A Chemical Crosslinking Mass Spectrometry-led exploration of Aurora A Kinase

Jaime Rosa Tarleton Pitts

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

August 2022

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to thank Howard Morris for funding my scholarship, and the University of Leeds Faculty of Biological Sciences for additional funding. The Biochemical Society and E-COST action also funded research trips, and the living/travel costs associated with these.

My gang of supervisors – Dr Megan Wright, Prof. Richard Bayliss and Prof. Frank Sobott – have all been immeasurably generous with their support, both scientific and pastoral. Their combined nurturing has made me the enquiring scientist I am today. I am grateful for their patience with my often half-formed experimental ideas, which has ultimately led me to completing this work. Reigning in and focus have been the watchwords. They and the science have also sustained me through the darkest parts of my life. Their gift of patience and understanding is one I will always treasure.

Great thanks is also owed to Prof. Andrea Sinz as a fourth unofficial supervisor, with the majority of MS measurement performed in Halle. During both trips I was made most welcome there.

I want to thank members of lab 1.49, the Bayliss group, Wright group, Sobott and Wilson groups for their support and instruction. Dr Martin Walko for schooling me in the art of peptide synthesis and Dr Eoin Leen for his persistent support through the biological aspects of my project. In addition to Dr James Ross dealing with those in silico experiments that “came in clutch”. I am grateful to the Sinz group members, especially Claudio, Anna and Christian, who made me welcome out there. Although entirely removed from the thesis, Josephina was critical to tissue culture, and I am grateful for Eoin, Selena and Mark, as their work was the foundational ideas for this projection.

I am indebted to data produced by Sammy Lawrence for collecting the IMS-MS data on my behalf which would have formed a pillar of a chapter in thesis if it weren’t for circumstances beyond my control. Also, to Lewis Jackson.

The work for this thesis spans peptide synthesis, mass spectrometry, tissue culture, flow cytometry, molecular dynamics simulations and too many coIPs which all required technical support. Named thanks is due to James and Rachel of the FBS MS faculty, Christian of Sinz group, J Hodrien, A Coleman and ARC Leeds more broadly, and Dr Sally Boxall and Dr Ruth Hughes for an extremely well-run Bio-imaging and Flow Cytometry facility. Thanks is also due to Mrs Honglin Rong and Mr Martin Huscroft.

All my thanks go to my friends, especially Jack and Sean, who seem to bear the brunt of my rage at science over coffee and zoom, helping me navigate my work.

I would finally like to thank my partner, Emmie Rose, for motivating me through my failures and supporting me in wrangling my rambunctious work. I am the woman I am because of her love, support and trust. She flew out to Germany to stop me leaving a trip early because of my suicidality, she paid for my healthcare so I didn’t quit, and she always met me with kindness and authenticity. Historians should never be subject to any discussion of kinases or XL-MS, and yet she has always helped in the formation of my ideas and the honing of my scholarship.

Abstract

Kinases function as signallers within the cell; using phosphorylation of the proteins with which they interact as molecular switches to perform functions in cells. Outside of their core catalytic domain, kinases can contain intrinsically disordered regions (IDRs) as part of their structure or interact with intrinsically disordered proteins (IDPs). The disorder, or lack of stable structure, of IDRs and IDPs can result in a protein or protein region being without any structure, or dynamically processing between stable structures, giving rise to an ensemble of conformations. The disorder and dynamic properties of such proteins makes the structural biologists' task of elucidating their molecular mechanism harder. But this is pressing work in the context of kinases, since their dis- and mis- regulation is a common trait of cancer, and drugging them is a topic of considerable interest. In this thesis, I investigate the value of chemical crosslinking mass spectrometry (XL-MS) in studying dysregulated kinases that feature IDRs or dynamic conformations. We used the mitotic serine/threonine kinase Aurora-A (AurA) and its activating intrinsically disordered partner protein TPX2 as a testbed for kinases as a target type, to explore the practicalities and limits of applying XL-MS to such a dynamic system.

XL-MS is an in-solution technique that uses a bi-functional reagent to generate distance restraints between the reactive side chains of residues that can be mapped to previously generated structures or models. We used the mass cleavable amine and hydroxyl reactive crosslinking disuccinimidyl dibutyric urea (DSBU), in conjunction with the software MeroX, to demonstrate that it was possible to generate useful structural data for dynamic kinases and their complexes, with this requiring good crosslink coverage. We found that the use of crosslink and monolink scoring in conjunction with MD simulations of kinase structures/conformers and literature data were required to adequately describe this system.

Ultimately, this thesis demonstrates the value of this methodology, as one that is both cost-effective and time efficient to study kinase structure and function, so long as it is used in combination with structural techniques that disentangle the conformational ensemble.

Table of Contents

List of Tables

List of Figures

Amino Acid Abbreviations

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Glossary and Abbreviations

AP-MS	Affinity-purification mass spectrometry
αC-helix	αC helix is located in the N-lobe beside the active site.
A loop	Activation loop of a kinase
Abl	Tyrosine-protein kinase ABL1; kinase linked to networks such as cell growth and survival
AGC kinase	The AGC family is named after the protein kinase A, G, and C families (PKA, PKC, PKG)
APC/C	anaphase promoting complex or cyclosome
AP-MS	Affinity-purification mass spectrometry
Apo(enzyme)	Apoenzyme is a protein part that binds to a non-protein cofactor to form an active conjugated enzyme known as the holoenzyme.
AST	Active site tether
ATP	Adenosine triphosphate
BCR	breakpoint cluster region protein
bHLHLZ	Basic helix–loop–helix Leucine Zipper
Bis-Tris	2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol
co-IP	Co-immunoprecipitation
Coomassie stain	Coomassie Brilliant Blue, protein stain used for SDS-PAGE gels
COSMIC (database)	Catalogue of Somatic Mutations in Cancer database of somatically acquired mutations found in human cancer
CPC	Chromosomal Passenger Complex
cPKA	cAMP-dependent protein kinase catalytic subunit
CSM	cross-link peptide-spectra match
CSS	Collision cross-section
Cα-Cα distance	measuring a crosslink distance using a straight line between the alpha carbon of each of the residue
dici/dico/doco	conformer nomenclature of kinases: DFG-in and αC-in or DFG-in and αC-out or DFG-out and αC-out
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EML4-ALK	a tumorigenic fusion of echinoderm microtubule-associated protein-like 4 and anaplastic lymphoma kinase
ESI-MS	Electrospray Ionisation Mass Spectrometry
FDR	False discovery rate
FLAG-tag	terminal protein tag, 8 amino acids in length, which high affinity monoclonal antibodies have been raised against and hence can be used for protein purification by affinity chromatography
(TR)-FRET	Time-resolved fluorescence energy transfer
GST-tag	A Glutathione S-transferase tag is a terminal protein tag which 220 amino acids in length
HA-tagged	Nine amino acid long tag derived from Human influenza hemagglutinin protein
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HER2	human epidermal growth factor receptor 2
His tag	Polyhistidine-tag, Six histidine residues placed terminally to allow a expressed protein to be purified by nickel-affinity resin
Histone H3	Protein that is part of the structure of chromatin
HM	hydrophobic motif (HM) and active-site tether (AST). The HM, also called the N-lobe tether (NLT),
HRMS	High resolution mass spectrometry
IDP	Intrinsically disordered protein are proteins that lack a fixed or ordered structure.
IDPRs	intrinsically disordered regions
IDPs	intrinsically disordered proteins
IMS-MS	Ion-mobility spectrometry–mass spectrometry
LC	Liquid chromatography
LCMS	Liquid chromatography–mass spectrometry
LCRs	low complexity region
Max	Myc associated factor X, can form homodimers or heterodimers with Myc or other proteins
MLN-like	A inactive-like conformer found with the crystallographic aurora in complex with MLN-8237 or CD532
MoRFs	molecular recognitions features
MSM	mono-link peptide-spectra match
MT	microtubules
NHS-ester	N-Hydroxysuccinimide Ester
ORDPs	ordered proteins
PBST	Phosphate buffered saline with Tween-20
PBST	phosphate-buffered saline with Tween 20
PDK1	3-phosphoinositide-dependent protein kinase-1

PELDOR Spectroscopy	Pulsed electron–electron double resonance (PELDOR, also frequently called DEER
Photo-Leu	Photo-Leucine, a photo-crosslinking amino acid, mimicking Leucine
Photo-Met	Photo- Methionine, a photo-crosslinking amino acid, mimicking Methionine
PIF	PDK1-interacting fragment
Pin1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1. It specific isomerizes phosphorylated Ser/Thr-Pro motifs
PLK1	polo-like kinase 1
Ponceau stain	diazo dye rapid reversible detection of protein bands on nitrocellulose
PPI	Protein-protein interaction
pS/pT	Phospho-(Ser/Thr)
pseudo b-factor	B factor, or temperature factor or Debye–Waller factor is a measure of the uncertainty for an atom's postiton in a crystallographic structure
PTM	Post-translational modification
pXL	Photo-crosslinker
RMSD	Root Mean Square Deviation
RMSF	Root mean square fluctuation
RPLC	Reverse Phase Liquid Chromatography
SASD	Solvent accessible surface distance, measurement of the distance of a crosslink respecting the surface and residue of a protein
SCF^{Fbw7}	A ubiquitin ligase that targets Myc-T58 phosphorylated for degradation. Skp1-Cdc53/Cullin-F-box-protein complex- F-box and WD repeat domain containing 7
SDS/PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling with amino acids in cell culture. A Quantitative proteomic methodology.
SLiMs	small linear motifs
SM	Small molecule
TAD	Transactivation domain is a transcription factor scaffold domain which contains binding sites for other proteins such as transcription coregulators
TFA	Trifluoroacetic acid
TFIIIC	Transcription factor IIIC seven member complex with two subunits; Tau-A and Tau-B. Essential for transcription of a subset of genes
Tris	tris(hydroxymethyl)aminomethane, buffer component
TRRAP	Transformation/transcription domain-associated protein. Adapter protein, which is found in various multiprotein chromatin complexes

1 Introduction

1.1 AGC kinases as drug targets in cancer

1.1.1 Contextualizing information

Aurora-A (AurA) is a protein kinase and a key regulator of mitosis.[1] It is activated by a phosphorylation cascade and locates to the centrosome during its role in the G_2 to M transition of the cell cycle.[2,3,4] It stabilises the oncogenic transcription factor Myc by disrupting ubiquitination and disconnecting Myc from growth factor-dependent control in neuroblastoma[5], such that Aurora-A overexpression increases expression of Myc.[6] Under non-oncogenic, regulated conditions, the interaction between AurA and Myc peaks during the S phase of the cell cycle.[7] Dysregulation of Myc or Aurora-A reinforces the oncogenic potential of either of the two proteins.[6] In cancers, such as neuroblastoma[8] or ovarian cancer [9], AurA is overexpressed and diffusely distributed in both the nucleus and cytoplasm, regardless of the cell-cycle phase.[10]

N-Myc is one of a family of three transcription factors (N-Myc, C-Myc and L-Myc). The Myc family are thought to regulate the transcription of one third of the genome.[11] Structurally, N-Myc presents as a classical disordered transcription factor[12], with an N-terminus with a TransActivation Domain (TAD), formed of highly conserved transcriptional regulation elements, a central region and a C-terminal DNA binding region.[13] Myc binds to DNA as an obligate-heterodimer, with a C-terminal basic helix–loop–helix leucine zipper (bHLHLZ) domain. The role of Myc in transcriptional regulation and active chromatin remodelling is mediated through recruitment of multiple proteins and complexes to the TAD and binding to DNA, as Myc/MAX heterodimers bind to the Enhancer box (E-box) elements in the promoters of target genes. The conserved regions in the TAD are highly promiscuous regions that are called Myc boxes; they mediate protein-protein interactions (PPIs) with hundreds of proteins, regulating chromatin remodelling, transcription and Myc stability, and are controlled by post-translational modifications (PTMs).[14] Myc has earned the moniker as a super controller of cancer as it regulates many aspects of oncogenesis.[15]

The mis- and de-regulation of kinases that lead to their over-activation is a known driver of oncogenesis. [16] This can be due to a somatic mutation, the generation of gene-fusions or gene amplification. For example, a single

amino acid substitution (L858R) in the receptor tyrosine kinase EGFR allows for the circumvention of its activation regulation and thus drives oncogenesis. [17] Its more common that mutations alter the copy number of a kinase, such as HER2 in breast cancer, or generate gene fusions, such as EML4-ALK in non-small cell lung cancer. These kinases are targets of anti-cancer studies.

A foundational event that marked the development of kinases as a target for clinical cancer therapy was the development and success of the tyrosine kinase inhibitor imatinib. Chronic myeloid leukaemia is commonly driven by the fusion of ABL kinase to breakpoint cluster region protein (BCR) due to chromosome rearrangement. Imatinib binds to the active site of this fusion kinase and inhibits kinase activity by stabilising the inactive conformation. [18]

The development of new treatments for kinase and kinase PPI driven cancer traditionally requires both target identification, showing the modulation kinase or kinase-complex PPI will successfully treat the diseases, and target determination, by understanding the molecular basis for that interaction. Consequently, structural biologists play a critical role in assisting the development of small molecules (SM) that target kinases by providing structural data that chemists can exploit for molecular design (Figure 1-1). [19]

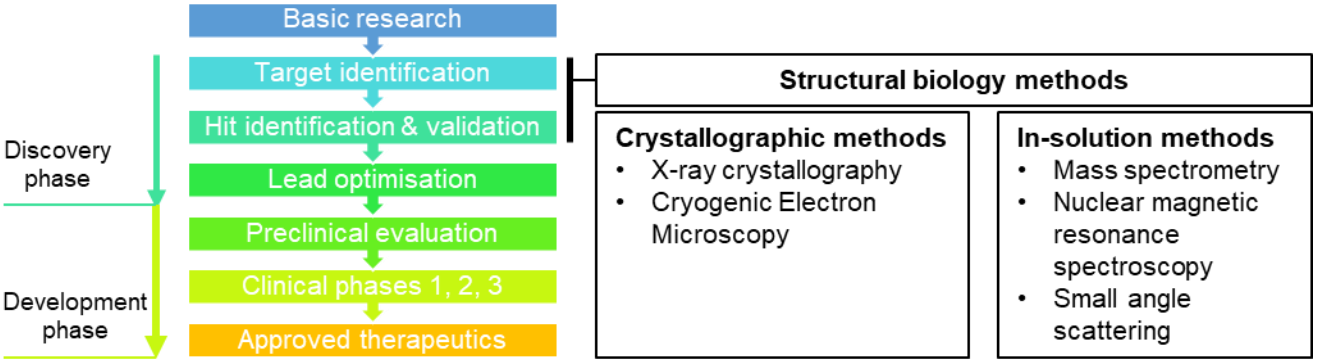


Figure 1-1 - Where structural biology fits into the anti-cancer drug development pipeline

Structural biology methods and their place within the discovery phase of drug development. The biophysical characterization of protein targets is a requirement to both generate and validate hit molecule engagement with the target and then assess the quality of modulation.

1.1.1.1 Phosphorylation as a mechanism of cell cycle control?

Phosphorylation regulates protein activity, both to control enzymatic turnover and more broadly defined *activity*, by regulating interactions with other proteins, manipulating the tertiary and quaternary structure of a protein, controlling subcellular distribution, or targeting proteins for degradation. [20]

1.1.1.2 Phosphorylation reaction – mechanism of phospho-transfer

PTMs are one of the key regulatory mechanisms used during cell signalling.[21] Phosphorylation, the covalent attachment of a phosphate group to the side chains of amino acid residues (most commonly Serine, Threonine or Tyrosine) provides a dynamic and reversible mechanism to regulate cell systems.[22 23] The addition of a phosphate to a protein can regulate a protein's activity or role in a cellular system by different molecular mechanisms. Commonly, phosphorylation can alter the tertiary structure of a protein by stabilising a particular protein conformer or ensemble of conformers.[24] This could be a short-range effect, for example a direct change to the helix, loop or sheet the phosphate was added, or long-range, allosteric effect where a protein's entire tertiary structure is changed. Alternatively, the PPIs that the phosphorylated protein can perform could differ from the unmodified protein, by the phosphate creating a new binding site for interactor proteins to bind to. These effects could work in combination, for example phosphorylation may stabilise a disordered loop so that it forms a helix, which could then bind an interactor. Kinases, facilitated by adaptor proteins, (Figure 1-2) catalyse the reaction that transfers a phosphate from adenosine triphosphate (ATP) to the substrate over a five-step process.[25] Figure 1-3 shows this multi-step process, which requires the kinase to be in an active conformation to allow a precise inter-molecular network of hydrogen bonding and hydrophobic interactions between the substrate, kinase, ATP and two magnesium ions. Disruption of residues that form contacts with the components of the reaction can leads to de-regulation of phosphorylation and both under or over-activity.

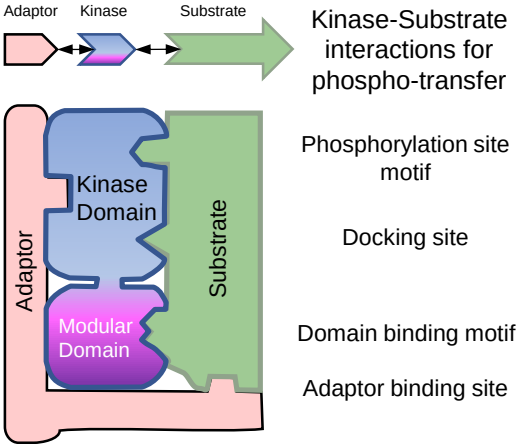


Figure 1-2 Interactions between a kinase, adaptor proteins and substrates

The kinase holoenzyme complex comprises inter-protein interactions local and distal to the phosphorylation site. The kinase domain (catalytic domain) participates in interactions with the c-lobe of the kinase and phosophorylation-site motif. Distal interactions are formed between the substrate and adaptor proteins or domains. Interactions between the catalytic and regulatory domains, and adaptor protein are required to form the active holoenzyme.

