperature with a Zeiss confocal system (LSM510) equipped with an Axiovert 100M microscope and an Apochromat ×63/1.4 numerical aperture oil objective, in ×2.5 zoom mode. The color balances of the images were adjusted by using Adobe Photoshop (Adobe, San Jose, CA). We checked the staining at least 3 worms / each staining.

Swimming assays

For the swimming assays, synchronized young adult animals were transferred in M9 buffer (85 mM NaCl, 42 mM Na_2HPO_4, 22 mM KH_2PO_4 and 1 mM $MgSO_4$).

Body bends of individual young adult animals were counted for 1 minute. A body bend was scored as a complete deflection of the anterior portion of the nematode from the midline. At lease 15 animals of each genotype were analyzed.

Crawling on a plate and optogenetic assays

For locomotion assay, synchronized young adult animals were transferred onto fresh 5.5 cm NGM plates. After a 2 min. acclimation period, movies were recorded to analyze the behaviors of each animal. To induce backward motion, their head was gently prodded with a platinum wire. The movies were post-processed to extract the worm skeleton using custom software written in MATLAB. The velocity of backward crawling was measured, and normalized by the length of each worm. For optogenetics assays of muscle kinetics, synchronized young adult animals were loaded into a two-layer microfluidic device. To induce ChR2 photo-activation, the animals were illuminated with blue light (450-590 nm; 0.3 mW/mm^2) for 15 sec. on the device. For measurements of projected body area, movies were recorded. The movies were post-processed using custom software written in MATLAB.

Competitive Fitness Assays

Experimental *C. elegans* strains were competed against a GFP labeled strain with an N2 background (JK2735) that was obtained from the Caenorhabditis Genetics Center (University of Minnesota, Minneapolis, MN). Ten replicate competition assays were run for experimental strains: N2, wild type 6X oc, and *unc-22(KtoA)* 6X oc. Competition assays were run for four passages on a 100 MM petri

dish containing 30 mL NGM-lite (US Biological, Swampscott, MA) and seeded with 200 µl *Escherichia coli* (OP50), and stored at 20°C during competition. Prior to the assay nematode lines were bleach synchronized. Competition assays were initially set up at a ratio of 100 experimental L4 larvae: 100 GFP L4 larvae. The larvae were permitted to develop to adults and reproduce. Every 4 days (approximately 1 nematode generation), the entire dish, containing parents and offspring, was washed with 1 mL of M9 buffer and the nematodes were pelleted. Approximately 1000 individual nematodes were transferred to a new dish containing OP50, using methods described in Morran *et al* (2009). After the fourth passage, the frequency of the experimental strain relative to the GFP strain was measured in 200 individuals from a cross section of the petri dish to determine the relative fitness of the experimental strain to the GFP strain (Morran *et al*., 2009). An increase in the frequency of the experimental strain relative to the GFP indicates greater competitive fitness for the experimental strain. Additionally, ten replicates of only JK2735 individuals (no experimental strains) were run to test for the loss of GFP as a control. We observed no loss of GFP expression across all ten replicates over four passages.

A Kruskal-Wallis nonparametric test was performed in JMP 12 (SAS Institute, Cary, NC) on the proportion of experimental strain values measured after 4 generations to compare the competitive fitness of the experimental strains.

Nicotine sensitivity assays using a WMicrotracker device

The method of nicotine assay was described previously (Matsunaga *et al*., 2015). Synchronized young adult animals of each genotype were collected in M9 buffer. After a final wash with M9 buffer containing 0.01% Triton X-100 (M9T), a

worm slurry (worm pellet : M9T = 1 : 5) was prepared. For the assays, 96-wells plates were used. To each well was added 40 µl of M9T followed by 10 µl of slurry, delivering approximately 50-100 worms per well. The plate was placed in the dark at room temperature for 1 hour. Then, locomotive activities under no-nicotine conditions were measured for 1 hour using a WMicrotracker (Phylumtech, Argentina) with the program "*C. elegans* (>L4)", supplied by the manufacturer. After the measurement, 50 µl of M9T containing 0.2% or 0.1% nicotine solution was added to each well, and locomotive activities were measured using the WMicrotracker. For each strain and nicotine concentration, 8 independent wells were assayed. On the graphs of locomotive activity vs. time, each point represents the mean and standard error.

Western blotting

Equal amounts of total protein from wild type N2, and *unc-22(K6290A)* were separated by 5% polyacrylamide SDS gels (50V for 40min., 100V for 10 min., and 200V for 60min.), transferred to nitrocellulose membrane (Bio-Rad Laboratories, Inc.) using transfer buffer [25 mM Tris-Base, 192 mM glycine and 20% methanol (v/v)] at 100V for 2 h, blocked overnight in 5% skim-milk (w/v)-TBST buffer [0.1% Tween-20 (v/v), 1× TBS (pH 7.6)], reacted with anti-twitchin (Benian *et al*., 1996) at 1:4000 in the same buffer for 1 h at room temperature. After washing, the blot was incubated with anti-rabbit HRP-conjugated secondary antibody (CAT#: NA9340) (GE Healthcare UK Ltd., England). The signal from the membrane was obtained using a Pierce ECL Western Blotting Substrate (Thermo SCIENTIFIC) and HyBlot CL® film (Denville Scientific Inc. USA).

Transgenic animals for determining tissue expression of *unc-22* isoforms

The following plasmids were constructed:

unc-22p (ABFGHI)::*gfp* has a 3.0 kb genomic fragment that includes the putative promoter sequence upstream of *unc-22*, a putative 5′-untranslated region, initiator ATG and the first 24 nucleotides of coding sequence of *unc-22*, amplified by PCR from *C. elegans* genomic DNA, and cloned into the pPD_venus vector, which contains a multiple cloning site followed by *venus* coding sequence and *unc-54* 3′-untranslated region, using *Pst*I and *Sma*I sites.

unc-22p (D)::*gfp* has a 3.0 kb genomic fragment that includes the putative promoter sequence upstream of *unc-22*, a putative 5′-untranslated region, initiator ATG and the first 24 nucleotides of coding sequence of *unc-22*, amplified by PCR from *C. elegans* genomic DNA, and cloned into the pPD_venus vector, which contains a multiple cloning site followed by *venus* coding sequence and *unc-54* 3′-untranslated region, using *Pst*I and *Msc*I sites.

unc-22p (C)::*gfp* has a 3.0 kb genomic fragment that includes the putative promoter sequence upstream of *unc-22*, a putative 5′-untranslated region, initiator ATG and the first 24 nucleotides of coding sequence of *unc-22*, amplified by PCR from *C. elegans* genomic DNA, and cloned into the pPD_venus vector, which contains a multiple cloning site followed by *venus* coding sequence and *unc-54* 3′-untranslated region, using *Pst*I and *Sma*I sites.

unc-22p (E)::*gfp* has a 3.0 kb genomic fragment that includes the putative promoter sequence upstream of *unc-22*, a putative 5′-untranslated region, initiator ATG and the first 24 nucleotides of coding sequence of *unc-22*, amplified by PCR from *C. elegans* genomic DNA, and cloned into the pPD_venus vector, which contains a multiple cloning site followed by *venus* coding sequence and *unc-54* 3′-untranslated region, using *Pst*I and *Sma*I sites.