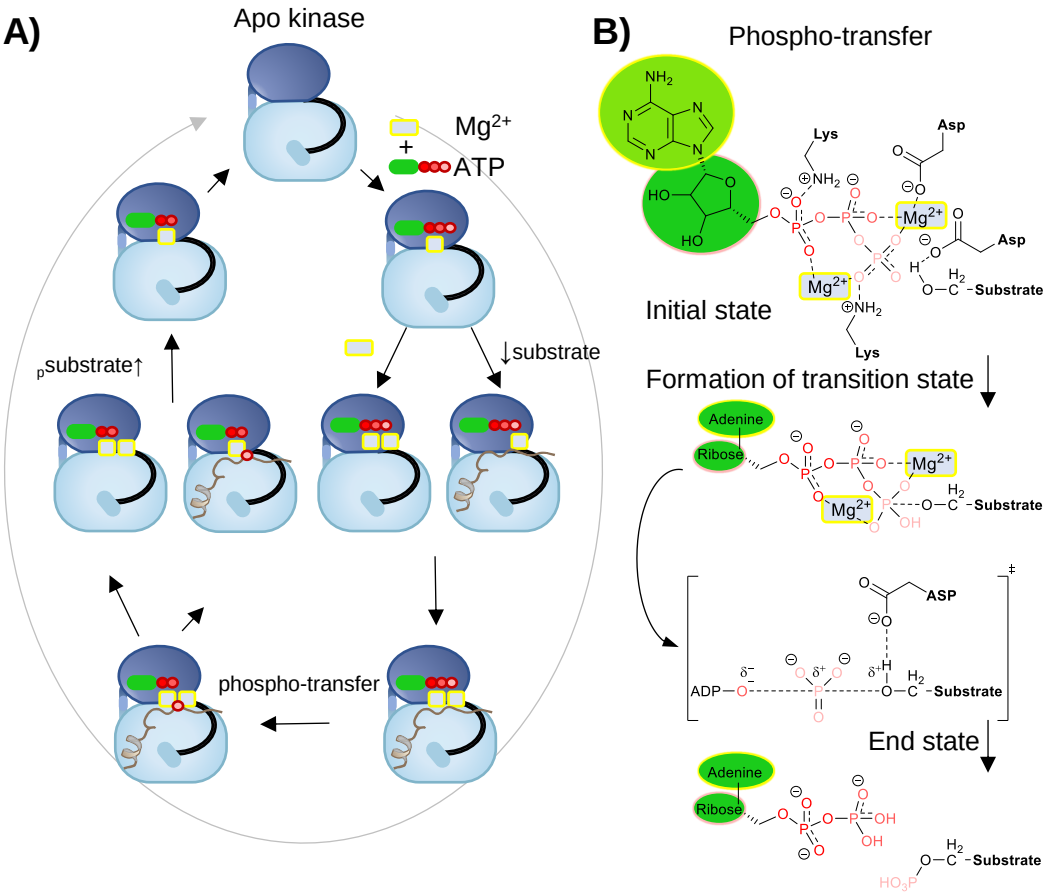


Figure 1-3 The kinase catalytic cycle

The phospho-transfer reaction mechanism. Panel A) shows the five step reaction involving the stepwise binding of and ATP+ Mg^{2+}, Mg^{2+} and the substrate. After the phospho-transfer and the generation of ADP, the substrate or Mg^{2+} will dissociate first dependent on the substrate binding strength. The ADP+Mg^{2+} dissociating gives the Apo kinase state re-generated. Panel B) shows the catalytic mechanism of phospho-transfer, the adenosine of the ATP is bound into a hydrophobic pocket. A series of charged sidechains and two Mg^{2+} will stabilise phosphates and Hydroxyl sidechains of the Thr/Ser of the substrate. The transition state is dissociative in character with Asp stabilizing Thr/Ser hydroxyl at the early stages of the reaction, which acts as a “proton trap” and serves as a proton donor for the release of the ADP.

The regulation of substrate binding to a kinase is vital to ensure signalling integrity is maintained. Figure 1-2 shows that these contacts can be local or distal to the site of phospho-transfer, involving domains of the kinase other than the catalytic domain or mediated between the kinase and substrate using adaptor proteins that form the active kinase complex.[26] The simplest kinase-substrate regulation is achieved by recognition motifs that typically consist of one to three residues directly up or downstream of the phosopho-site. These recognition motifs interact directly with kinase residues in the C-lobe and activation loop. These *local* or *proximal* recognition motifs can also be added to with docking sites on the kinase catalytic domain that involve

the binding of the substrate protein to the catalytic domain of the kinase, usually on the C-lobe. Distal recognition sites on the substrate protein may also interact with a kinase, often transiently, but instead of interacting with the catalytic domain, they interact with another domain of the kinase protein or to a scaffold protein that is part of the active kinase complex.[27] These can be contextual with subcellular localisation, liquid–liquid phase separation (LLPS) and substrate competition also allowing for the precise regulation of kinase activation and activity.[28 29]

1.1.2 What are the common structural features of AGC kinases?

1.1.2.1 Let's meet the family

The human kinome, ~2.5 % of the genome, is divided into families.[30 31] The AGC family consists of 63 evolutionarily related serine/threonine protein kinases, and it has recently been argued that the Aurora and PLK kinase families should be included in this grouping.[32] As kinase structural features and binding surfaces are critical to this thesis they will in introduced in detail in Section 1.3.3.1.

1.1.2.2 Common features of the core AGC kinase domain

Cyclic AMP-dependent protein kinase (cPKA) is the exemplar AGC kinase. It has been the subject of extensive biochemical and crystallographic experiments, and has served as the basis for the investigation of structure-function relationships and molecular links between activation, structure and catalysis in AGC kinases. It has both N-terminal and C-terminal extensions (NTE/CTE) from its catalytic kinase domain (which is formed of an N- and C-lobe) which undergo PTMs to form a catalysis-competent conformation. Many AGC kinases also have a PIF-pocket located on the N-Lobe,[33] so named for PDK1 Interacting Fragment (PIF) pocket discovered in protein kinase C-related kinase-2. AGC kinases are regulated through conformational control and regulating access to the PIF-pocket, which acts as a strict activity switch that has been the focus of PPI modulation.[34 35] An exemplary AGC kinase is shown in Figure 1-4. It shows the NTE, which is phosphorylated when active, and uses a conserved FXXW motif that anchors the NTE into a cleft formed between the C-helix and activation loop, stabilising the active conformation and protecting against de-phosphorylation.[36] The CTE features two conserved motifs: the hydrophobic motif (HM) and active-site tether (AST). The HM, also called the N-lobe tether (NLT) of the CTE binds into the PIF-pocket, formed by the αC-helix directly above the activation loop (Figure 1-4B). The AST uses a conserved

FDDY motif as part of the ATP binding pocket to form interactions with the nucleotide.[37] The AST is disordered in the absence of a bound nucleotide.[38]

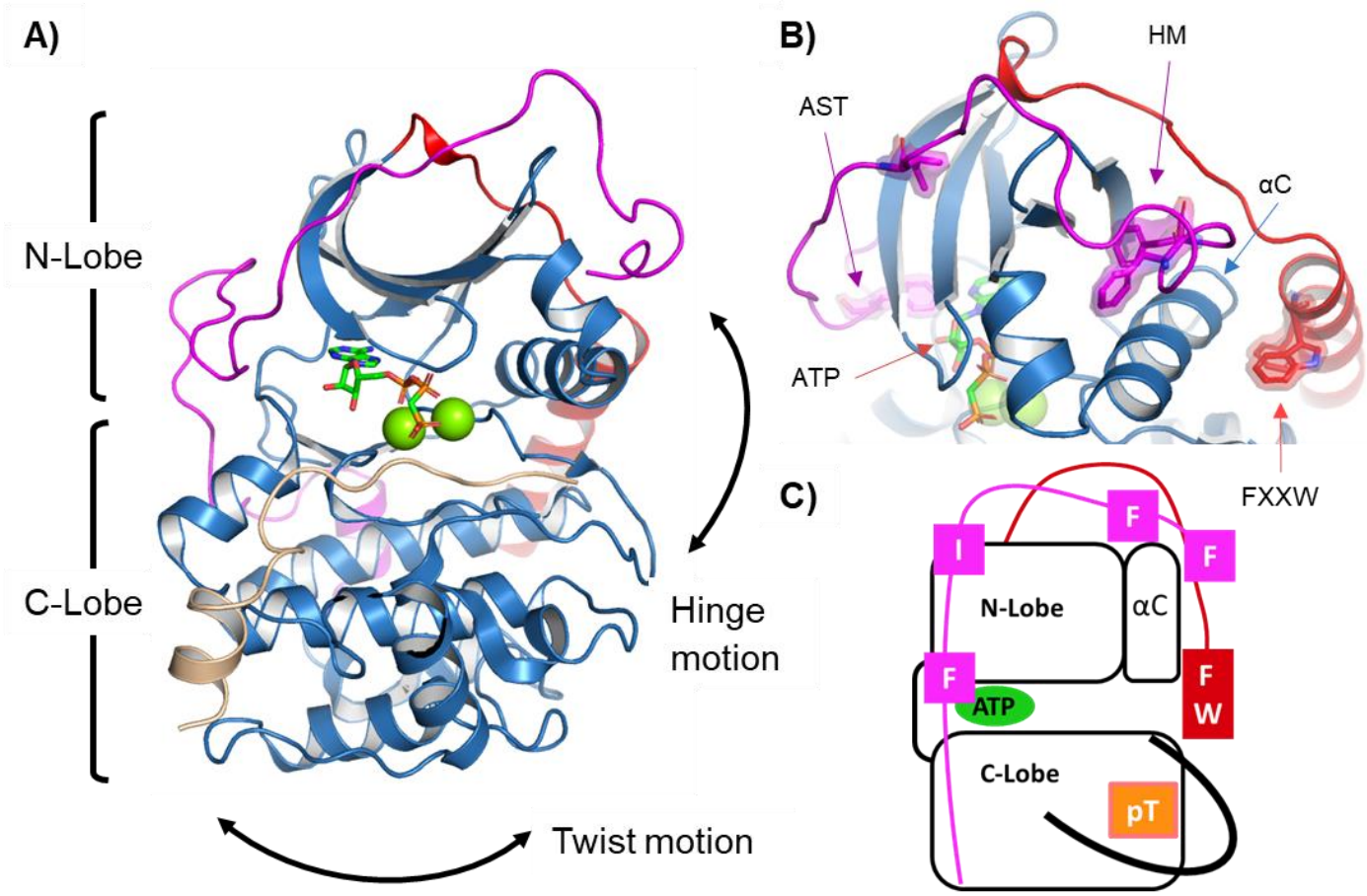


Figure 1-4 Structural features of AGC kinases

Features of AGC kinases annotated using cAMP-dependent protein kinase A (PBD: 1JBP). Panel A) shows the conserved features of kinases. The N- and C-lobe connected via the hinge. N-terminal extension (red), bound to the activation loop, using the FXXW motif. C-terminal extension (fuchsia) bound into the nucleotide pocket using the AST, onto the N-lobe crown using the turn motif and into the PIF-pocket with the HM. The substrate (wheat) is bound across the activation loop. The ATP and $2Mg^{2+}$ bound into the nucleotide pocket. Panel B) highlights the residues used by the NTE and the CTE. Additionally, the HM residues are placed directly on top of the αC-Helix Panel C) is a cartoon of the simplified AGC kinase structure. The fuchsia CTE shows the relative positioning of Phe of the AST, the Ile of the Turn motif and the two Phe residues of the HM. The red NTE places the FXXW motif into the cleft between the N-lobe and the activation loop.

1.1.2.3 Common residues of AGC kinases

A structured activation loop is stabilised by its phosphorylation, either through autophosphorylation or phosphorylation by another kinase.[39] The phosphoryl group coordinates to neighbouring basic residues that when absent increases its dynamics. This prevents the binding of the substrate for phosphorylation. Hence a structured activation loop, by phosphorylation, is a common indicator of kinase activity. The N-lobe nucleotide pocket is roofed by the glycine-rich loop that supports the binding of ATP. The Asp-Phe-Gly

(DFG) motif and its following two residues coordinate the two magnesium ions. The orientation of the DFG motif, relative to the catalytic lysine is one of a range of structural indictors commonly used to predict if a kinase conformation could catalyse the phosphorylation of a bound substrate (Figure 1-5). The ‘DFG-in’ conformation is the active conformation that is stabilised by hydrophobic interactions with ATP and coordinated magnesium ions.[40] In the ‘DFG-out’ conformation, the activation loop undergoes re-orientation, with the DFG positioned (*flipped*) towards the glycine-rich loop so that ATP and magnesium ions cannot be coordinated.

In active conformations the catalytic lysine forms a salt bridge with a Glu residue from the αC-helix. This electrostatic interaction places the sidechain of the lysine near to the phosphate groups of the bound ATP for phospho-transfer. The sequence of the His-Arg-Asp (HRD) motif, found on the loop directly under the DFG motif but part of the C-lobe, is less conserved overall than the DFG motif. During the phosphorylation reaction, the Asp of the DFG motif acts as a base, deprotonates the hydroxyl moiety of the residue to be phosphorylated and the His of the HRD motif interacts with the DFG motif, stabilising the active conformation.

An active kinase, as well as featuring a structured activation loop and the αC-helix, also usually contains a regulatory spine (R-spine). The aligned R-spine is formed from the His of the HRD, the Phe of the DFG, and the hydrophobic residues of the αC-helix and the N-lobe β-sheet. A misaligned R-Spine is another structural indicator of an inactive kinase, due to the incorrect configuration of residues critical to the phosphorylation reaction. The regulatory domains, if they are present, must also be in the correct conformation.[35 41] As its important to understand and categorise kinase domain conformation to predict potential activity, various groups have developed webservers for algorithmic categorisation methods.[42 43]

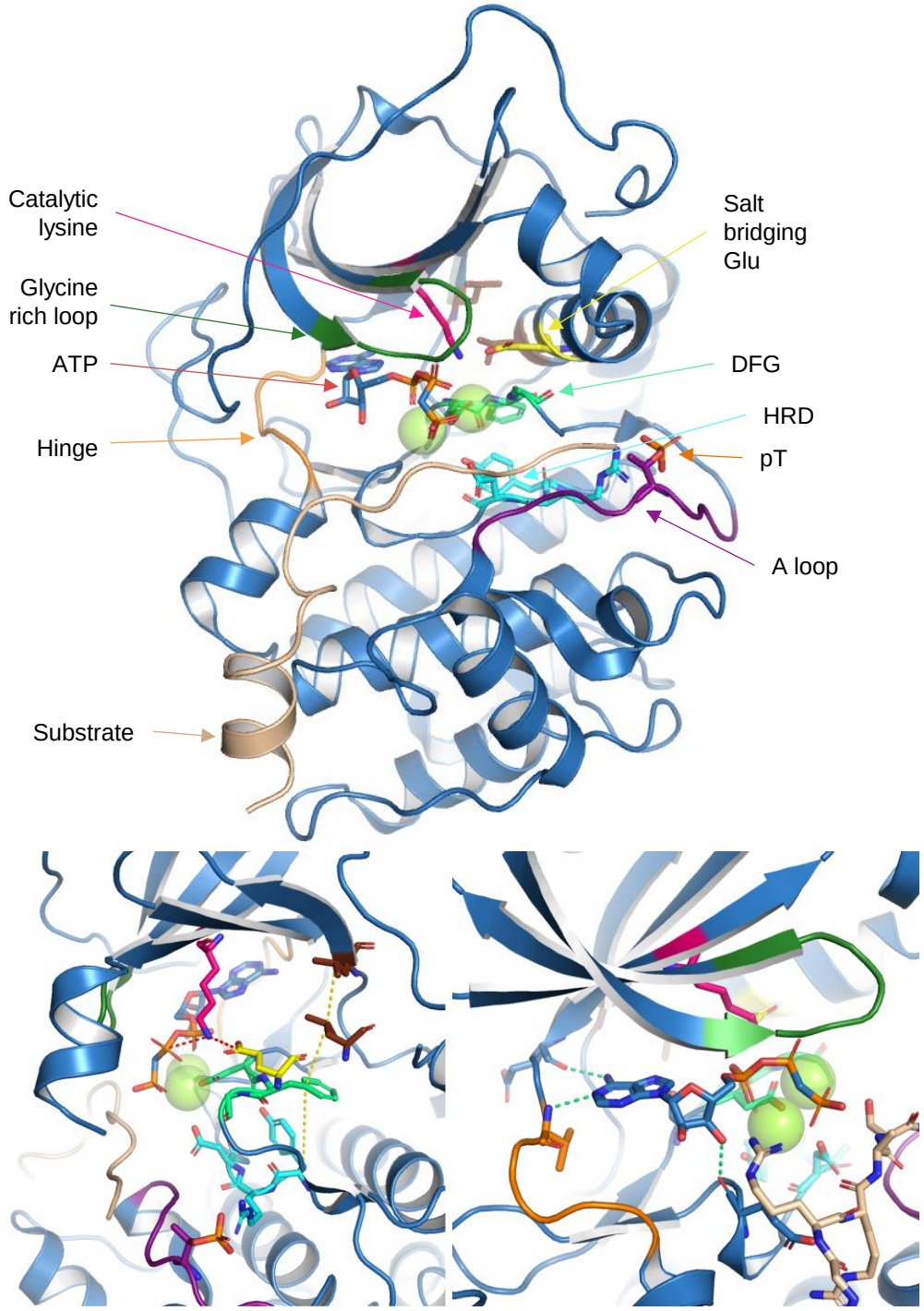


Figure 1-5 Structural features of an active AGC kinase

The visualization of key regions of conserved kinase architecture. Upper) PKA (blue) complexed with a substrate (wheat) (PBD: 1JBP). The hinge (orange) connects the N- and C-lobes. Glycine rich loop (green) roofs the nucleotide pocket with ATP and $2Mg^{2+}$ bound. The catalytic lysine (pink) forms a salt bridge with a Glu (yellow) on the αC-helix. The DFG motif (lime green) sits above the HRD motif (cyan, YRD in PKA). Behind this is the two other residues of the R-Spine in brown. Lower) magnified.

1.1.2.4 Spines of AGC kinases

Being able to visually inspect which kinase conformers can perform catalysis is useful to provide context for inhibitor/activator design for therapeutics,[44] whether these structures are generated from crystallographic methods, MD simulations or in-solution methods like NMR. This allows the comparison of structures for point mutants, or those with small molecules or interactors bound. Kornev *et al* used spatial alignment to show that certain residues in the catalytic domain of kinases, are highly spatially conserved (i.e. not sequence conserved) and their conformation and relative local interactions can be used to visually inspect structures for potential catalytic viably.[45] This is especially useful for studying activation by protein-protein interactions (PPI) and for drug design as it helps understand what changes to distal contacts affect the potential for active catalysis.[46] Figure 1-6 shows the active and inactive conformers with the R-spine (R for regulatory) and C-spine (C for catalytic) labelled. The R-spine being assembled means that the activation loop and αC-helix are structured, so the substrate can bind as the Arg of HRD is interacting with phosphate on activation loop. The C-spine consists of (top to bottom) the Ala G-loop, the hydrophobic residue G-loop and two hydrophobic residues from the C-lobe. This conformer can bind ATP with the assistance of the AST.

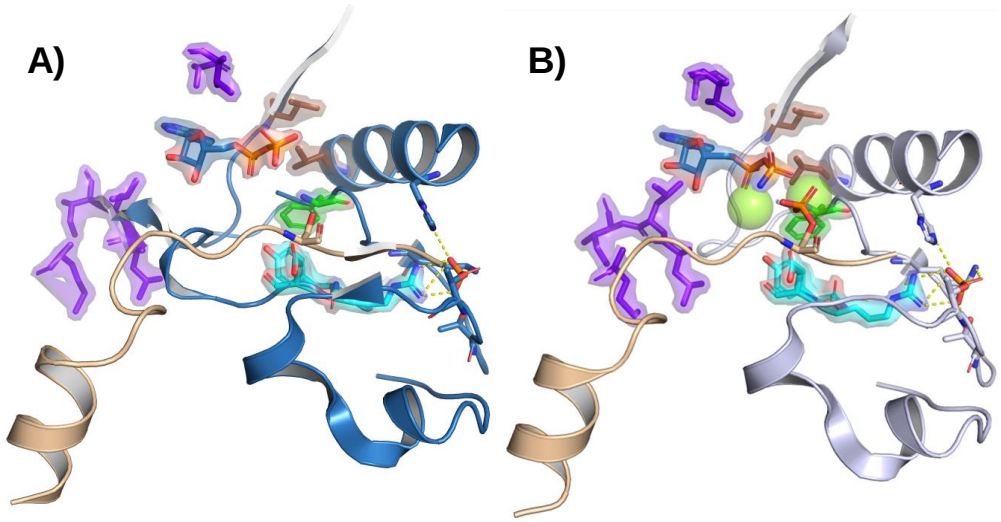


Figure 1-6 Positioning of catalytic residues for phospho-transfer

Panel A) shows highlighted residues of the C-spine (purple) in place allowing for the binding of ATP, the formation of the R-spine that permits, along with salt-bridge formation and the stabilized activation loop, the binding of the substrate. cAMP-dependent protein kinase A (PBD: 1JBP) with ADP and an inhibitor substrate. Panel B) shows the product of phospho-transfer, a substrate with phosphorylation. (PBD: 4HPT)

1.1.2.5 Inactive and active conformers of AGC kinases

Dodson *et al*, summarised the hallmarks for active mitotic kinases (including AGC kinases) as having the αC helix, the R-spine, and the activation loop structured and in the correct orientation for phospho-transfer.[47 48] This correctly positions the catalytic residues, substrate residues and ATP for phospho-transfer.[44] Deviation from this would reduce or eliminate catalytic turnover by preventing either the substrate or nucleotide from binding or by disrupting the transition steps of phospho-transfer. These hallmarks were refined into four structural indicators or tests to approximate if a conformer is an active kinase structure.[49] The formation and positioning of the R-spine (1) is joined by formation and positioning of the C-spine (2). The positioning of the αC helix is distilled into the formation of a salt bridge between the catalytic lysine and the Glu found in the αC helix (3). The final test remains the positioning and structure of the activation loop for binding of substrate (4). Figure 1-8 shows highlighted residues of an active catalytic domain and the product of phosphorylation. A simplified approach would use the position or orientation of the DFG motif as an identifier for kinase activity. For example, DFG-in conformations were all thought to be active, whereas DGF-out conformations were thought to be inactive. However, an ATP bound DFG-in kinase conformation that cannot bind substrate is unable to perform phosphorylation and therefore is inactive.

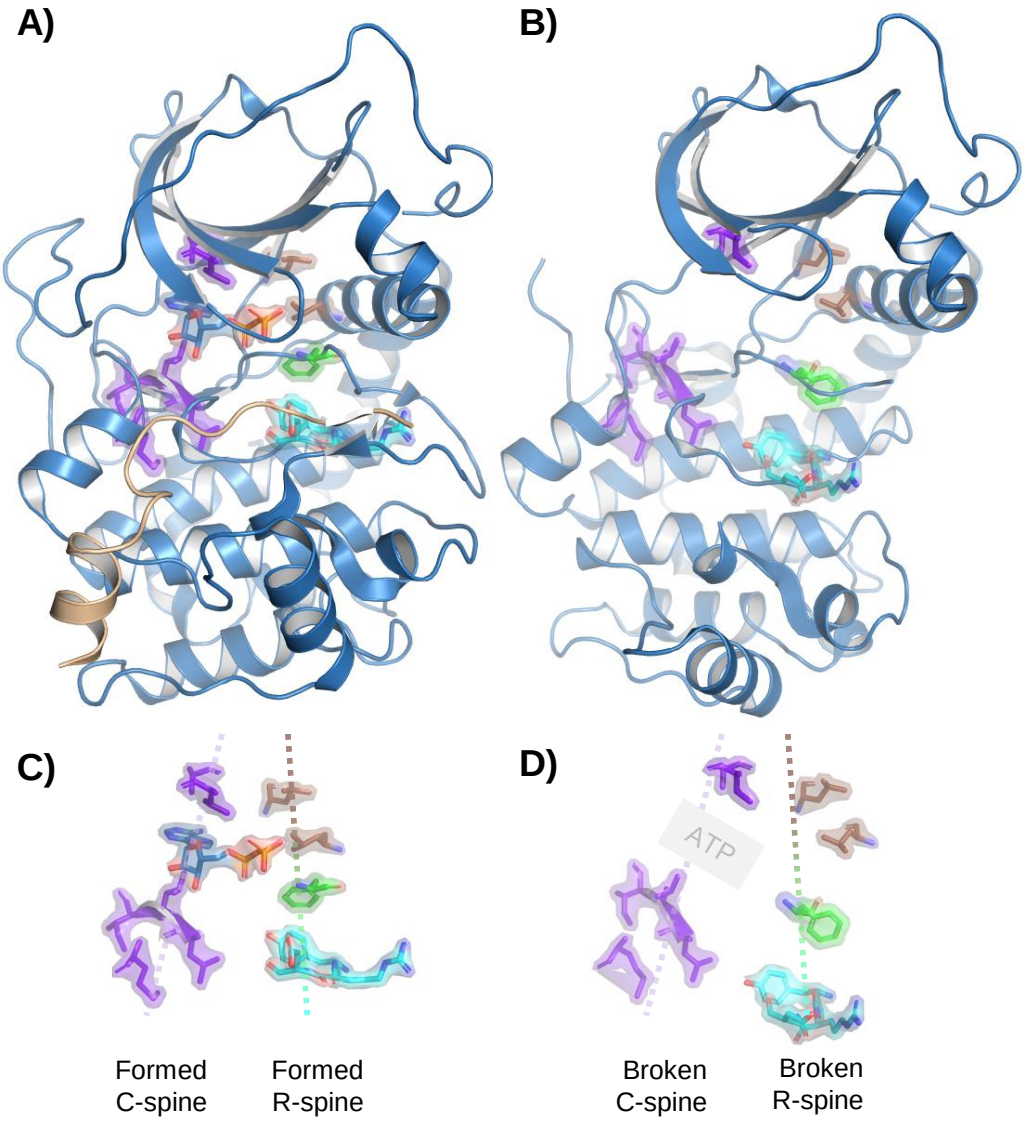


Figure 1-7 Active and inactive spines of AGC kinases

Active and inactive spines of AGC kinases annotated using cAMP-dependent protein kinase A. Panel A) shows the assembled C-spine in purpleblue formed from residues in the N-lobe, C-lobe and the nucleotide (PBD: 1JBP). Panel B) shows a broken C-spine with a missing nucleotide (PBD: 4DFY). The structure was obtained using R194A because destabilises interaction between the A Loop and the C-lobe. Parts of the CTE and activation loop are also unstructured. Panel C) has the relevant residues highlighted. Panel D) has the relevant residues highlighted from the inactive conformer. Note the nucleotide is missing and the R-spine no longer forms a straight line.

1.2 Disorder and dynamics within the context of kinase biology

1.2.1 Introduction to intrinsically disordered proteins

Proteins that lack a stable folded structure to perform their 'cellular tasks' are referred to as intrinsically disordered proteins (IDPs).[50] Babu *et al* argue that this paradigm is developed from the structure-function paradigm common to structural biology, namely that sequence leads to structure then structure to function.[56] This empowers hypotheses and predictions to be made about newly discovered proteins and their domains by analysis of their sequence and comparison with known sequence motifs. However, this paradigm fails for proteins that are somewhat (or completely) disordered but still have functional roles. Figure 1-9 shows the continuum of intrinsically disordered in proteins - from fully disordered to folded. A third of the proteome is predicted to consist of proteins that lack a stable fold but have functional roles within a diverse array of cellular systems.[51] IDPs and proteins with IDRs (intrinsically disordered region [of a protein]) are involved in many PPIs, with their sequence amino acid composition granting them more mobility that a folded protein, granting IDP(R)s the capacity to recognize multiple partners through distinct PPI modes.[52] In the presence of a binding partner or through addition/loss of a PTM, IDPs can gain 'contextual structure' to perform a cellular function.

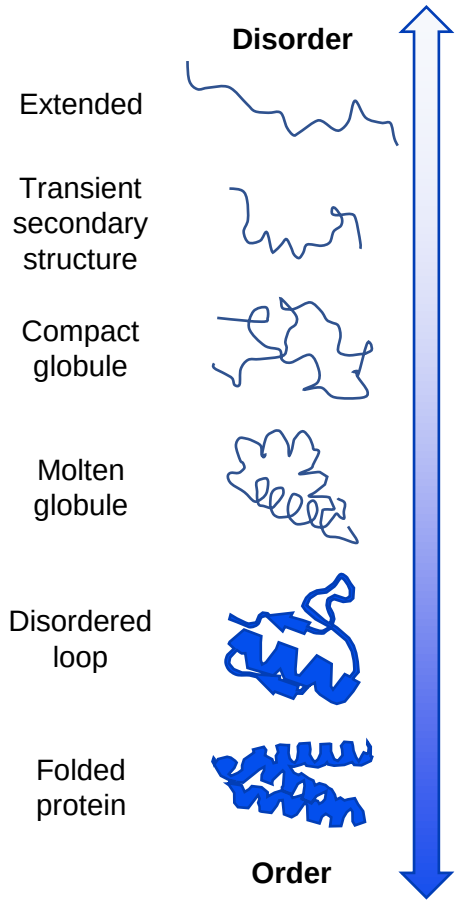


Figure 1-8 The order to disorder continuum of proteins

Disordered proteins and proteins with disordered domains exist in a continuum. These are classified in Table 1-1. Figure adapted from Classification of Intrinsically Disordered Regions and Proteins.

The energy landscapes of IDPs are flat, and as a result in solution they populate a continuous ensembles of structures, due to the low energy barriers of interconversion between these states. This is unlike dynamic proteins where there are much more defined, metastable conformations

within the energy landscape (Figure 1-10). Despite being structurally heterogeneous, IDPs facilitate PPIs, and this may be achieved by exploiting a conformational selection effect,[53] where the binding of an IDP to an interactor, small molecule or other biomolecule, will promote/favour binding competent conformations. Figure 1-10 shows the different energy landscapes of ORDPs (ordered proteins), IDPRs (intrinsically disordered protein regions) and IDPs.[54]

The paradigm of protein conformers in dynamic equilibria challenges the 'induced fit' model of PPI binding as the *functionally important conformer* of a protein may be a minor population of within a high energy well of the landscape, and not just the bulk low energy conformational heterogenous static conformer. An early example of this phenomenon is used to describe antibody/antigen binding.[53] For this thesis we maintain a conceptional division between dynamic proteins and disordered proteins whilst acknowledging that dynamic proteins may also contain IDRs.

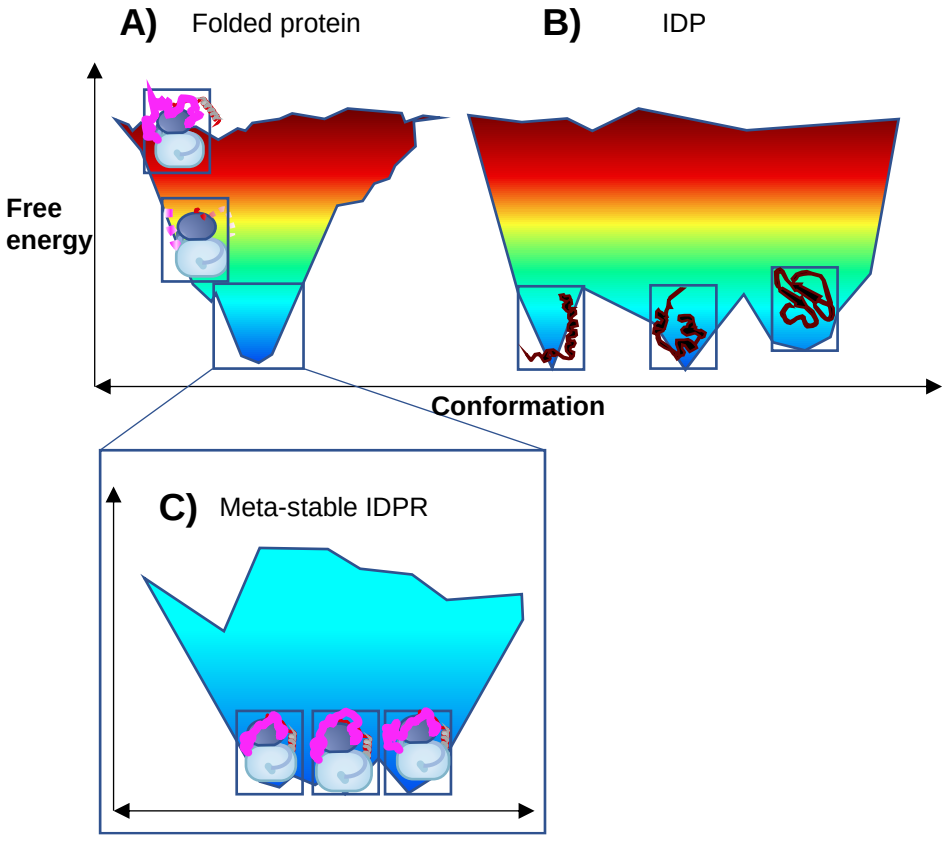


Figure 1-9 The energy landscape of Folded proteins

The energy landscape of Folded protein, IDP and IDPR. Panel A) shows the minimal free energy well that gives a protein with a single conformer. Panel B) shows an IDP with three low energy wells which can interconvert. Panel C) shows an IDPR protein with a single broad energy well that to be described as three inter-related conformers.

Deiana *et al* use nomenclature which lets us roughly bin proteins into three sub-classes, see Table 1-1.[55] They used an experimentally defined 30 % as the cut-off for disorder as Gsponer *et al* demonstrated that proteins become more susceptible to proteolytic degradation when more than 30% of the sequence consists of disordered residues.[56]

Table 1-1 Definitions of ordered and disordered proteins

Name	Acronym	Definition
Ordered proteins	ORDPs	<30 % of disordered residues in total. No terminal domains are >30 consecutive disordered residues. No internal domains with >40 consecutive disordered residues
Proteins with intrinsically disordered regions	IDPRs	<30 % of disordered residues in total. Feature either one terminal domain with >30 consecutive disordered residues or an internal domains with >40 consecutive disordered residues
Intrinsically disordered proteins	IDPs	They have more than 30% of disordered residues in the polypeptide chain

1.2.1.1 Introductory analysis of disordered domains

The primary sequence of an intrinsically disordered polypeptide, whether or not part of an IDR or IDP, may be classified into three types: short linear motifs (SLiMs), molecular recognition features (MoRFs), and low-complexity regions (LCRs) as defined in Table 1-2.[57] SLiMs mediate PPIs using their primary sequence and thus form extended disordered interfaces with folded proteins. MoRFs are commonly larger and form some structural feature upon binding that, when disrupted, ablates the PPI. Recently, LCRs have been investigated for their phase-separation propensities; for example, transcription factors use phase-separation to form local high-concentration interaction hubs using LCRs. However their position within a protein sequence can alter the role of the protein.[58] LCRs have been implicated in aiding structure formation. SLiMs serve two generalised roles: firstly, they are foci for PTMs and secondly they are involved in the formation of PPIs.[59] The lack of static fold allows easier recognition of their primary sequence by the catalytic cleft of an enzyme for post-translational modification. [59] Additional SLiMs are used to scaffold the formation of a complex with a protein, repeating the motif to increase avidity of the interaction. [59]

Recognising the type of disordered domain is important to structural biologists attempting to understand the molecular basis of an interaction between a protein or complex and an IDPRs or IDP. This is especially important when trying to replicate the minimum binding domains for a PPI in

vitro, as even if a sequence is unstructured in a crystallographic structure it maybe functionally important.

Table 1-2 Definitions of disordered domains

Name	Acronym	Definition
Short linear motifs	SLiMs	<10 amino acid residues long
Molecular recognition features	MoRFs	10–70 amino acids that are conserved but use structure to form specific interactions. The is commonly called a disorder-to-order transition upon molecular recognition
low-complexity regions	LCRs	<10 amino acid residues Homo-polymeric repeats of a single amino acid or hetero-polymeric short repeats of a few amino acids residues that have low sequence conservation and high structural flexibility

1.2.1.2 PPIs using disordered regions

Disordered domains can mediate several different interaction types which if not recognised can lead to incorrect hypotheses or incorrect estimation of the relative strength of the interaction. Olsen *et al* characterized the properties as potentially involving avidity, allovalency and fuzziness for a two-protein model.[60] An IDR can gain multi-domain like behaviour with the formation of a clamp or flanking binding sites in the PPI. Additionally, the formation of a complex can meditate a disorder to order transition (folding upon binding) of an IDP or IDPR. This is summarised in Table 1-3.