The constructed plasmids were injected at 20 ng/µl into the gonads of wild-type animals along with the pRF4 [*rol-6*(*su1006*)] plasmid as co-injection marker (80 ng/µl)(Mello and Fire, 1995). At least three independent stable transgenic lines were generated for each plasmid. Fluorescence and DIC images were obtained with Axioskop microscope and AxioCam MR microscope camera (Carl Zeiss).

Pumping assays

Synchronized young adult animals (15-20 worms) are placed on a 2% agar pad with 5 µl of M9 buffer containing OP50 bacteria. A cover slip was placed on top of the worms. After a 15 min. acclimation period, pharyngeal pumping was recorded using a 40X objective lens (DIC optics) equipped with Axioskop microscope (Carl Zeiss) and AxioCam MR microscope camera (Carl Zeiss). Pharyngeal pumping of individual young adult animals were counted for 20 sec. At least 20 animals of each genotype were analyzed.

Lifespan measurements

Ampicillin-killed E. coli strain OP50 was used as a food source in the experiment to avoid any effects of live E. coli. Lifespan with killed E. coli is longer compared to that with live E. coli probably because toxicity from growing bacteria is avoided (Garigan *et al*. 2002). Worms were raised until the L4 molt and were then transferred onto a new plate containing 40 μM 5-fluoro-2′-deoxyuridine (Sigma Aldrich) to prevent self-fertilization (Hosono, 1978; Mitchell *et al*. 1979). The day of transfer at the L4 molt was counted as adult day 0. Worms were judged to be dead when they did not respond to a mechanical stimulus. To focus on aging, worms displaying extruded internal organs were excluded from our analysis. The results of the survival assays were analyzed using the Kaplan-Meier method, and significance was measured with the log-rank test using GraphPad Prism (GraphPad Software).

Brood size measurements

Synchronized L4 stage animals were transferred onto 3 cm NGM plates (1 worm/plate) with fresh OP50 bacteria. The worms were transferred everyday (24 h.) onto the new plates for 6 days. The number of egg was counted on each plate. After 24 hours at 20 ºC, hatched worms were counted as the number of progeny on the plate.

SUPPLEMENTAL INFORMATION

Supplemental information includes Supplemental Experimental Procedures, seven figures and one table.

ACKNOWLEDGMENTS

G.M.B., H.L. and O.M. gratefully acknowledge support from a Human Frontier Science Program grant (RGP0044/2012).

REFERENCES

Ayme-Southgate, A., Southgate, R., Saide, J., Benian, G.M., and Pardue, M.L. (1995). Both synchronous and asynchronous muscle isoforms of projectin (the Drosophila bent locus product) contain functional kinase domains. J. Cell Biol. 128, 393-403.

Benian, G.M., Kiff, J.E., Neckelmann, N., Moerman, D.G., and Waterston, R.H. (1989). Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. Nature *342*, 45-50.

Benian, G.M., Tinley, T.L., Tang, X., and Borodovsky, M. (1996). The Caenorhabditis elegans gene unc-89, required for muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains. J. Cell Biol. *132*, 835-848.

Epstein H.F., and Thomson J.N. (1974). Temperature-sensitive mutation affecting myofilament assembly in Caenorhabditis elegans. Nature *250*, 579-580.

Francis, R., and Waterston, R.H. (1991). Muscle cell attachment in Caenorhabditis elegans. J. Cell Biol. *114*, 465-479.

Funabara, D., Hamamoto, C., Yamamoto, K., Inoue, A., Ueda, M., Osawa, R., Kanoh, S., Hartshorne, D.J., Suzuki, S., and Watabe, S. (2007). Unphosphorylated twitchin forms a complex with actin and myosin that may contribute to tension maintenance in catch. J. Exp. Biol. *210*, 4399-4410.

Garigan, D. Hsu, A.-L., A. G. Fraser, A.G., Kamath, R.S., Abringet, J., and Kenyon, C. (2002). Genetic analysis of tissue aging in *Caenorhabditis elegans*: a role for heat-shock factor and bacterial proliferation. Genetics *161*, 1101–1112.

Heierhorst, J., Probst, W.C., Kohanski, R.A., Buku, A., and Weiss, K.R. (1995). Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase. Eur. J. Biochem. 233, 426-431.

Heierhorst, J., Tang, X., Lei, J., Probst, W.C., Weiss, K.R., Kemp, B.E., and Benian, G.M. (1996). Substrate specificity and inhibitor sensitivity of Ca2+/S100-dependent twitchin kinases. Eur. J. Biochem. *242*, 454-459.

Hosono, R. (1978). Sterilization and growth inhibition of *Caenorhabditis elegans* by 5-fluorodeoxyuridine. Exper. Geron. *13*, 369–374.

Hwang, H., Barnes, D.E., Matsunaga, Y., Benian, G.M., Ono, S., and Lu, H. (2016). Muscle contraction phenotypic analysis enabled by optogenetics reveals functional relationships of sarcomere components in Caenorhabditis elegans. Sci. Rep. *6*:19900. doi: 10.1038/srep19900.

Iyer, G.H., Garrod, S., Woods, V.L., Jr., and Taylor, S.S. (2005). Catalytic independent functions of a protein kinase as revealed by a kinase-dead mutant: study of the Lys72His mutant of cAMP-dependent kinase. J. Mol. Biol. *351*, 1110-1122.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. *36*, 1277–1282.

Lei, J., Tang, X., Chambers, T.C., Pohl, J., and Benian, G.M. (1994). Protein kinase domain of twitchin has protein kinase activity and an autoinhibitory region. J. Biol. Chem. *269*, 21078-21085.

Liewald, J.F., Brauner, M., Stephens, G.J., Bouhours, M., Schultheis, C., Zhen, M., and Gottschalk, A. (2008). Optogenetic analysis of synaptic function. Nat. Methods *5*, 895-902.

Matsunaga, Y., Qadota, H., Furukawa, M., Choe, H., and Benian, G.M. (2015). Twitchin kinase interacts with MAPKAP kinase 2 in *C. elegans* striated muscle. Mol. Biol. Cell. *26*, 2096-2111.

Mello, C., Fire, A. (1995). DNA transformation. Methods Cell Biol *48*, 451-482.

Miller, D.M., 3rd, Ortiz, I., Berliner, G.C., and Epstein, H.F. (1983). Differential localization of two myosins within nematode thick filaments. Cell *34*, 477-490.

Mitchell, D.H., Stiles, J.W., Santelli, J., and Sanadi, D.R. (1979). Synchronous growth and aging of *Caenorhabditis elegans* in the presence of fluorodeoxyuridine. J. Geron. *34*, 28–36.

Moerman, D.G., and Baillie, D.L. (1979). Genetic organization in *C. elegans*: fine-structure analysis of the *unc-22* gene. Genetics *91*, 95-103.

Moerman, D.G., Benian, G.M., Barstead, R.J., Schriefer, L.A., and Waterston, R.H. (1988). Identification and intracellular localization of the *unc-22* gene product of *Caenorhabditis elegans*. Genes Dev *2*, 93-105.

Morran, L.T., Parmenter, M.D., and Phillips, P.C. (2009). Mutation load and rapid adaptation favour outcrossing over self-fertilization. Nature *462*, 350-352.

Nonet, M.L., Grundahl, K., Meyer, B.J., and Rand, J.B. (1993). Synaptic function is impaired but not eliminated in *C. elegans* mutants lacking synaptotagmin. Cell *73*, 1291-1305.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis J. Appl. Cryst. *40*, s223-s228.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Cryst. *45*, 342-350.

Probst, W.C., Cropper, E.C., Heierhorst, J., Hooper, S.L., Jaffe, H., Vilim, F., Beushausen, S., Kupfermann, I., and Weiss, K.R. (1994). cAMP-dependent phosphorylation of Aplysia twitchin may mediate modulation of muscle contractions by neuropeptide cotransmitters. Proc. Natl. Acad. Sci. USA *91*, 8487-8491.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477–481.

Siegman, M.J., Funabara, D., Kinoshita, S., Watabe, S., Hartshorne, D.J., and Butler, T.M. (1998). Phosphorylation of a twitchin-related protein controls catch and calcium sensitivity of force production in invertebrate smooth muscle. Proc Natl Acad Sci USA *95*, 5383-5388.

von Castelmur, E., Strumpfer, J., Franke, B., Bogomolovas, J., Barbieri, S., Qadota, H., Konarev, P.V., Svergun, D.I., Labeit, S., Benian, G.M., Schulten, K., and Mayans, O. (2012). Identification of an N-terminal inhibitory extension as the primary mechanosensory regulator of twitchin kinase. Proc. Natl. Acad. Sci. USA *109*, 13608-13613.

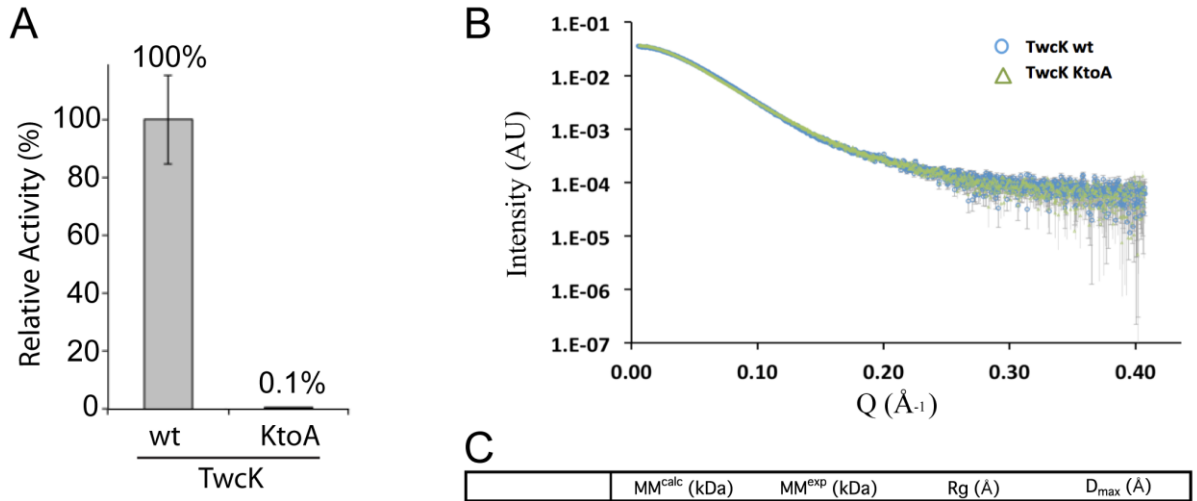
Warner, A., Xiong, G., Qadota, H., Rogalski, T., Vogl, A.W., Moerman, D.G., and Benian, G.M. (2013). CPNA-1, a copine domain protein, is located at integrin adhesion sites and is required for myofilament stability in *C. elegans*. Mol. Biol. Cell *24*, 601-616.

Wilson, K.J., Qadota, H., and Benian, G.M. (2012). Immunofluorescent localization of proteins in Caenorhabditis elegans muscle. Meth. Mol. Biol. *798*, 171-181.

Yamada, A., Yoshio, M., Kojima, H., and Oiwa, K. (2001). An in vitro assay reveals essential protein components for the "catch" state of invertebrate smooth muscle. Proc. Natl. Acad. Sci. USA *98*, 6635-6640.

FIGURES

Fig. 1



C

	MM^{calc} (kDa)	MM^{exp} (kDa)	Rg (Å)	D_{max} (Å)
TwcK wt	66.1	55.0	31.79±2.47	120
Twck KtoA	66.1	57.0	33.12±1.28	122

Figure 1. A lysine involved in ATP binding when mutated to alanine (KtoA) eliminates catalytic activity but has no effect on the structure of twitchin kinase in vitro.

(A) The catalytic core of twitchin kinase was expressed and purified from E. coli as either the wild type or with the KtoA mutation. The ability to phosphorylate a model peptide was measured with □-32P-ATP. As shown, the KtoA mutation nearly eliminates catalytic activity.