Table 1-3 Types of interactions involving disorder

Name	Definition
Avidity	Where two or more epitopes are found on the IDP and a corresponding number of binding sites on the globular protein will interact. This allows multiple weak sites of interaction to build to a strong PPI as one interaction is established, a second binding event is more likely to occur due to the proximity of the additional interaction site(s).
Allovalency	A single binding site on the globular protein can bind to several, identical epitopes along the IDP/IDR. When a bound site is released, the chance of rebinding is higher than anticipated from the IDP/IDR concentration alone, because of the very high epitope concentration close to the binding site on the receptor.
Fuzziness	A fuzzy complex is where both members of the PPI have a number of interaction sites along an extended surface or cleft. A similar concept to avidity but the multiple weak sites are proximal and an interaction site on the ligand is not restricted to connect with one specific interaction site on the receptor and vice versa. Consequently, the effect of high local concentration of both ligand and receptor binding sites leads to high likelihood of reforming the PPI than what one would anticipate from ligand concentration alone.

Figure 1-11 highlights how disorder complicates binding with both inter- and intra-molecular effects. Disordered regions within ORDPs/IDRPs can act as a linker between domains that can be modulated using PTMs or via complex formation. If the IDR linker undergoes a disorder to order transition it will

alter its effective length thereby changing effective local concentration of the inter-domain interaction. IDRs can be auto-inhibitory, the disordered region acting as a regulatory domain by forming a fuzzy complex that serves as an intramolecular competitive partner for the binding surface. PPIs can also involve a fuzzy region that increases the local concentration of a weak-affinity binding domain near the target or anchors it via transient interactions.[61]

These concepts highlight how the involvement of disorder in a PPI can dramatically increase the complexity and hence the need for orthogonal validation. The interactions can have complex binding thermodynamics and kinetics with highly concentration and context dependent phenomena.[62] Additionally attempting to simplify the interaction in vitro by, for example, generating minimum binding domain constructs for assays or crystallisation, may disrupt weak interaction sites critical to the overall strength of binding. IDRs require less input energy to alter their conformation as the energy difference between the conformers in the ensemble is small.[63]

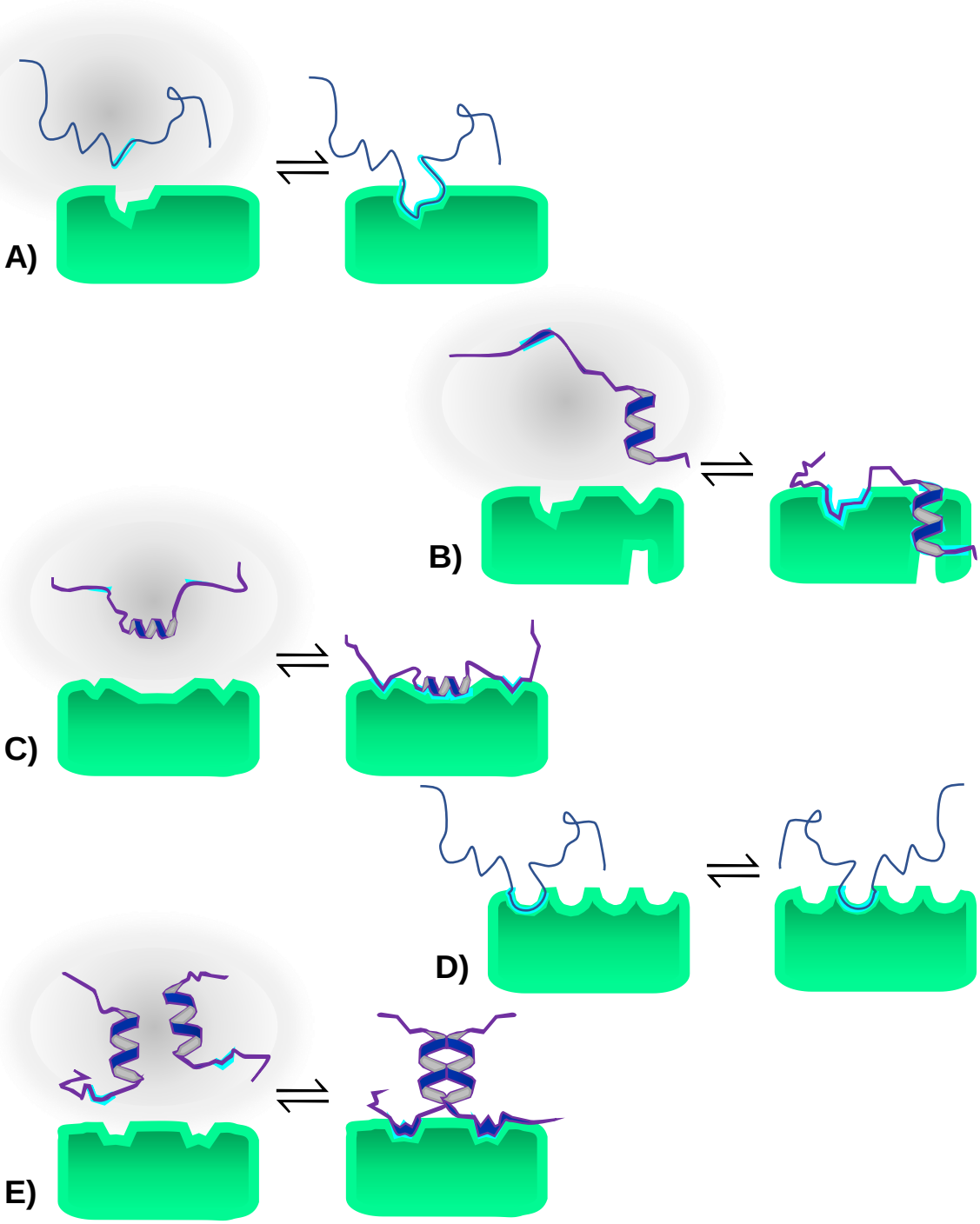


Figure 1-10 Modes of disordered interactions

Modes of Disordered interactions A) A fuzzy complex. The blue region of the IDP binds to a specific site on a binding partner. B) A tethered complex, with two binding site (blue) separated by a disordered linker C) An IDP binding to a globular protein bound in clamp mode D) IDP binding to a globular protein bound showing an allovalent PPI (multiple identical binding sites in one IDP) E) higher stoichiometry complex (1+2) with multiple interactions sites and modes, facilitated by protein oligomerisation.

1.2.2 Conformational dynamics

As previously mentioned, proteins can exist in an equilibrium of conformations that can be characterized by an energy landscape.[64] All soluble proteins in-solution will be subject to structural fluctuations that are determined by the temperature of the solvent, the energy of conformer and the energy barriers between conformers. Proteins can participate in a continuum of dynamic motions over a range of timescales (Figure 1-12).[65] These motions can as be described using statistical thermodynamics and quantitatively described computationally by all-atom molecular dynamics simulations.[66]

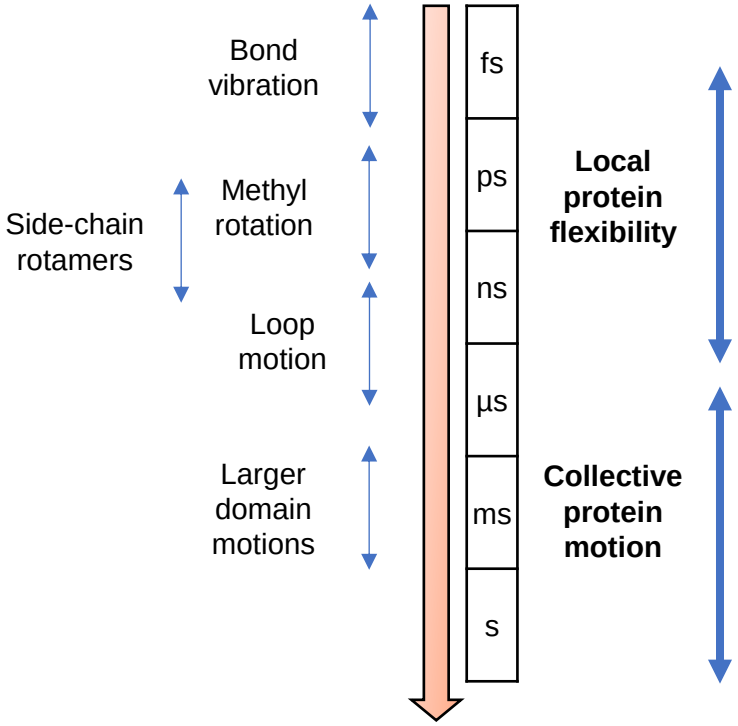


Figure 1-11 The time scales of dynamic processes

The time scales of dynamic processes in proteins and their relevance to conformational. Adapted from Wei *et al*, 2016.

Unlike structural dynamics borne of a protein lacking a stable fold, a protein can form a heterogeneous ensemble of conformations that are challenging to characterize but are in a temperature dependent distribution. This model of protein conformational dynamics allows us to interpret how the relative populations of the interconverting conformers will change following allosteric events. Allosteric events perturb the relative energy state of a conformer and thus shift the ensemble of the protein. An allosteric event could be the binding of a cofactor or the addition of a PTM. This is why a combination of structural biology approaches is important to elucidating the molecular basis of a PPI as a low population conformer could be the functionally relevant conformer.[67]

1.2.3 Disorder and dynamics in kinases

The catalytic and regulatory domains of kinases are ORDPs and IDPRs in addition to featuring disordered linkers. These disordered domains and regions can impart a fitness advantage to protein kinases by allowing the same domain to interact with diverse modulators and substrates. Removal of these structurally challenging elements may hamper efforts to investigate the full functionality of a kinase.[68]

The catalytic domains of kinases require dynamics to facilitate turnover of phospho-transfer, as a substate should not be so stable that significant energy is required to overcome the kinetic barrier, and the kinase cannot risk getting *stuck* in a part of the catalytic cycle.[69] Under the model of interconverting conformers, the population of active conformer is controlled by a stimulus, such as a PTM administered by another enzyme in a signalling cascade or binding of an interactor so that it allosterically stabilises the active conformer to prevent over-activation.

Structurally, kinases feature, often family specific, globular functional domains that administer catalytic activity and localization in addition to modulation of the activation loop. These can bind ligands, modulate PPIs, or undergo additional post-translational modification. Yet regulatory regions can also be present without any secondary structure or a fixed structure. Avidity and allovalency are important to allow for prompt activation and signalling cascades. For example, SH3 domains are common in certain families of kinases in multidomain PPI modules or cassettes. They form part of the regulatory mechanisms of these kinases where an auto-inhibited fuzzy complex forms between a phosphorylated SLiM on the interdomain linker of the SH3 domains and the rest of the module. Upon de-phosphorylation the interaction is lost, and the kinase domain released.[70] The number and organization of SH3 domains and inhibitory SLiMs affects the pathway of activation.

Commonly kinases can use an intrinsically disordered N- or C-terminal extensions that lack stable, single conformations but after the addition of a binding partner or PTM present as single static conformer (Figure 1-13). These domains often resist structural methods as they lack a single low-energy conformer that is present after crystallisation.[71]

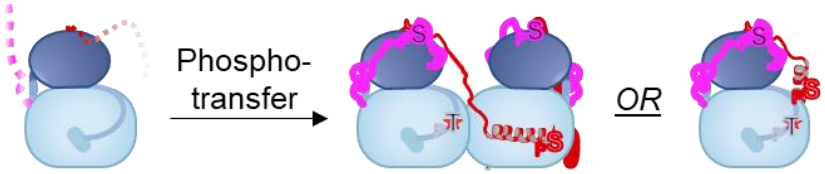


Figure 1-12 Kinase IDRs that regulate kinase activity undergo disorder to order transitions

An exemplary kinase with NTE (red) and CTE (pink) as an IDR which becomes ordered upon phospho-transfer. This could allow the formation of an active homodimer species where the CTE becomes a structured domain mediating the PPI. Also phosphorylation can also cause the disorder to order transition regulatory elements bound by a flexible linker.

Another regulatory mechanism of kinases that concerns structural dynamics is auto-phosphorylation. This is the activation of the kinase by the phosphorylation of the activation loop by a kinase itself. This occurs either where one molecule of kinase phosphorylates another (trans auto-phosphorylation), or the same molecule phosphorylates itself (cis auto-phosphorylation).[72] This raises a problem, as how can an inactive kinase molecule perform phosphorylation to generate the active kinase species capable of phosphorylation? The problem can be resolved in three ways. Firstly, the activation loop of the kinase contains the substrate consensus sequence of the kinase so that, despite very slow turnover an increase in local concentration of the kinase due to localisation can create enough transient dimer to allow the wholesale activation of the local kinase population. Secondly, the structure of the kinase allows for the formation of a transient dimer that is either symmetric or asymmetric. An asymmetric transient dimer utilizes the low-population *active-like* conformer to phosphorylate a molecule in a receiving conformer. Thirdly, a transient dimer can form that mediates cis auto-phosphorylation by acting as an allosteric activator. Finally, formation of a unique conformer can allow the phosphorylation of the activation loop of a kinase molecule by itself.[73] These mechanisms must be tightly regulated to prevent stochastic activation. This can be achieved using inhibitory domains that require PTMs or third-party proteins to remove them.

Hence dynamics in the formation of the holoenzyme and its activation and localisation will always be dependent on the precise control of the conformation ensemble and disordered regions for timely performance of its cellular role. [74] [75]

1.3 Aurora A kinase

1.3.1 Introduction to AurA – role in cells

1.3.1.1 History of AurA study

The Aurora kinase family consists of serine/threonine kinases essential for mitosis in all eukaryotes. It was originally named Breast Tumour Amplified Kinase (BTAK) due to its overexpression in breast cancer.[76] Aurora A kinase (AurA) is the most studied of the three family members. It is so named because mutations of the AurA homologue in Drosophila show incorrect mitotic spindle pole formation, and under fluorescence microscopy this resembles the northern lights.[77] The human family is composed of three members that show 70 % sequence conservation, where AURKA and AURKB are found in nearly every cell type in the body, while AURKC is limited to normal germ cells.[78] Increased levels of AurA have been reported in diverse types of cancers.[79] Intriguingly, for both neuroblastoma and triple negative breast cancer, AurA appears to regulate the ubiquitination of N-Myc in order to fulfil its pro-tumorigenic function in cancer cells. High levels of AurA increase the half-life of the N-Myc protein by protecting it from proteasome-dependent degradation.[80 81 82 83 84 85 86]

1.3.1.2 How is AurA regulated?

AurA lacking activation loop phosphorylation is minimally active and has been shown to be in an equilibrium of inactive and ‘pre-active’ conformers. The binding of an interactor, here using the exemplary targeting protein Xklp2 (TPX2), begins the activation cascade. Although the un-phosphorylated AurA+TPX2 complex is minimally active, this complex can undergo autophosphorylation to generate the fully active phosphorylated AurA+TPX2.[87] The precise structural conformations found in the activation pathway are unknown.[49] AurA regulation is complicated by different activation partners of AurA using diverging mechanisms, with these occurring in particular subcellular locations.[1] For example the binding protein aurora borealis (commonly known as Bora) activates un-phosphorylated AurA by substituting for the lack of auto-phosphorylation, and appears to trigger localisation of AurA to the nucleus and mitotic entry.[88] These alternate mechanism have been challenging to elucidate as they feature disordered interactors.[89] AurA also can be subject to redox promoted activation mediated through the Cys found in the activation loop.[90] In cancerous cells, this mechanisms can be mis-regulated to co-opt the role of AurA related to mitotic activity for the promotion of tumorigenesis.[82]

1.3.1.2.1 Localisation and degradation

As AurA has many roles which are mediated by many activators and phosphorylation of client substrates, precise control is needed over its localisation as well as timely removal. AurA doesn't feature any conserved, structured regulatory domains and the presence of a canonical nuclear localisation sequence is contested. Therefore substrate phosphorylation is PPI dependent.[80]

During the progression through mitosis, AurA is initially located to the centrosomes and then spreads on the spindle microtubules in metaphase. It is then moved to the central spindle in anaphase prior to degradation as its role in the cell cycle is finished. For its major role in mitotic spindle assembly, TPX2 binds to AurA and both promotes activation through auto-phosphorylation and regulates its localisation to the spindle. The N-terminal extension (NTE) of AurA targets it to the centrosomes in a microtubule (MT)-dependent manner and deletion of the NTE affects spindle formation and the progression of mitosis.[91 92 93 94]

Additionally, as the conserved substrate motifs of AurA and AurB are similar it has been demonstrated through chimeric proteins of AurA/B that their largely unstructured terminal extensions direct their respective cellular localizations through the selection of cofactor PPIs and hence target substrates and functions.[95]

Mitotic AurA degradation increases during Anaphase as it is degraded through a ubiquitin E3-ligase complex where hCDH1, also known as the Fizzy-related protein homolog, is an activator and substrate adaptor of the Anaphase Promoting Complex/cyclostome (APC/C).[96] hCDH recognizes SLiM degrons (known as KEN-, A- and D-boxes) present in AurA that target AurA for ubiquitination by APC/C and ultimately leads to timely proteasome-mediated degradation.[97 98] Dephosphorylation of Ser51, at the degron 'A-box' of AurA, by phosphatases plays a role in APC/C, regulating degradation.[99] Recently however, Abdelbaki *et al* demonstrated that the traditionally reported C-terminal D-box of AurA does not act as a degron and instead mediates essential structural features of the protein and that the N-terminal intrinsically disordered region of AurA containing the "A-box" is sufficient to confer FZR1-dependent mitotic degradation.[100] Giubettini *et al* presented evidence that TPX2 protects AurA from proteolytic degradation, keeping enough active AurA+TPX2 complex throughout mitosis and helping to control the cellular levels of the kinase.[101] Low levels of AurA proteins can

be found often enriched in the MT and sparsely in the cytoplasm of interphase cells.[102]

1.3.1.2.2 Substrates of AurA phosphorylation

Figure 1-14 shows the local sequence alignment of known substrates of AurA and the AGC kinase PKA$_{CA}$. This builds on the quantitative phospho-proteomics approach that gives strict motif enrichment of the amino acid sequences surrounding phosphorylated RR[L/X]p[S/T][L/X] and less strict motif as RRXp[S/T].[103] Modelling of selected AurA substrates shows distal binding interactions to the N-lobe of AurA.[104] This is not unexpected given the binding modes of other cofactors.