(B) SAXS analysis shows that the KtoA mutation does not affect the overall structure of the twitchin kinase catalytic core. Experimental scattering curves from TwcK-wt (blue circle; capped error bars) and TwcK-KA (green triangles; uncapped error bars) scaled using DATADJUST (Petoukhov *et al*. 2012). (C) Molecular parameters calculated from scattering data. MM, Rg and D_{max} denote molecular mass, radius of gyration and maximal particle size, respectively. MM^{calc} is the theoretical MM of the constructs computed from primary sequence. Rg was calculated using AUTORG (Petoukhov *et al*. 2007).

Fig. 2

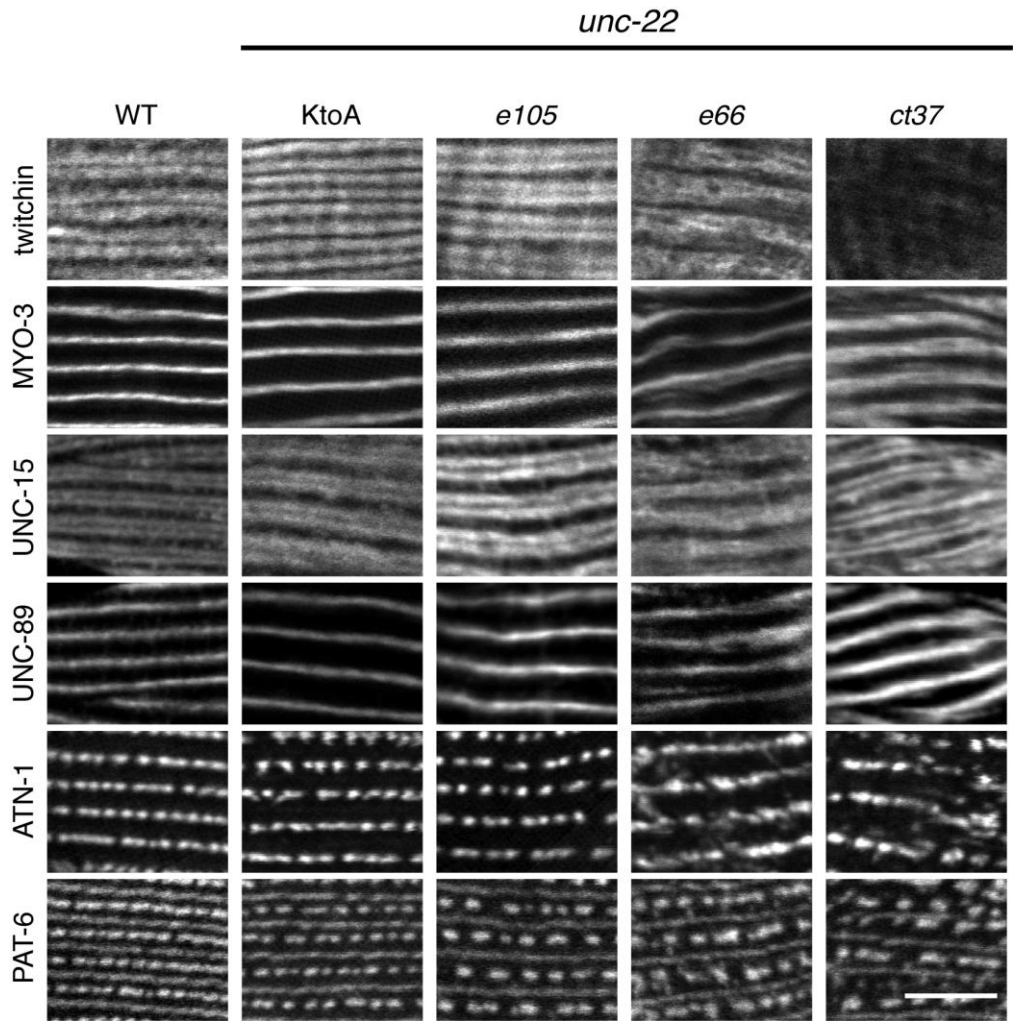


Figure 2. Expression of twitchin containing the KtoA mutation in its kinase domain shows normal muscle sarcomere structure, including normal localization of twitchin.

Wild type (WT) and the indicated *unc-22* mutant nematodes were fixed and immunostained with antibodies to the indicated sarcomeric proteins; twitchin, MYO-3 (myosin heavy chain A) and UNC-15 (paramyosin) of the A-bands; UNC-89 (obscurin) of the M-lines; ATN-1 (α-actinin) of dense bodies (Z-disk analogs); and PAT-6 (α-parvin) of M-lines and dense bodies. As indicated, *unc-22(KtoA)* shows normal localization of every sarcomeric protein tested, including twitchin. For comparison, *unc-22(e105)* shows the same normal localization pattern as *unc-22(KtoA)*, but the loss of function allele *unc-22(e66)* and the null allele *unc-22(ct37)* show disorganization of each of these marker proteins. Scale bar, 10 μm.