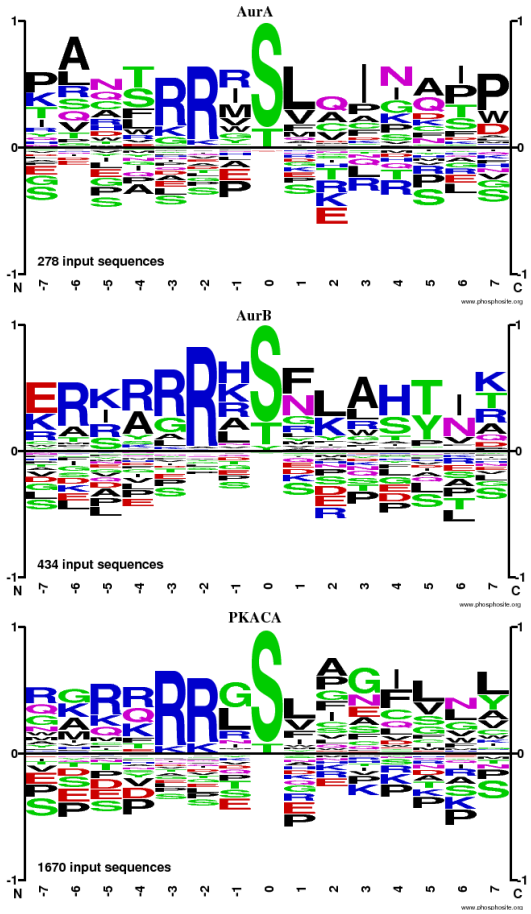


Figure 1-13 Comparison of phosphorylation sequences

Comparison of the phosphorylation sequences of the known substrates of human AurA, AurB and PKA from PhosphositePlusDatabase

1.3.1.3 Role of AurA during the cell cycle in healthy cells

1.3.1.3.1 Overview

The mammalian cell goes through the following stages in order to divide: interphase, synthesis phase, the mitotic phase and cytokinesis (inset to figure 1-15). The freshly divided daughter cell, in Gap 1 (G1) phase, begins duplicating its cellular contents except the chromosomes. In synthesis (S) phase the chromosomes are duplicated and in Gap 2 (G2) phase the cell undergoes further growth to prepare for entry into mitosis and must pass certain checkpoints before committing to mitosis. Mitosis is the part of the cell cycle where the replicated chromosomes are divided into two new nuclei and two daughter nuclei are formed. This is followed by cytokinesis, where the two daughter cells undergo cytoplasmic division and form separated cells. Figure 1-15 shows the localisation and activation of AurA during the different phases of the cell cycle.[105]

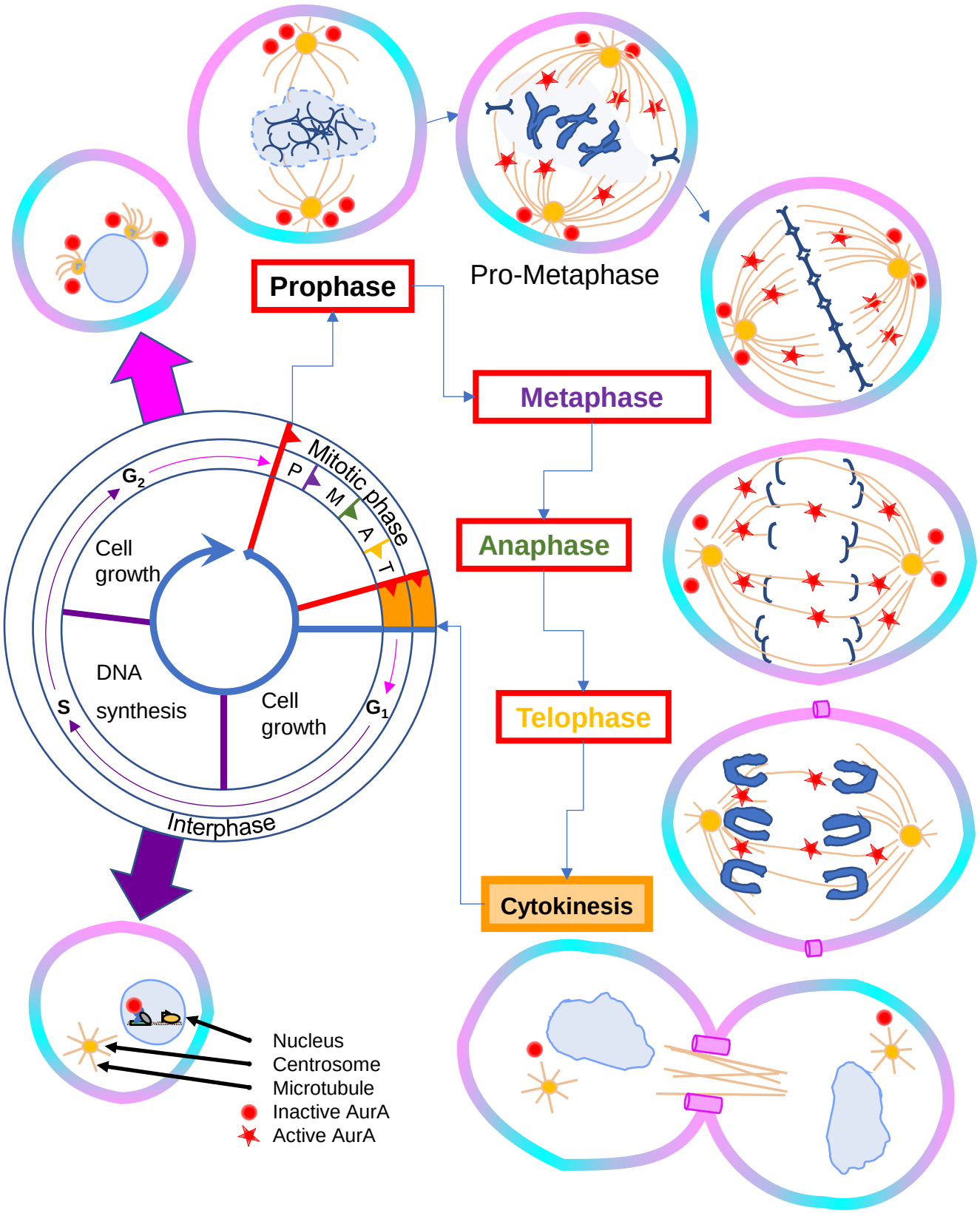


Figure 1-14 AurA in the cell cycle

During S phase AurA begins to amass at centrosomes and is activated on the progression of M phase. During the metaphase, activated AurA moves along the mitotic spindle to the midzone. Most of the population of AurA is inactivated and degraded before cytokinesis with only low levels detectable in early G1 cells.

The three human Aurora kinases share about ≈60 % sequence identity in their catalytic domain at the C terminus and follow the same domain organisation (Figure 1-16).[106] AurB and AurC appear to be able to phosphorylate similar substrates.[107] Localisation and binding partners modulate the substrate specificity of AurA compared with AurB/C.[108]

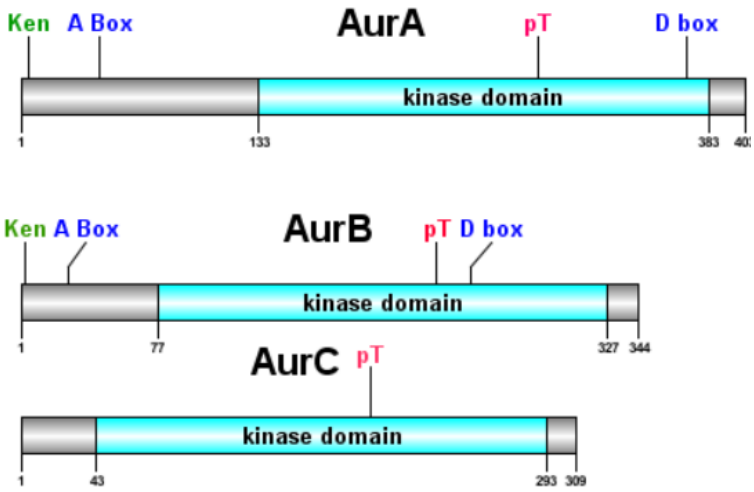


Figure 1-15 Domain layout of Aurora family kinases

Domain layout of Aurora kinases/Structure of AurA, -B, and -C showing the kinase domain in cyan and the NTE and CTE in silver. Ken (Lys-Glu-Asn) and A (QVRL) boxes are SLiMs for cell cycle-dependent proteolysis. D-box (Destruction Box) is likened to consensus motifs (RxxLxxx) in other cell cycle kinases but its validity as a true degron motif for AurA has been questioned.

Aurora kinase C (AurC) expression in mammals is largely localised to the germ cells and the kinase lacks Ken and A boxes in its N-terminal extensions. AurC−/− mice are viable but have impaired male fertility as spermiogenesis is disturbed.[109] AurC can rescue Aurora kinase B (AurB) function in AurB depleted cells, indicating that there is not a strict division of labour between AurB/C to carry out mitotic functions.

AurA and AurB both coordinate the progression of mitosis, but AurA manages the initial stages of mitosis and AurB late mitosis.[110] The distinct mitotic functions of AurA and AurB are partially regulated by their spatial regulation. AurA activity peaks during prometaphase and is involved in the initial, critical stages of mitosis, such as the maturation and separation of the centrosome and formation of bipolar spindle. The activity of AurB trails behind that of AurA, reaching a maximum from metaphase as AurB is a part of the chromosomal passenger complex (CPC) involving in the regulation of late mitosis, including the spindle-assembly checkpoint, and finally cytokinesis. AurA achieves its distinct and diverse roles within mitosis by pairing off with cofactors that act as AurA signalling modules.[111]

There is some evidence that AurA may continue to regulate the closing stages of mitosis.[112]

1.3.1.3.2 Interphase – Synthesis

AurA inhibits the ‘pause release’ of RNA polymerase II during S phase, stalling N-Myc driven transcriptional elongation.[113] AurA starts to localise on the duplicated centrosomes at the end of S phase and is highly enriched at the centrosomes during G2 phase.[114] However, regulatory phosphatases prevent the built up of active AurA, from autophosphorylation, in the cytoplasm and at spindle microtubules.[115] During S-phase AurA blocks the binding of the meditator complex preventing N-Myc mediated transcription. The AurA+N-Myc complex has therefore become a key therapeutic target in neuroblastoma.[116] This will be discussed further in Section 1.3.4. Additionally, it is speculated that the interaction of AurA with RNA-binding proteins mediates its non-catalytic functions in S-phase.

1.3.1.3.3 Interphase Gap 2

Active CyclinA/B-Cdk1 kinase triggers the G2-M transition by phosphorylating Bora, which then promotes AurA auto-phosphorylation and activation of Plk1.[88 117] At G2/M, AurA localizes in the centrosome and also contributes to centrosomal maturation before mitotic entry. In the late G2 phase AurA interacts with and is activated by Ajuba in a feedforward loop as Ajuba is both a substrate and activator for AurA. This PPI promotes mitotic commitment.[118]

1.3.1.3.4 Mitotic phase – Prophase

At the centrosomes, centrosomal protein of 192 kDa (CEP192) recruits and activates AurA to initiate a multistep signalling cascade that underlies the maturation and separation of the centrosome and MT nucleation. The assembly of the CPC is progressed by the phosphorylation of Haspin by AurA. During the early stages of mitosis, an increase of Guanosine triphosphate (GTP) activates the transporter protein Ras-related Nuclear protein (Ran). This releases TPX2 from the inhibitory complex with importins.[119] The GTP-bound form of Ran also imports and releases crucial spindle assembly factors into the nucleus that AurA requires in mitosis such as TPX2 and nuclear mitotic apparatus protein (NuMA).[120] The nuclear envelope breakdown (NEBD) is completed by the end of prophase after which AurA is released throughout the cytoplasm. Then, AurA mainly concentrates in centrosomes and in the nearby mitotic spindles. It is only after NEBD that TPX2 can reach the centrosome to bind AurA, to activate and re-localize AurA protein on microtubules at the spindle pole.[121]

1.3.1.3.5 Mitotic phase – Prometaphase and Metaphase

Re-localisation of AurA is coupled to activation by TPX2 through auto-phosphorylation of Thr288. This dramatically increases the catalytic activity of AurA. Recently, NuMA has been shown to be in direct interaction with the AurA/TPX2 complex for NuMA-dependent spindle orientation at spindle MTs.[122] The AurA-mediated phosphorylation of NuMA allows a sub-population of NuMA to reach the cortical region above the spindle poles. Additionally, the AurA+TPX2 complex phosphorylates TACC3.[123] P-TACC3 enhances MT polymerase activity of chTOG both near centrosomes and chromatin for its role in spindle formation, counteracting depolymerization. As kinetochore MTs become established, TACC3 recruits clathrin to these MTs. Long-lived kinetochore MTs allow local activation of AurA by TACC3 and phosphorylation at S558 stabilises the TACC3-clathrin complex, which in turn crosslinks these MTs. At metaphase, AurA moves to the proximal MT and phosphorylates MT-associated proteins substrates to organize the mitotic spindle.

1.3.1.3.6 Mitotic phase – Anaphase and Telophase

Anaphase begins with the formation of sister chromatids from the cleavage of chromosomes and AurA is targeted for proteasomal degradation via its association with APC/C.[124] AurA's role in mitosis thus ends and it is found in low abundance with the onset of telophase as the nuclear envelope re-forms with the formation of the contact ring.

1.3.1.3.7 Cytokinesis

During the formation of the daughter cells by fission the activity of AurA is lowest.[125] AurA and TACC3 have been shown to be active and interact on the central spindle but the function and substrates of AurA at this time are unclear.[123]

1.3.1.3.8 Interphase Gap 1

Venkitaraman *et al* have shown that AurA is needed in the G1-S phase transition but it appears to be a non-catalytic function.[126] Also, there is some evidence indicating that the reabsorption of the primary cilium during G1 phase is controlled by AurA.[127]

1.3.1.4 The role of AurA in cancer

Increased AurA expression levels have been found in a range of cancers, including breast cancer, but arguably the most detailed mechanistic studies have been done in neuroblastoma.[128] High AurA levels have been correlated

with highly proliferative cancers with drug-resistance and tumour metastasis formation.[129] [130] These high expression levels are commonly associated with the amplification of the chromosomal region that bears AurA 20q13.[76] Other mechanisms of unregulated AurA activity include gene mutation,[131] microRNA regulation,[132] transcriptional and/or PTM de-regulation.[133]

A review of the Cancer Genome Atlas (TCGA) UALCAN database by Naso *et al.* found significantly higher expression in cancer tissues than in normal control tissues for multiple tumour types.[80] This is concerning for clinicians as over-expression of AurA can increase genomic and chromosomal instability and enhance proliferation which leads to tumorigenesis.[134] Despite the continued effort to develop anti-cancer therapies that exploit AurA expression there has yet to be any small molecule shown to improve on current treatments other than potentially in cancers with combined AurA/N-Myc overexpression. However, this may be related to the questioned utility of mitotic kinase inhibitors as treatments.[135] However, exploiting a synthetic lethality of AurA in concert with other validated tumour suppressor-oriented treatments, as has been shown with BRCA-PARP axis, could bring successful approaches in the coming years.[136] For example, BRCA2-deficient cancer cells have shown enhanced sensitivity to de-activation of the AurA+TPX2 complex.[137]

The non-mitotic role of AurA driven cancers has been most studied in the case of AurA and the transcription factor N-Myc, as AurA has the ability to protect N-Myc from degradation independent of its kinase activity.[116]

Oncogenic AurA mutations have not been the primary focus of research and development when compared to amplification. However, the review of the COSMIC database by Naso *et al.* found several cancer-associated mutations and truncations linked to the deregulation of AurA. The mutation S155R reduces the binding of TPX2 and thus disrupts the progression of the mitotic spindle by AurA+TPX2 complex.[138] The AurA+TACC3 complex is also disrupted by the G136 point mutant.[123]

1.3.2 The structural features of AurA and its common interactors

1.3.2.1 AurA as a 'broken' AGC kinase

The Aurora family of kinases form their own branch of the kinase tree. However, Levinson *et al,* building on work by Leroux *et al*, argued for the inclusion of AurA along with Polo-like kinases into the AGC kinase family as *broken* AGC kinases for two notable reasons.[139 140] First, Aurora kinases lack the true NTE/CTE which serve as sites for phosphorylation in other kinases. Phosphorylation of NTE/CTE typically would result in docking of these disordered regions onto the catalytic domain, resulting in active kinase conformations. However, for AurA, its interactors perform this role, and as a result it can be thought of a kinase in which the catalytic and regulatory elements are "broken" apart. Second, residue sequence analysis of the activation loop of AurA shows that it lacks key *latch* residues, and hence the activation loop is also "broken". Furthermore, the APE motif is found as PPE, which is unique to AurA. This gives us the expectation that AurA will be more dynamic than closely related "complete" AGC kinases.

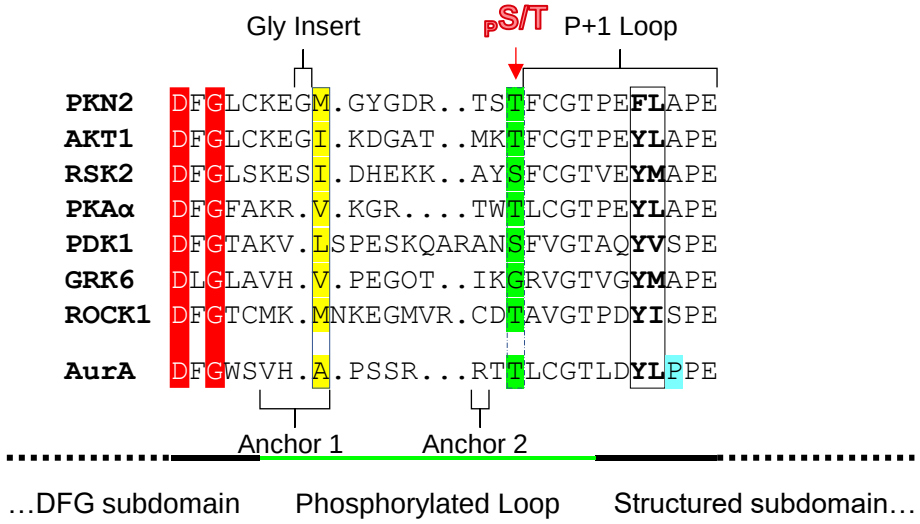


Figure 1-16 Sequence alignment of AurA and select kinases

Sequence alignment of the activation loops of AurA and members of eight AGC kinase subfamilies. Highlighted are the upper DFG triad (red), glycine insert, site phosopho-site (green) and P+1 Loop. Lower, Anchor point 1 (yellow) with AurA lacking hydrophobic latch residue. Anchor point 2 and the conserved hydrophobic residue pair are highlighted. Additionally, the mutated residue of the APE motif is highlighted in cyan. Adapted from Levinson *et al*, 2018.

Figure 1-17 shows the alignment used by Levinson *et al* to compare the activation loops of exemplary AGC kinases, manually curated using X-ray structures, and showing that anchor 1 is absent in AurA.[139] Additionally, the

P+1 loops are strongly conserved across AGC kinases with the exception of the Ala of the APE motif.

1.3.2.2 Naming of the pockets of AurA and surfaces

As AurA can be classified as an AGC kinase, its interactors and the binding pockets can be compared and analysed against other familial, conserved binding interactions (Figure 1-18). TPX2 and Transforming Acidic Coiled-Coil Containing Protein 3 (TACC3) have both been used extensively *in vitro* as model interactors to study AurA activation, PPIs and catalytic activity, using a suite of molecular biology techniques. Figure 1-19 relates the structures of AurA-TPX2 and AurA-TACC3 complexes to hPKA and provides a nomenclature for AurA pockets. The complex of AurA+TPX2 has been described previously as an oncogenic holoenzyme.[86] The complex of TPX2 and AurA is structurally comparable to conserved AGC kinases with TPX2 performing the role of the NTE and CTE extensions.[87]

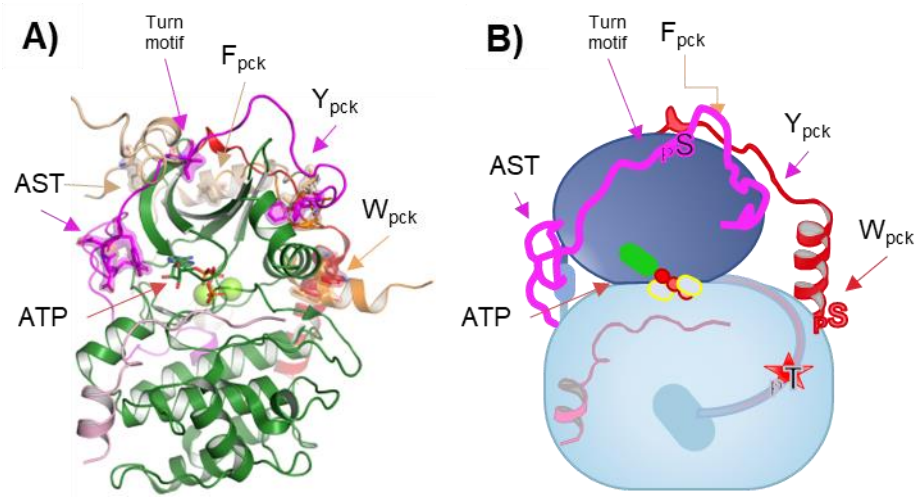


Figure 1-17 Pockets of AurA

Naming of the pockets of AurA and their relationship to hPKA (PBD: 4WB5). Panel A) shows the names of the pockets for AurA binding sites where F_{pck} is the F-pocket, W_{pck} is the W-pocket, AST is the active site tether, and Y_{pck} is the Y-pocket. Panel B) is a simplified cartoon of the relative locations of these pockets on hPKA. The N-terminal extension (red), bound to activation loop, using the FXXW motif but on AurA this is the W-pocket. C-terminal extension (fuchsia) bound into the nucleotide pocket using the AST, onto the N-lobe crown at the turn motif and into the PIF-pocket on AurA this is the Y-pocket.

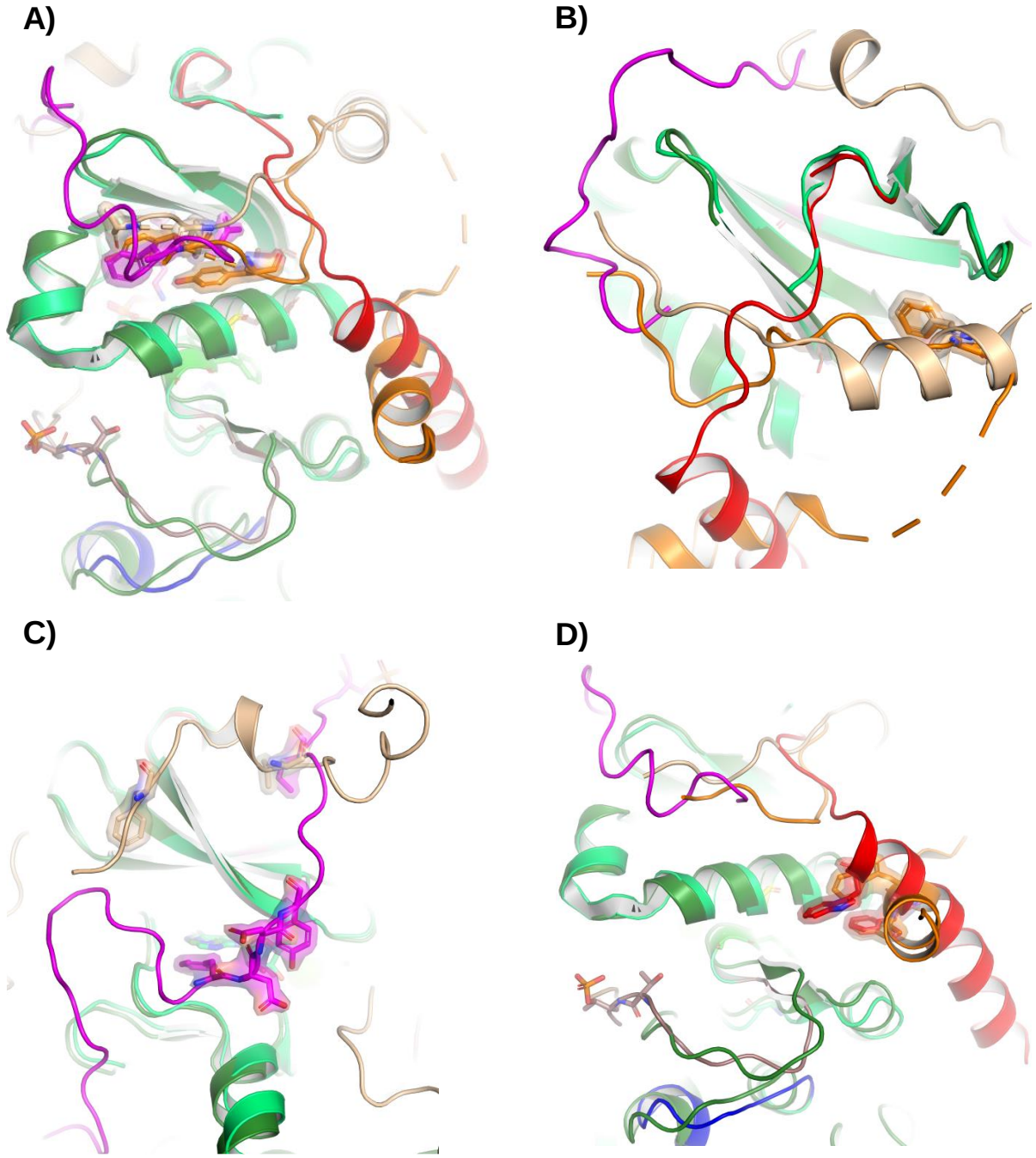


Figure 1-18 Naming pockets of AurA

Naming pockets of AurA and their relationship to hPKA (PBD: 4WB5). Panel A) shows the Y-pocket with the highlighted residues of Tacc3 (wheat; PBD 1OL5) and TPX2 (orange; PBD 5ODT). This is the PIF pocket on PKA. Panel B) shows the F-pocket with the highlighted residues of Tacc3 (wheat; PBD 1OL5) and TPX2 (orange; PBD 5ODT). There is no similar pocket on PKA. Panel C) shows the AST and the turn motif with the highlighted residues of TACC3 (wheat; PBD 1OL5) and hPKA (PBD: 4WB5). Tacc3 doesn't direct into the nucleotide pocket. Panel D) shows the W-pocket with the highlighted residues of TPX2 (orange; PBD 5ODT) and hPKA (PBD: 4WB5).

1.3.2.3 Structural basis for the binding of activators of AurA

1.3.2.3.1 Introduction

To investigate the molecular basis of novel PPIs involving AurA it is critical to understand what binding surfaces and clefts are used by known interactors and how their modes of binding relate to conserved, familial AGC kinase-like interactions. However, as these PPIs are dynamic and involve disordered interactors, crystallographic structures may not reveal the entire interaction, (Figure 1-21).

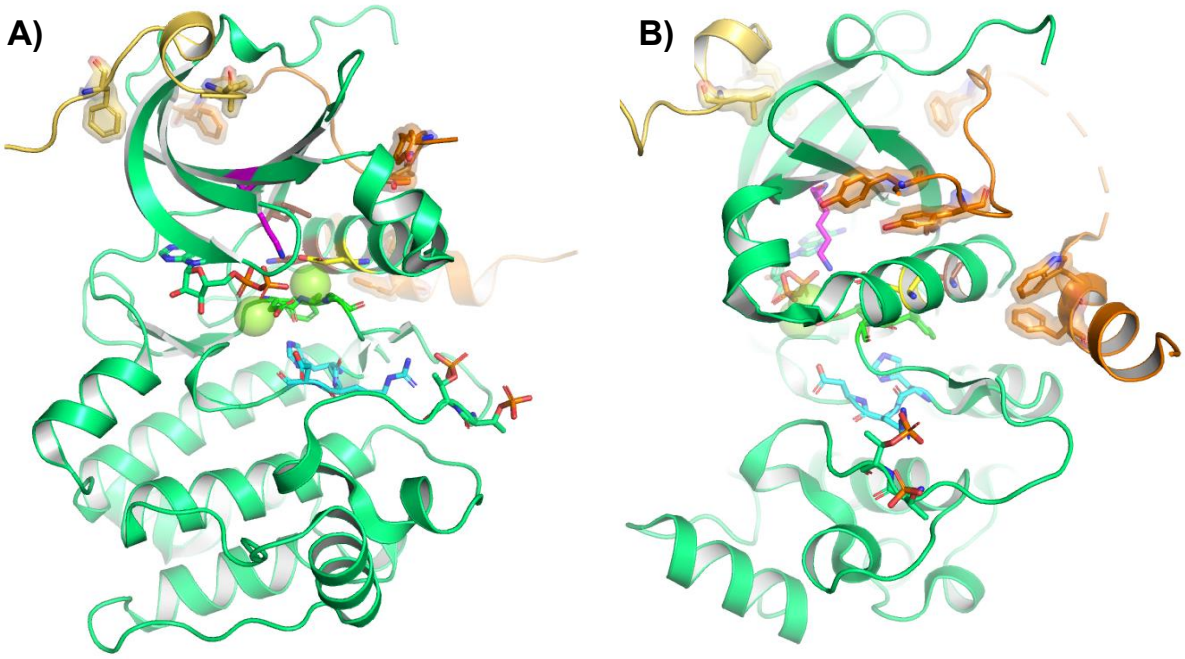


Figure 1-19 Pockets of AurA made to show the interaction of Tacc3 and TPX2

Composite of TACC3 (wheat) and TPX2 (orange) bound to AurA KD demonstrating pockets that AurA can utilize for binding (PBD 1OL5, 5ODT). A) Shows key residues of TACC3 interaction: TACC3 Leu532 bound as a Turn motif into N-lobe of AurA and Phe535 binds as an AST. B) TPX2 Tyr8 and 10 binds in the pocket formed from Helix B and C (aka αC helix) known as Y pocket. TPX2 Trp34 binds into a pocket formed from helix C and activation loop; W pocket. TPX2 Phe19 binds into the F-pocket formed by beta sheet 4 and 2 the AurA N-lobe.

1.3.2.3.2 Alignment of known interactors

As the pockets of AurA are related to those found in AGC kinases, interactors, and especially activators of AurA, have conserved sequences to bind to those pockets. Tavernier *et al* demonstrated using alignment to TPX2, that Bora uses conserved residues to bind to AurA using pockets also used by TPX2.[88] These pockets have been secondarily confirmed with crystallographic structures of AurA with Monobody and vNar binders.[141 142] Figure 1-22 shows these experimentally validated, conserved sequence alignments and their relationships to hPKA.

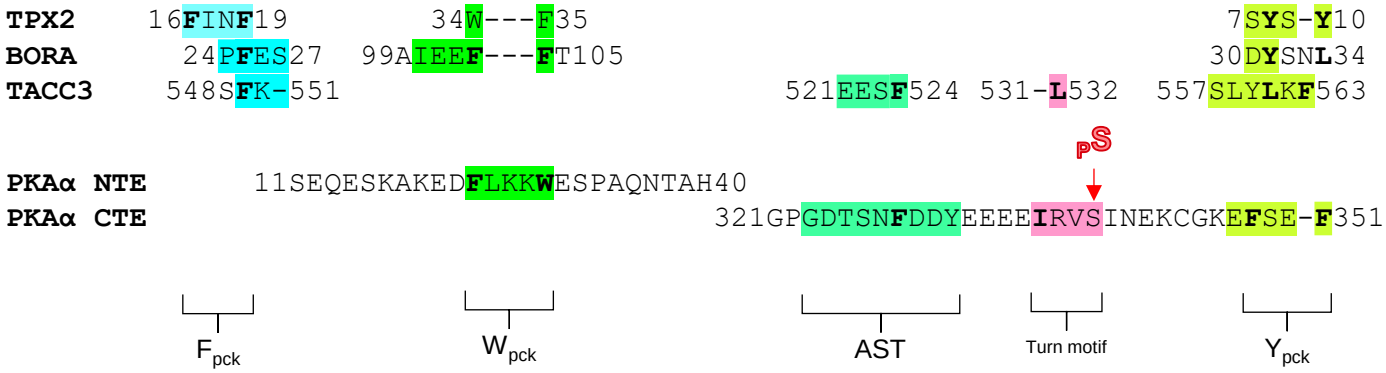


Figure 1-20 Alignment AurA of selected activator interactors

The alignment AurA of selected activator interactors against the relevant Human cAMP-dependent protein kinase catalytic subunit alpha terminal extensions. F_{pck} is the F-pocket, W_{pck} is the W-pocket, AST is the active site tether, and Y_{pck} is the Y-pocket. The Ser338 phospho-site is critical to forming the active enzyme conformation. The formal multiple protein sequence alignment can be found in the appendix section 5.1.

1.3.2.3.3 Key AurA cell cycle interactors

Different PPIs use diverse sequences and binding modes despite using a similar combination of AurA pockets. Figure 1-23 shows these different mechanisms that occur in different parts of the cell cycle.

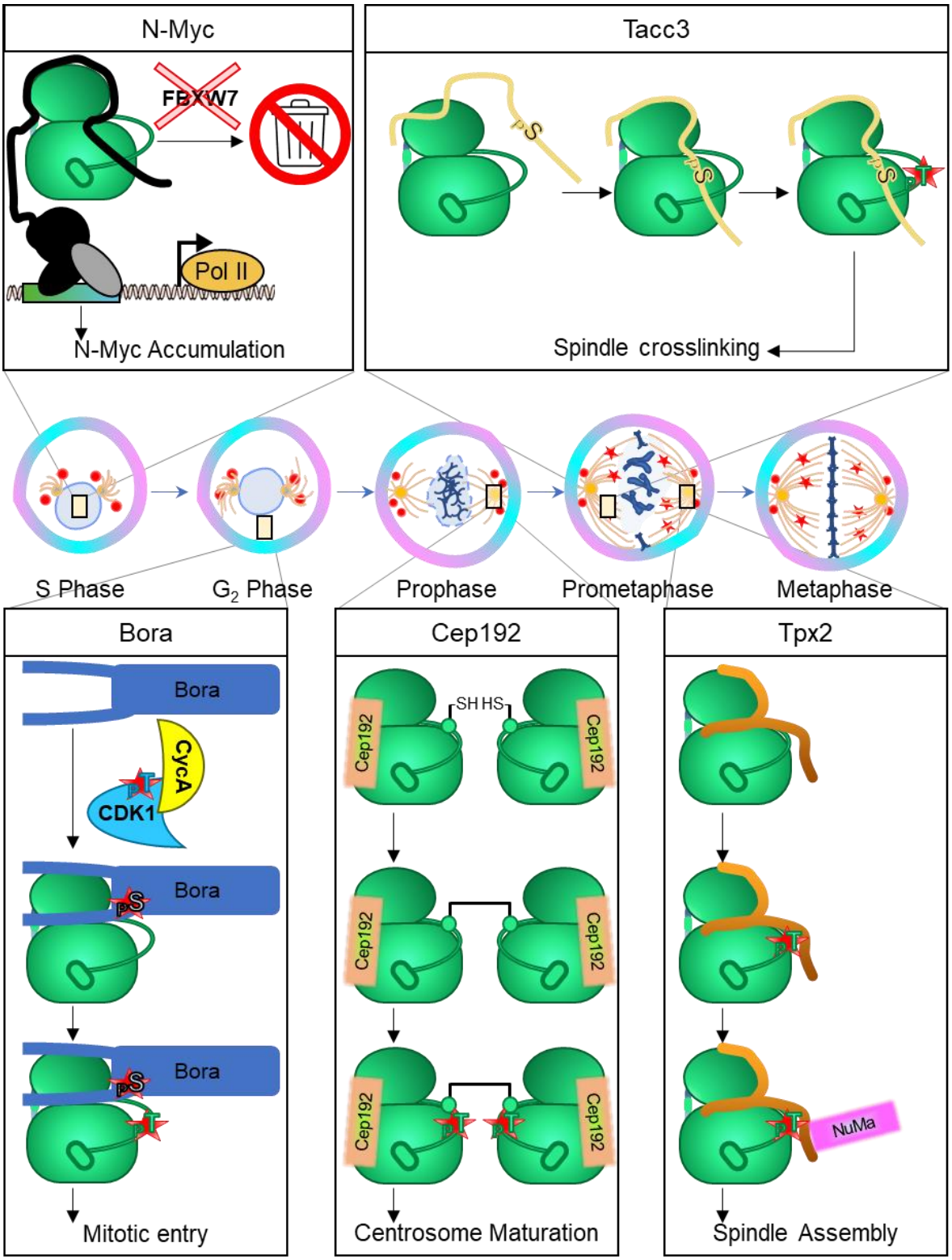


Figure 1-21 The diversity of AurA activation during cell cycle

To activate AurA, Bora must be phosphorylated by CyclinA2-Cdk2, pBora then forms a complex with PLK1 to induce AurA activation using the same pockets as TPX2. TPX2 doesn't require phosphorylation but may induce dimerization to activate AurA. Cep192-dependent oligomerization and oxidation of Cys290 by ROS at the centrosomes activates AurA. Tacc3 binds to the N-lobe of AurA, which undergoes phosphorylation and then activates AurA.