Fig. 3

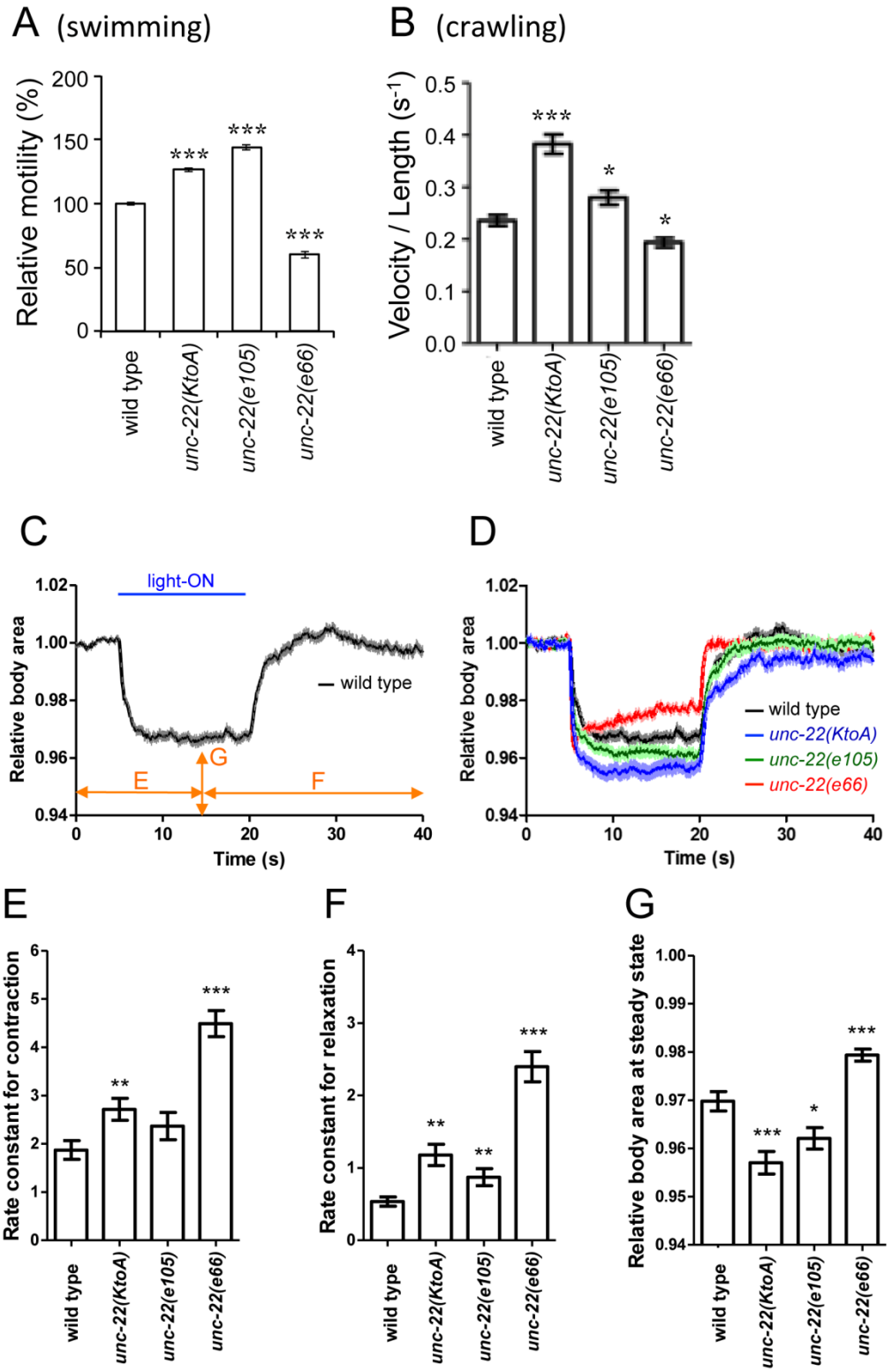

A (swimming)
Relative motility (%)
200
150
100
50
0
***
***
***
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)
B (crawling)
Velocity / Length (s-1)
0.5
0.4
0.3
0.2
0.1
0.0
***
*
*
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)
C
light-ON
Relative body area
1.02
1.00
0.98
0.96
0.94
wild type
G
E
F
0
10
20
30
40
Time (s)
D
Relative body area
1.02
1.00
0.98
0.96
0.94
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)
0
10
20
30
40
Time (s)
E
Rate constant for contraction
6
5
4
3
2
1
0
**
***
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)
F
Rate constant for relaxation
4
3
2
1
0
**
**
***
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)
G
Relative body area at steady state
1.00
0.99
0.98
0.97
0.96
0.95
0.94
***
*
***
wild type
unc-22(KtoA)
unc-22(e105)
unc-22(e66)

Figure 3. Nematodes lacking twitchin kinase activity move faster and show a greater degree of muscle contraction.

(A) Results of a swimming assay in which the number of times an animal moves back and forth in liquid are counted. The slower motility of the loss of function allele *unc-22(e66)* is typical of *unc-22* mutants. Both *unc-22(KtoA)* and *unc-22(e105)* display, unexpectedly, faster motility.

(B) Measurement of crawling velocity of nematodes on the surface of agar; results are normalized to the length of the animal. Results are similar to those obtained for swimming.

(C) Typical results of an optogenetic experiment to measure contraction and relaxation of wild type nematodes. Relative body area is a measure of the contraction state of body wall muscle.

(D) Optogenetic experiments on wild type and three *unc-22* mutants. As compared to wild type animals, *unc-22(KtoA)* and *unc-22(e105)* display a greater degree of muscle contraction, whereas *unc-22(e66)* shows less muscle contraction and an inability to maintain the contracted state.

(E, F) Rate constants for contraction (E) and for relaxation (F) derived from fitting curves to data shown in part (D).

(G) The relative body area at steady state, the part of the optogenetic curve labeled “g” in part (C) is displayed. These numbers emphasize that *unc-22(KtoA)* and *unc-22(e105)* contract more than wild type.

Fig. 4

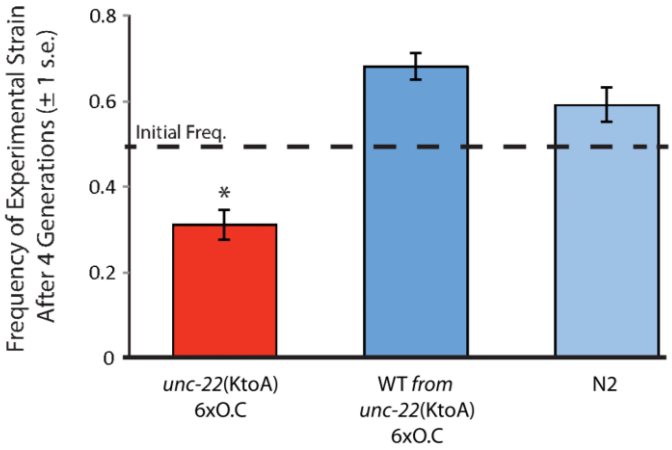


Figure 4. A competition assay demonstrates that *unc-22(KtoA)* has a selective disadvantage to wild type.

Equal numbers of GFP tagged wild type and the indicated strains were cultured on plates with E. coli food, and after four generations the frequency of each strain was determined. "*unc-22(KtoA)* 6X oc", is the *unc-22(KtoA)* after six outcrosses to wild type; "wild type 6X oc" is an isogenic line containing wild type sequence in the *unc-22* gene derived from the same sixth outcross; "wild type" is our usual laboratory wild type strain.

Supplemental Fig. 1

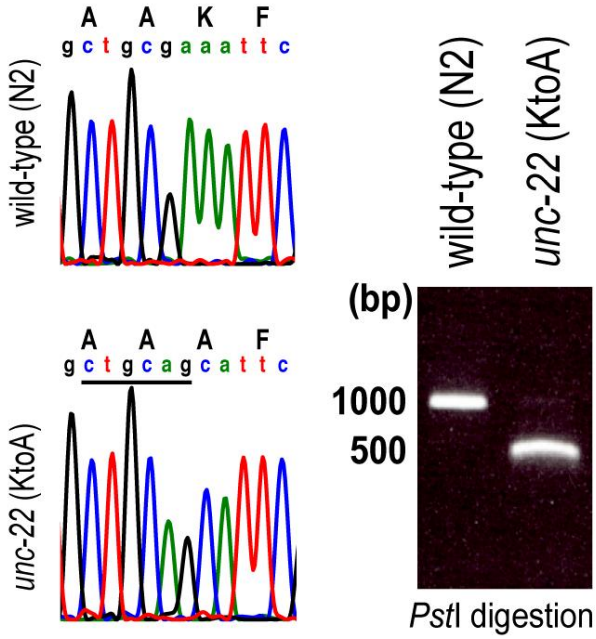


Figure S1. Generation of *unc-22(KtoA)* by CRISPR/Cas9.

On the right is an agarose gel showing PCR fragments digested with PstI that allow discrimination of wild type from the *unc-22(KtoA)* sequence. On the left are portions of the DNA sequence chromatograms of these fragments together with conceptual translation; the underline denotes the PstI recognition site.

Supplemental Fig. 2

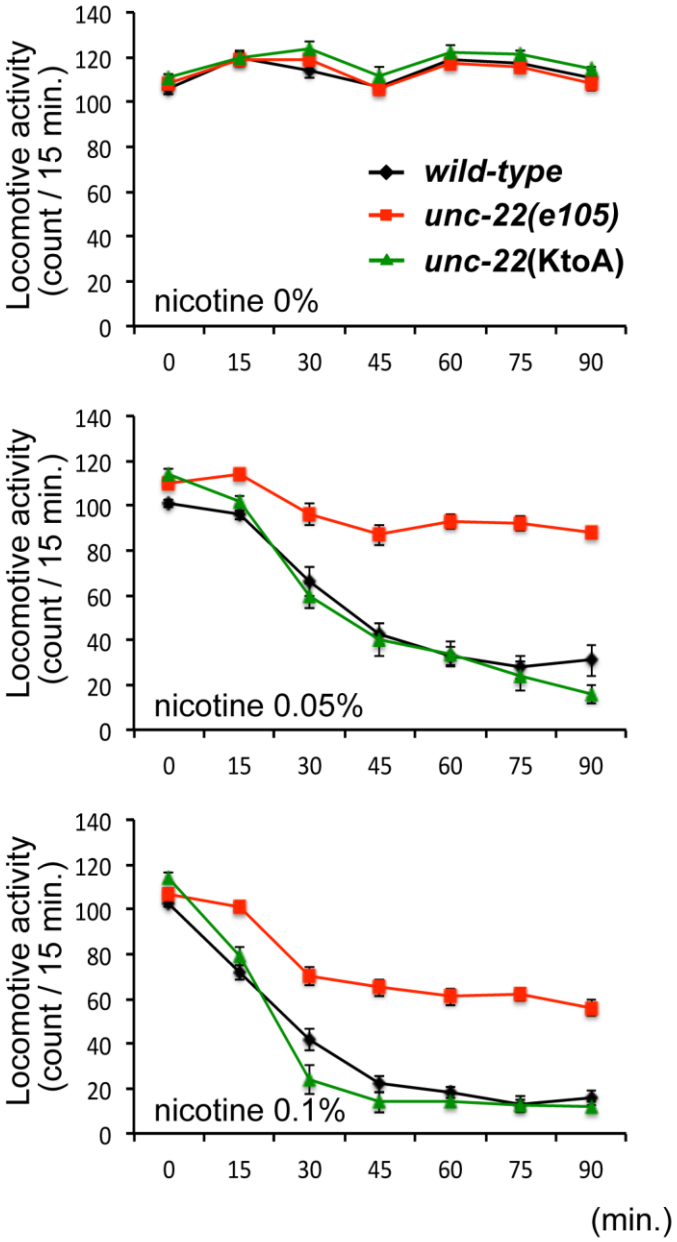


Figure S2. Response of wild type, *unc-22(e105)* and *unc-22(KtoA)* nematodes to nicotine.

A WMicrotracker (DesignPlus) was used to monitor the locomotion of multiple worms per well in a microtiter dish over time during exposure to a solution of 0.00%, 0.05% or 0.1% nicotine.

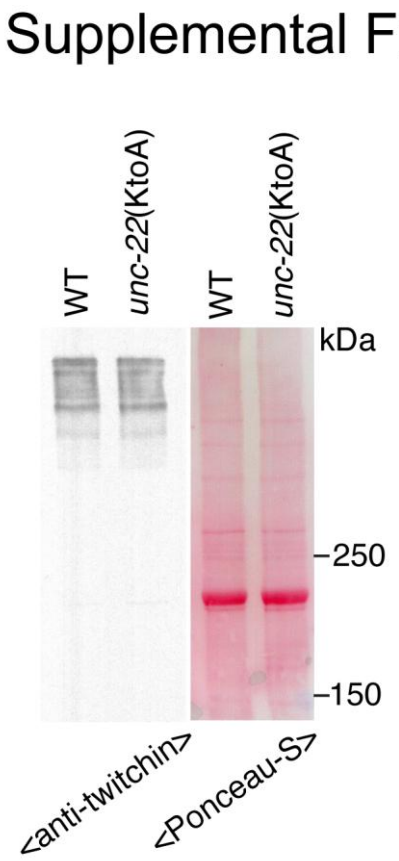


Figure S3. By western blot, *unc-22(KtoA)* expresses normal levels of twitchin isoforms of the appropriate size.

Extracts from wild type (WT) or *unc-22(KtoA)* were separated on a 5% SDS-PAGE and transferred to a membrane. On the right is shown the blot after Ponceau-S staining; the positions of molecular weight markers are indicated. On the left is shown the result of reaction to anti-twitchin antibodies.