1.3.2.3.4 Interaction with TPX2

Targeting protein for Xenopus kinesin-like protein 2 (TPX2) is the best-characterized Aurora-A activator and is expressed from G1/S transition to cytokinesis and then rapidly degraded. It was first identified as a binding partner of the Xenopus kinesin-like protein 2 (Xklp2) and it localises Xklp2 to the spindle pole in prometaphase and metaphase.[143] Only the first 43 residues of 747 residues of TPX2 are necessary for interaction with and activation of AurA.[144] Zorba *et al* argued that only the upstream domain comprised of residues 7-21 of TPX2 is required to activate AurA as they produced biochemically validated crystallographic structures of a dimer of heterodimers (AurA+TPX2) capable of generating active AurA.[145]

AurA nuclear localization is promoted by TPX2 and counteracted by protein degradation as free AurA is deactivated by dephosphorylation by phosphatases.[146]

Despite the range of potential AurA activation loop conformers, the binding of TPX2 has been shown to not alter the conformational range of AurA, rather its appears to select and stabilise existing AurA conformations.[49] Hence current work has focused on designing ATP-competitive kinase inhibitors to facilitate kinase inactivation, in the quest to develop novel cancer therapeutics, and it has not been possible to develop molecules that prevent activation of AurA by modulating its conformation allosterically to favour inactive states.[147] Previously it has been shown that crystal structures of AurA+TPX2 bound to small molecules don't represent the structure adopted in solution with respect to the positioning of the DFG motif and activation loop, demonstrating the challenge of using X-Ray crystallography in structure-based drug design studies that do not account for dynamics in solution, and understanding this is key for kinase activation.[144 148 149]

1.3.2.3.5 Interaction with TACC3

TACC3 is a member of the transforming acidic coiled coil (TACC) family of centrosome proteins that enhance MT stability in mitosis. TACC3 519-563 is the minimum binding domain required for interaction with AurA.[150] The interaction was originally discovered in Xenopus laevis, where AurA phosphorylation of Maskin (Xenopus homologue of Tacc3) was required for mRNA translation control. It was shown, as with the human homologue, to be an activator of kinase activity, as binding of Maskin to AurA induces an increase in AurA kinase activity.[151] Phosphorylation of TACC3 S558 induces TACC3 localisation on the spindle microtubules and is critical to its binding

to clathrin,[152] as it induces a local conformation change of TACC3 from disordered to helical. Structurally, the interaction of TACC3 is two-step process, with TACC3 binding to AurA using motifs on F525 and L532. After this activated AurA phosphorylates TACC3 on S558 and reorganises the binding surfaces.

1.3.2.3.6 Interaction with BORA

The gene was named Bora for Aurora-A Borealis because the phenotypes of Bora and Aurora-A mutants were similar, and they localise in patterns that resemble the northern lights. The phosphorylated form of Bora activates cytoplasmic AurA, in a manner predicted to be similar to the TPX2 activation mechanism; however the Bora interaction occurs much earlier than the TPX2 one in mitosis.[88] Bora requires prior activation by Cyclin A–Cdk1 and during the S/G2 transition pBora is exported to the cytoplasm, allowing it to phosphorylate AurA, and interact with PLK1.[153]

PLK1, AurA and Bora are involved in a multi-step cascade but do not form a ternary complex.[154] Instead, during early G2 pBora promotes a dimerization of cytoplasmic PLK1, relieving PLK1 auto-inhibition and hyper-phosphorylating pBora. This allows AurA to phosphorylate PLK1 that is translocated into nucleus for mitosis.[155]

Using NMR, point mutations and other biochemical experiments, the phosphorylated N-terminal region of Bora was shown to bind AurA to form a fuzzy complex using the sites of TPX2 (Y-, F- and W-pockets) but critically pBora stabilises unphosphorylated AurA to form an active conformation. The pSer112 of Bora ‘stands in’ for phosphorylated Thr288 of AurA, forming the stabilising contacts of pThr288 to allow AurA to activate and phosphorylate substrates.[88]

This model of intrinsically disordered Bora activating AurA using a highly allovalent and dynamic interaction was built using the alignment to TPX2 binding surfaces.

1.3.3 Introduction to AurA – structural dynamics

1.3.3.1 The structural features of AurA in detail

1.3.3.1.1 Introduction

Here we directly characterise AurA as an AGC kinase to be able to compare AurA to a wider range of potential kinase conformations.[49] As previously explained AurA has a twostep activation mechanism which is thus typified as two conformations; an inactive, un-phosphorylated conformation and an

active, phosphorylated AurA conformation. Cyphers *et al* have also presented a third conformation of AurA with the activation loop bound in an active conformation, but the DFG motif rotated preventing kinase activity.[156]

1.3.3.1.2 Structure – Pre-defining kinase nomenclature

1.3.3.1.2.1 Why do we have to adapt the naming nomenclatures?

For this thesis we precisely define several AurA conformations, as structures that we explore fall outside the common kinase nomenclatures.[157] At times we will refer to the common names of the conformations, and also use the Modi and Dunbrack nomenclature to define DFG motif conformations relative to other kinames.[158]

1.3.3.1.2.2 Naming each helix, loop and sheet

Figure 1-24 shows the naming system arrived at for every beta sheet and helix found in AurA. Loops are named using the adjoining sheet or helix. In the text the name αC helix is used to describe Helix C. NTE or CTE, N-Terminal or C-Terminal Extensions, are located at the N-terminus or C-terminus of AurA.

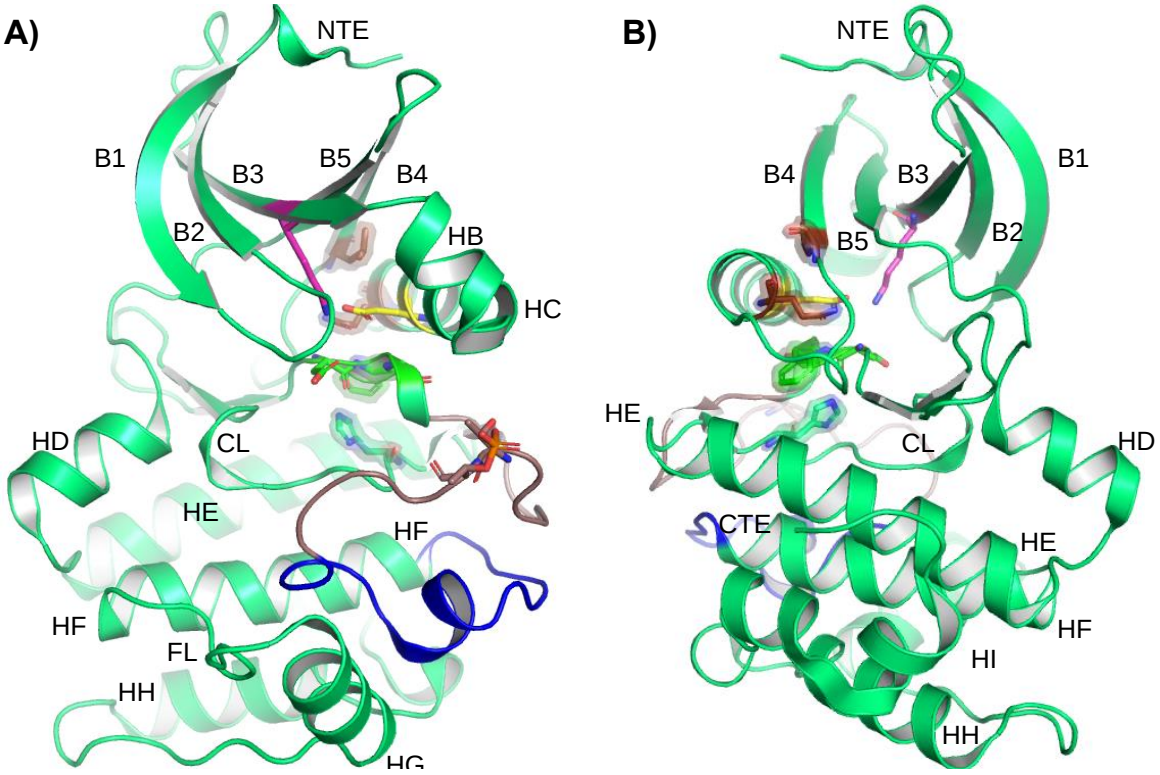


Figure 1-22 Names of all the AurA loops, sheets and helices

The canonically active AurA structure (PBD: 1OL5) annotated with helix and beta-sheet names. Residues 122-387 commonly found structured with residues 1-119 removed for crystallographic constructs as this truncated version is significantly more amenable to structural and molecular biology.

1.3.3.1.2.3 Naming activation loop conformations in Aurora-A

Figure 1-26 describes the range of AurA conformations explored in this thesis.

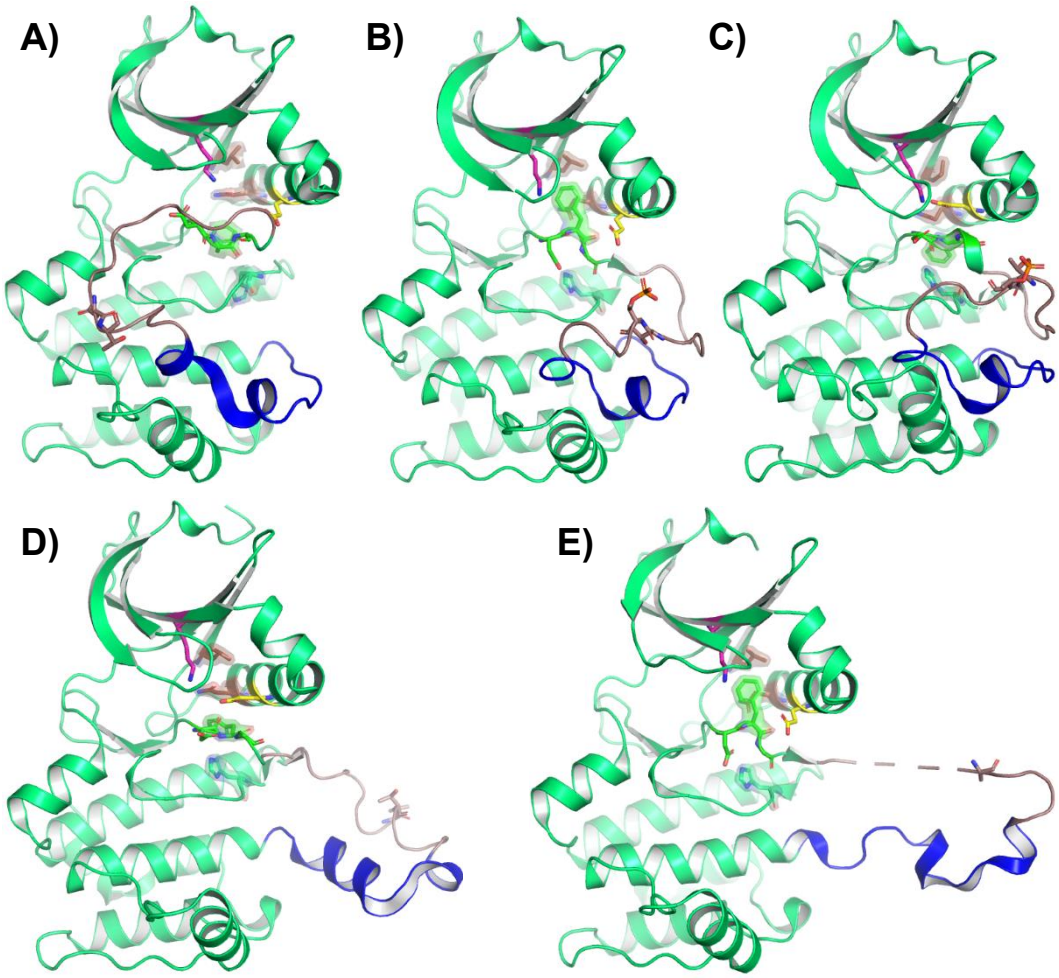


Figure 1-23 Activation loop position nomenclature
Examples of A) Mps1-like or MLN-like pose (PDB code: 4J8M) aligned but incorrectly stacked R-spine and a broken salt bridge. B) Active (PDB code: 4CEG). C) Inactive active (PDB code: 5L8K), autoinhibited DFG-out. D) Continuous helix (PDB code: 4C3P). E) Discontinuous helix (PDB code: 6C2T).

1.3.3.1.2.4 What are the common structures of AurA?

A brief examination of structures present in the PDB shows a plethora of structures of AurA. The biggest variation is in the activation loop position, and then twisting and hinging motions. The NTE and CTE are not generally structured. The majority of structures are monomeric, with some domain swapped or crystallographic dimers. The nomenclature available to describe the activation loop position relative to the N-/C-lobe focuses on the DFG position. However, DFG-in, -inter and -out nomenclature fails to describe the positioning of the entire activation loop length. The activation loop is sometimes described as ‘open’ or ‘closed’. When nomenclature around

DFG-in/out is used, there is also an implicit description of the activation status of the conformer, as an incorrect DFG orientation means that the salt bridge will be broken, and thus the conformer will be inactive.

1.3.3.1.2.5 Three subdomains of the activation loop

For this thesis, we developed a nomenclature to describe the position of the activation loop more fully by dividing the activation loop into three subdomains (Figure 1-27). Residues 272-277 of the activation loop were named as part of the **DFG subdomain;** residues 278-292, the **phosphorylated loop** subdomain, and residues 293-308 form the **structured subdomain**. We also noted that some but not all AurA structures were able to pass some activation tests. The activation loop terminates at Helix F, which is formed of residues 304-325. AurA structures appear to always have a structured helix B and C, unlike other kinases where it can present as disordered in inactive structures. From this, we named two families of structures – those with semi-dynamic and those with fully-dynamic activation loops.

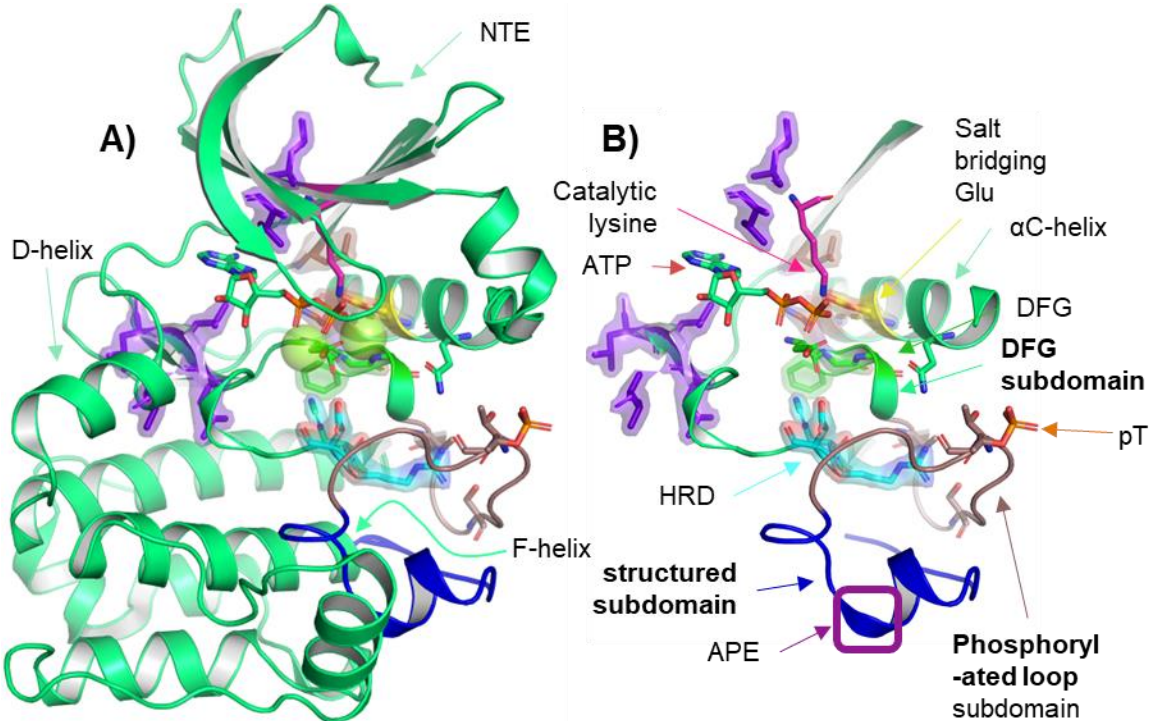


Figure 1-24 Structure labels used in nomenclature of AurA

Structure labels used in nomenclature of AurA (PDB code: 4CEG). Panel A) shows AurA active structure with the D- and F-helices labelled Panel B) shows selected residues and motifs relevant to subdomains of the Activation loop.