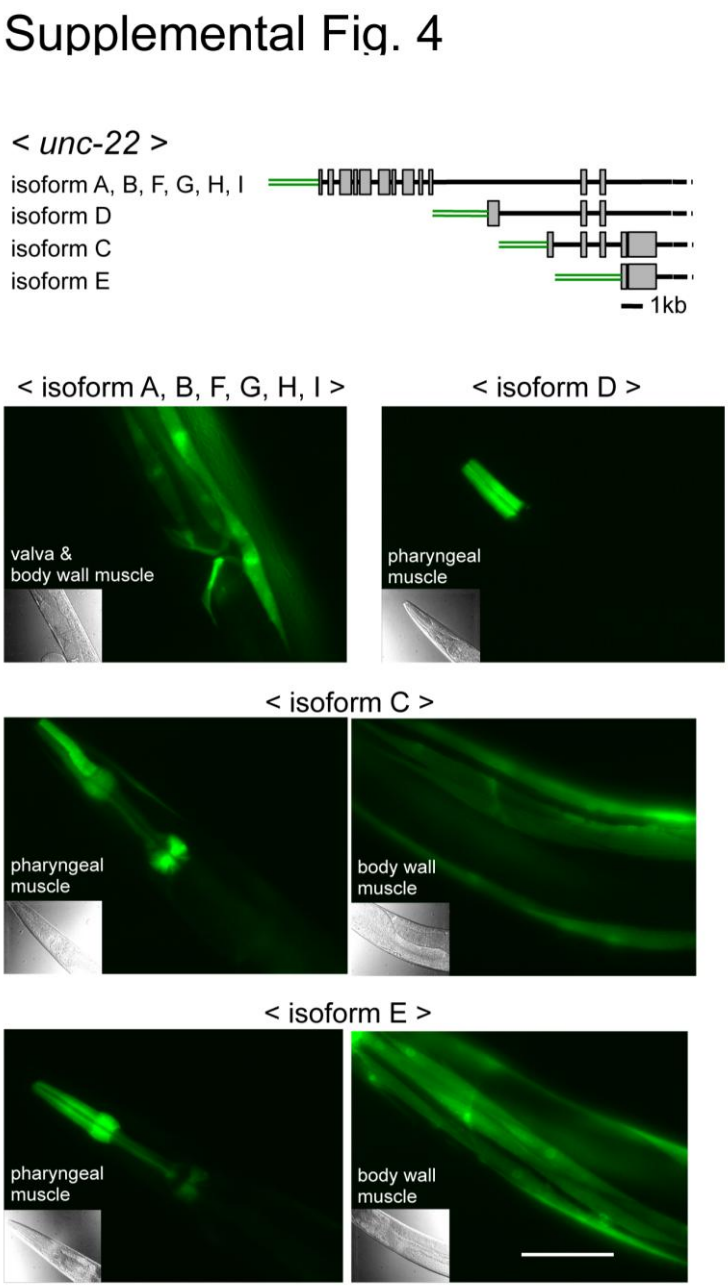


Figure S4. Expression patterns of 4 putative promoters for *unc-22*.

On top is a schematic depicting the 5' ends of the 9 isoforms of *unc-22* mRNAs predicted on WormBase; grey boxes represent exons, black lines, introns, and green lines represent the 3 kb of upstream sequence, together with the putative 5′-untranslated region, initiator ATG and the first 24 nucleotides of coding sequence of *unc-22* fused in-frame to GFP. Fluorescent images of expression patterns are shown below; insets show Nomarski images of the given anatomical regions. Scale bar, 30 μm.

Supplemental Fig. 5

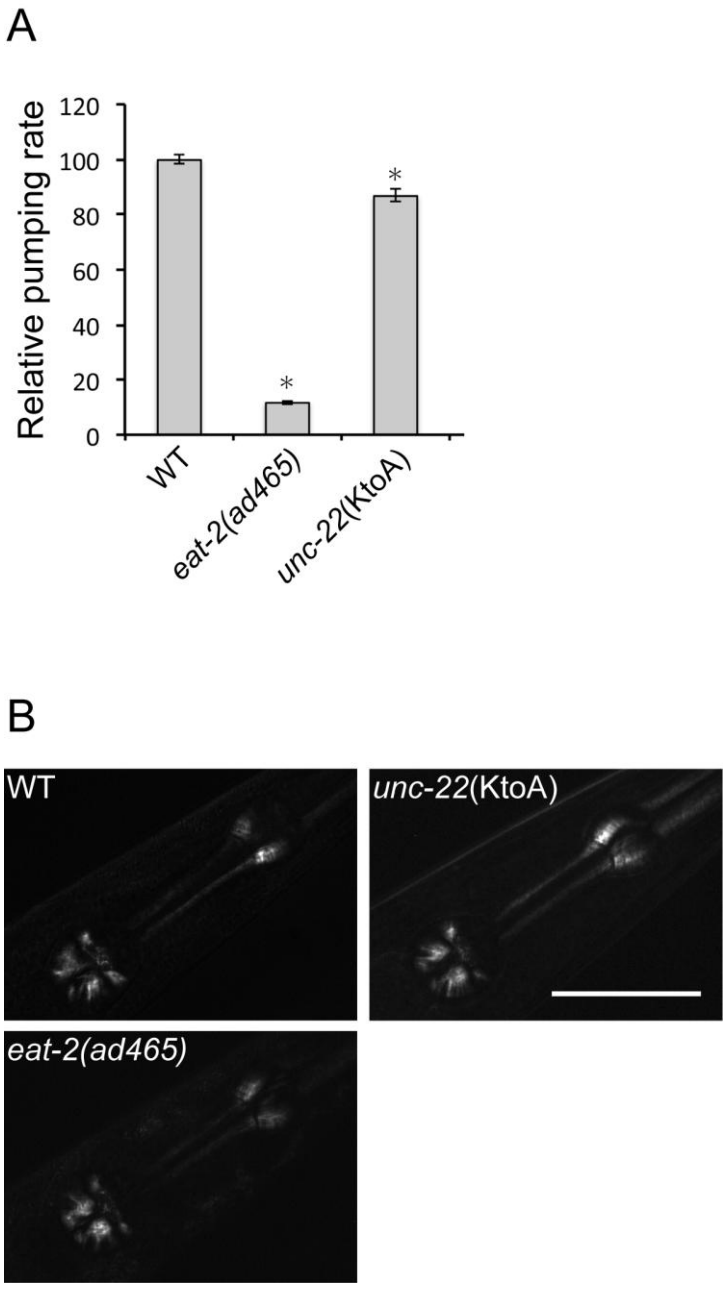


Figure S5. *unc-22(KtoA)* has slightly reduced pharyngeal pumping but normal pharyngeal muscle structure.

(A) Pharyngeal pumping rates, normalized to wild type (WT) of wild type, *unc-22(KtoA)*, and *eat-2(ad465)*(as a "positive control"). (B) Polarized light microscopy of the pharynxes of the same mutants. Scale bar, 30 µm.

Supplemental Figure 6

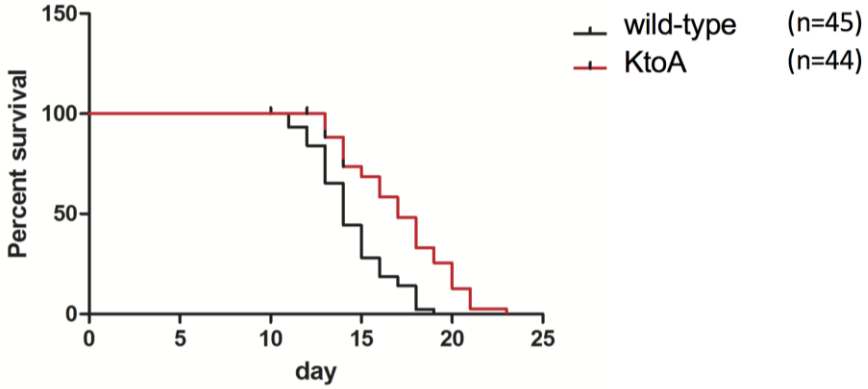


Figure S6. Adult lifespan of *unc-22(KtoA)* is slightly increased as compared to wild type.

The survival curves of N2 (wild type) and *unc-22(KtoA)* on NGM with ampicillin-killed OP50 E. coli are shown. The percentage of live worms is plotted against adult age. Day 0 corresponds to the L4 molt. The mean lifespan of wild type is 14 days, whereas the mean lifespan of *unc-22(KtoA)* is 17 days. By log-rank test, this is statistically significant with $P<0.01$.

Supplemental Figure 7

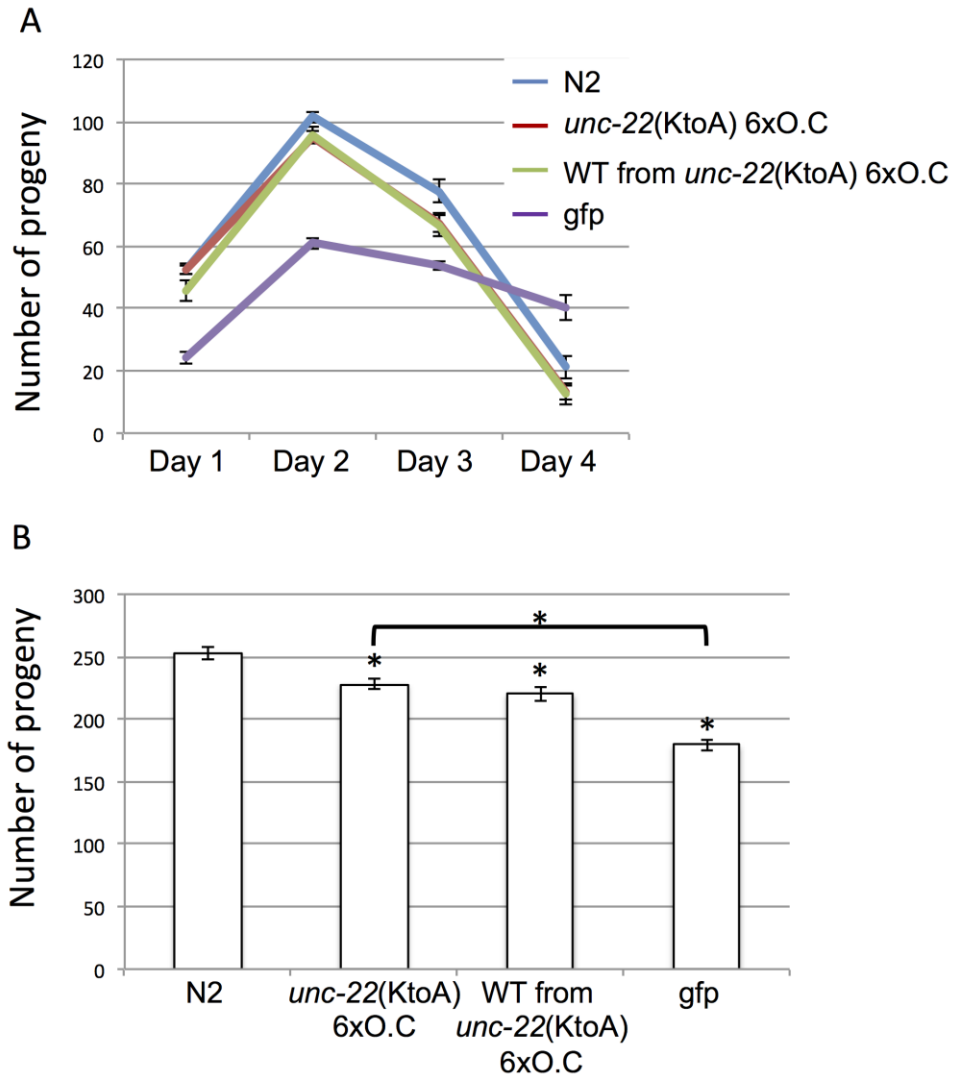


Figure S7. Brood size measurements of wild type (N2), *unc-22(KtoA)* 6X O.C., wild type derived from the 6th O.C., and a GFP tagged wild type strain.

(A) Brood sizes plotted for the first 4 days. Note that the GFP tagged wild type strain shows consistently fewer progeny, and also shows a delay or lag is producing progeny as compared to the other strains. (B) Total numbers of progeny shown as a bar graph. Note that all the strains have reduced brood sizes as compared to wild type (N2). More significantly in regards to the competition experiment, the GFP tagged wild type strain produces a smaller brood than unc-22(KtoA) 6X O.C. *: statistical significance at $p<0.5$.

7.2 Appendix II - Copyright permissions

Figure 1.1.1a was reprinted from *Journal of Muscle Research and Cell Motility*, **30**, Luther, P. The vertebrate muscle Z-disc: Sarcomere anchor for structure and signalling, 171-185, 2009, with permission from Springer.

Figure 4.1.2.1 was reprinted from *Annual Review of Biophysics,* **38**, Zocchi, G. Controlling proteins through molecular springs, 75-88, 2009, with permission from Annual Reviews.