1.3.3.1.2.6 Family 1

The first family of structures are those with semi-dynamic activation loops. In these, only the DFG and phosphorylated loop subdomain are mobile; the structured subdomain maintains its position between Helix F and G. This family includes structures that are canonically active and structures that involve the rotation of the DFG subdomain, in addition to structures with variation in the position of the phosphorylated loop subdomain, either in contact with Helix D, B or E. These can be quantified by measuring the distance between a residue on Helix D and the phosphosite on the activation loop. These structures are those that are sometimes referred to as DFG-in, activation loop open or 'dici' and 'dico'.[159]

The structures with the shortest distance between the phosphorylated loop subdomain and Helix D are informally named '**MLN-like**' structures, as these structures were first produced by AurA complexed with the MLN-series of inhibitors. These so-called MLN-like structures have been seen in a range of conformations, depending on the inhibitor they have been complexed with. The activation loop is either bound directly to Helix D or is unstructured. In fact, in the majority of these types of structures, the activation loop – or a portion of it – is unstructured in the crystal structure, for example structures with PDB codes 2X81 or 6HJK. This conformer can be presented as a face-to-face dimer or a monomer as the biological unit. No crystallographic structures of this conformer have been found without the use of a conformationally disrupting drug, since the activation loop is too disordered to be modelled. However, a strikingly similar structure has been found for the human protein kinase, MSP1, revealing an ordered conformation of the activation loop. Hence a better name is **MSP1-like,** and is the one used in this thesis**.**[160]

Also in this family are structured subdomain in, alpha-helix contact structures. In these structures the activation loop is unstructured but appears to be in contact with helix beta or the glycine rich loop, such as PDB 2C6D.

Finally in family 1 are the active or **active-like** structures, where there is contact between the phosphorylated loop subdomain and Helix E. T287-288 sits either between the structured subdomain and the helix B or in a fully active position. There are some subtle variations in where residues 286-291 are placed, which seems to be dependent on the construct and the crystallographic ligands. We see structures where this part of the phosphorylated loop subdomain involves Cys290 interacting with Lys143,

such as the structure with PBD code 5DT3, or alternatively, as with the structure with PBD code 1OL5, Cys290 is bound to Arg255.

1.3.3.1.2.7 Family 2

The second family identified is the reason why we are defining a new conformer nomenclature. In family 2, the structured subdomain (activation loop 293-308) is displaced into an extended position. The PDB contains domain swapped dimer AurA structures, so we have to be clear not to overstate the biological relevance of the monomeric structure, informally named pseudo-dimer or dimer-like structures. Dimers are formed either by crystal contacts, such as those seen in structure PBD code 6C2T, or where domain swapping is mediated by TPX2 binding, as with PBD code 4C3P, or by a disulfide bridges. PBD code 3UOK is a similar also a crystal contact dimer.

These conformers of AurA can be categorised loosely by three poses of the activation loop. **Structured subdomain out, alpha-helix contact** conformers, like PBD code 6VPJ. This is different from structured subdomain in pose, as the structured subdomain is rotated out of contact with Helix F and Helix G.

The structure with PBD code 4C3P can be described as **structured subdomain out, continuous helix.** Here, the structured subdomain is helical, but in contact with the phosphorylated loop subdomain. This structure is also seen other kinases such as DAPK3 (PBD code 5A6O).

Finally**,** structures like PBD code 6C2T are characterised as **structured subdomain out, discontinuous helix.** The loop is disordered between the structured subdomain and the Helix F, because the phosphorylated loop subdomain and structured subdomain are behaving as a separated domain. This is seen with the Unc-51-like kinase 2 structures, where the activation loop exchange enables dimeric assembly. A singular structure of AurB complexed with an inhibitor and activator, like TPX2, features a discontinuous helix due to a domain swapped dimer structure (PDB code 4AF3).

This nomenclature allows us to precisely describe and group potential AurA models and relate the inactive structures to active ones. For instance, comparing structures from PBD codes 4C3P and 1OL7, the former has a formed R-spine, DFG-in (DFGin-ABAminus) while the latter does not (DFGin-BLAminus), but both structures appear active, when based just on the active site, but have dramatically divergent phosphorylated loop and

structured subdomains, which directly relate to the biological role of the structure and kinase activity. DFG-in/out were therefore inadequate descriptors, and our new nomenclature necessary.

1.3.3.1.2.8 Summary table

The described conformers are summarised in table 1-4.

Table 1-4 Summary of conformers described

Name	Family	DFG	Phosphorylated loop	Structured subdomain	Example
Active-like	1	DFG-IN	Held in active conformer	Held in active conformer	1OL7
Active-like autoinhibited	1	DFG-IN / -INTER	Held in active conformer	Held in active conformer	5L8K
MSP1-like	1	DFG-Out	Held against D-helix or unstructured	Held in active conformer	4J8M
Structured subdomain out, alpha-helix contact	2	NA	Compact in contact with the αC helix and the structured subdomain	Rotated out of contact with Helix F and G	6VPJ
Structured subdomain out, continuous helix	2	NA	Ordered or disordered in contact with the structured subdomain	Extended and continuous with Helix F	4C3P
Structured subdomain out, discontinuous helix	2	NA	Extended, Ordered and in contact with the structured subdomain	Extended and discontinuous with Helix F	6C2T

1.3.3.2 Modelling the conformational dynamics of AurA

As the region of AurA 1-119 is unstructured, it can be classified as an IDPR and as a result it is often omitted from AurA constructs used for structural biology (e.g. X-ray crystallography) and high throughput screening of small molecule binders. Its characterized as being in a conformational equilibrium of conformers due to its dynamic activation loop lacking conserved features and the disordered NTE and CTE. Due to AurA being an oncogenic target there has been vast library of x-ray structures, and a plethora of in-solution techniques and MD have been used to generate models of AurA dynamics.

Crystallographic structures seem to be biased and lack the NTE, with interactors often ejected during crystallization. Additionally, domain swapped dimers are found under certain conditions. Many groups exploring AurA using in-solution approaches have adapted crystallographic structures which are processed using MD simulations to understand experimental data.

In 2011, kinetic analysis of the catalytic activity of AurA in the presence and absence of TPX2 found that TPX2, synergistically with activation loop phosphorylation, stabilises the active conformation of AurA.[161] Later,

PELDOR spectroscopy using a crystallographic construct of AurA labelled on the αC-Helix and the activation loop, demonstrated that TPX2 locks the activation loop in a fixed conformation.[162] Hence, it is not unexpected that the binding of TPX2 to AurA decreases the K_m, i.e. increases the affinity, for nucleotide binding.[163]

Using the crystallographic construct of AurA labelled on the Helix D and the activation loop, Lake *et al* demonstrated that inactive AurA presents as an equilibrium of MSP1-like and Active-like conformer.[148] These in-solution observations were fitted to crystallographic structures. However, the addition of TPX2 and a nucleotide greatly shifted the equilibrium in favour of the active conformer. Striking they show that a unique linkage between the C-helix, R-spine and activation loop is mediated through an allosteric network of solvent water molecules.[156] Additionally, Ruff *et al* found, using other in-solution MD simulations and EPR/PELDOR experiments, that activation loop phosphorylation initiates a further equilibrium shift from an Active-like autoinhibited conformer to the *true* Active-like conformer.[164]

Gilbert et al also studied, using single-molecule fluorescence spectroscopy, the AurA dynamic equilibrium and the effect of phosphorylation and TPX2 binding.[165] Interestingly, they found that neither substrate nor nucleotides shifted the equilibrium but together they increased the proportion of the inactive conformer AurA.

Together this suggests a two or three conformer model, where AurA is inactive, interconverting between MSP1-like and Active-like conformers and upon TPX2 binding this shift towards an Active-like autoinhibited conformer which requires phosphorylation to trigger the transition to an Active-like conformer and onto catalytic activation.

Re-evaluation of PELDOR traces and generated αC-Helix-Activation loop distance distributions presented by Burgess *et al.* showed a shouldered main peak and a minor second peak at a longer distance. This suggests the overlapping of two conformers with a definitive main conformer and a minor conformer. This is reminiscent of the MSP1-like, Active-like autoinhibited conformer and Active-like conformer ensemble by the FRET.

In 2014, Zorba *et al* proposed an AurA activation mechanism that used intermolecular autophosphorylation in a transient but long-lived domain-swapped dimer structure produced by X-ray crystallography that was validated by functional assays.[145] They thus propose asymmetric, trans-auto-phosphorylation in dimeric AurA as an activation mechanism. However,

the only in-solution, orthogonal data for this type of dimer comes from the redox mediated formation of a face-to-face dimer stabilized by a Cys290–Cys290 disulfide bond capable of auto-phosphorylation.[166]

One would have expected to see reports of a dimer in the native mass spectrometry experiments of Tomlinson *et al.* but no such evidence appears.[167] They do report three conformers of active AurA, a major conformer that matches the size of the canonical Active-like conformer and two minor conformers, one smaller and one larger.

Akamine *et al* found the Apo form of PKA to have its Gly-loop and n-lobe latched open relative to its nucleotide bound form suggesting that Apo AurA would also undergo this stabilising *lid* closure upon the addition of nucleotide.[168] Indeed, Musavizadeh *et al* demonstrated, using MD simulations, that the stability of the Active-like conformer was enhanced by phosphorylation of the Activation loop but that the addition of ADP/Mg to the simulations increased the dynamic motions of the activation loop but brought the N- and C-lobes of Apo closer together.[169]

This is notable as Pitsawong *et al* found that nucleotides can bind to both MSP1-like and Active-like conformers of AurA using a two-step mechanism of binding.[89] Additionally, they agreed with Cyphers *et al.* who previously showed that the conformational ensemble of AurA isn't altered by nucleotide binding.[156] Interestingly, the binding of TPX2 to phosphorylated AurA increases the binding affinity of nucleotides, in line with previous kinase assay kinetic experiments.[163]

ADP can be turned over by AurA to phosphorylate substrates but not be used in AurA auto-phosphorylation.[170]

Questions are outstanding about the dynamics of AurA and its mechanism of action although there is broad agreement that TPX2 and phosphorylation stabilise AurA in an Active-like conformer capable of substrate phosphorylation that shifts the conformational ensemble of AurA. AurA does not undergo dramatic domain re-organisation to perform catalysis. The mechanism of AurA activation and other minor conformers that might contribute to that process is still unclear.

1.3.4 High value drug target – AurA+N-Myc TAD

1.3.4.1 Introduction to the interaction

The N-Myc and AurA complex represents a key target for anti-cancer drug discovery as up regulation of N-Myc has been shown to be protected by the

interaction with AurA.[50] The oncogenic, feed forward interaction occurs in S-phase when nuclear localised AurA binds to the transactivation domain of N-Myc, effectively preventing the proteasomal degradation of N-Myc by SCF^{FBXW7} ubiquitin ligase. This provides protein protection and stabilisation and its effect is independent of kinase function. N-Myc, in complex with MAX, binds to the majority of promoters with an open chromatin structure and promotes transcriptional elongation by RNA polymerase II using a meditator complex.[171] [116] This is striking in *MYCN*-amplified neuroblastoma, where binding of AurA to an overlapping region reduces FBXW7 binding to N-Myc.

1.3.4.2 Introduction to N-Myc

1.3.4.2.1 N-Myc structure TAD

Myc proteins are divided into three regions, the transactivation domain (Figure 1-28), the nuclear localisation sequence (green) and the DNA binding region (purple). The basic helix-loop-helix leucine-zipper (bHLHZ) domain binds DNA. The smaller, conserved sequence motifs are known as Myc boxes (MB)**.** MB0-IV are important for protein-protein interactions**.** The highlighted pathway shows the common route of Myc degradation.[172] Cell stimulatory signals drive the activation of mitogen-activated protein kinases (MAPK) and thus S62 is phosphorylated by extracellular signal-regulated kinase (ERK). Phosphatidylinositol 3-kinase (PI3K) leads to T58 phosphorylation by glycogen synthase kinase 3β. This down regulates Myc but prior phosphorylation at S62 is required. Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1) isomerises P63 to facilitate removal of the stabilizing phosphate at S62 by the trans-specific Protein phosphatase 2A (PP2A). The E3 ubiquitin ligase SCF^{FBXW7} then targets pT58-Myc for proteasomal degradation.[173]

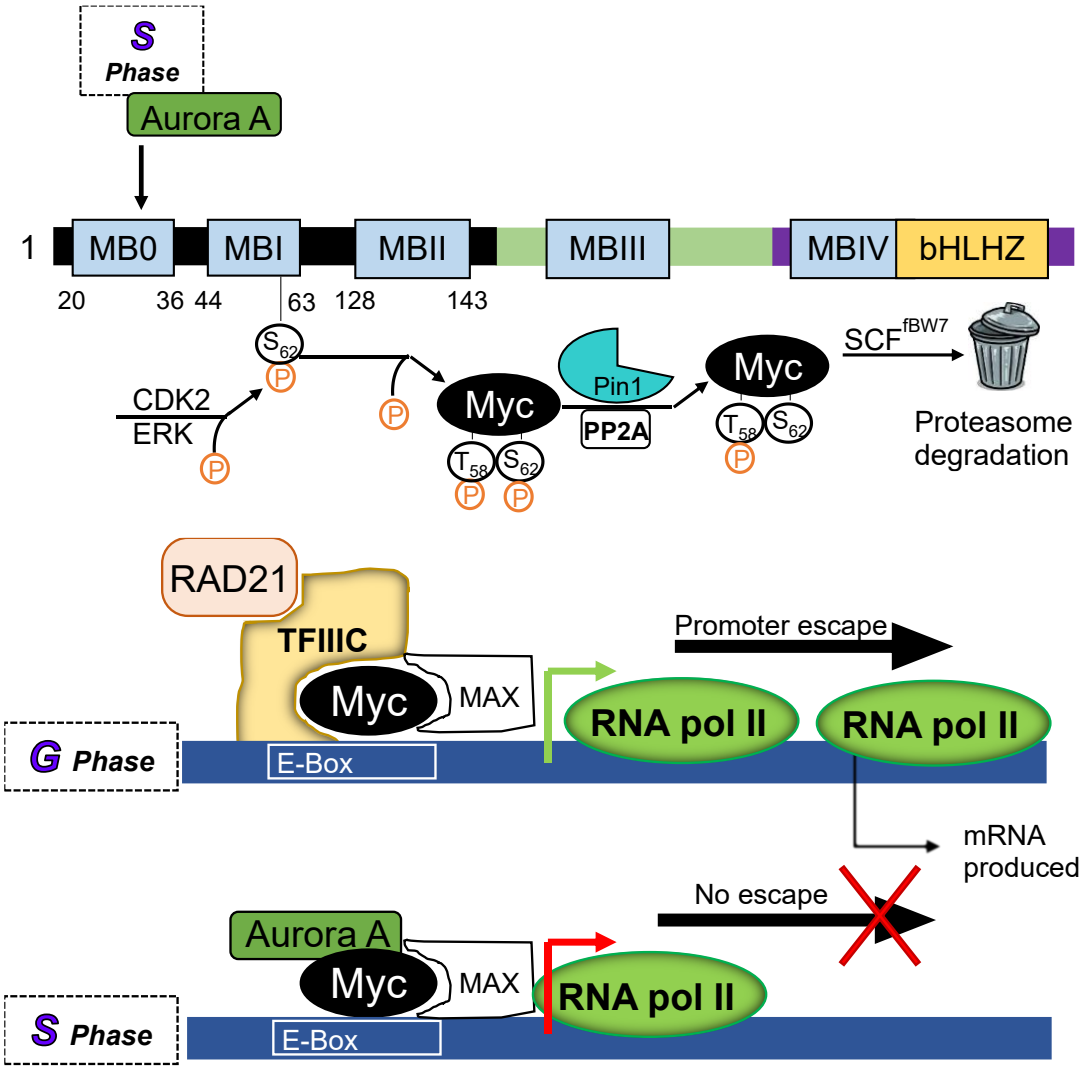


Figure 1-25 Introduction to N-Myc

The structure of N-Myc and its cell cycle dependent interactions. Upper) The structure of N-Myc is divided into TAD, central region and DNA binding region. Myc boxes annotated. The proteasomal degradation pathway that AurA disrupts. Lower) Myc binding to an E-box in DNA with cofactors that allow mRNA synthesis. During G phase, when AurA activity is low, TFIIIC and RAD21, aids the 'pause release' of RNA pol II. Conversely, when AurA activity is high AurA outcompetes TFIIIC and RAD21 in binding to Myc preventing transcription.

Myc *addiction* and dysregulation is a common hallmark of cancer and has thus been subjected to a diverse campaigns of drug discovery to develop small molecule anti-cancer therapies in the hope that treatments will be applicable to multiple diseases. Transgenic mouse models have shown that reducing Myc activity can lead to tumour regression.[174] Myc has earned the moniker of "undruggable" partly due to the number of treatments that have failed in phase I/II trial.[54] The drugs perform poorly because of low bioavailability, rapid metabolism and off-target effects. Difficulties arise from Myc lacking an enzymatic pocket to target or screen against, its localisation

in the nucleus as a heterodimer and, its functions as a super controller, meaning it is critical to healthy cells. To meet these challenges, novel therapeutics have been developed such as G-quadruplex stabilizers, e.g. Quarfloxin, that seek to reduce Myc mRNA transcription, or antisense oligonucleotides, e.g. AVI-4126, to reduce Myc translation.[175 176 177] Targeting the Myc transcription factor, NME2, with a natural product small molecule analogue showed that Myc may be tractable to small molecules via upstream inhibition.[178]

1.3.4.2.2 The molecular basis of the AurA KD+N-Myc TAD structural features

Figure 1-29 shows the structure of the crystallised complex AurA KD 122-403 and a portion of the TAD N-Myc. However, this structure does not tell the full story. Richards *et al* reported that fragments corresponding to the C-terminal and N-terminal ends of N-Myc 28-89 independently bound to AurA with an affinity of approximately 10 µM, and that the entire span of the N-Myc 28-89 fragment also bound with an affinity of ≈10 µM.[171] The incomplete crystal structure has been the basis for multiple ongoing diverging drug discovery campaigns.[178] The stoichiometry and valency of the interaction is unclear.[171] As well as developing inhibitors of the catalytic activity of AurA, several inhibitors of AurA that disrupt of the AurA+N-Myc complex using conformation modulation have been presented.[11] It's also of note the AurB has also been identified as a target that stabilizes Myc.[179]

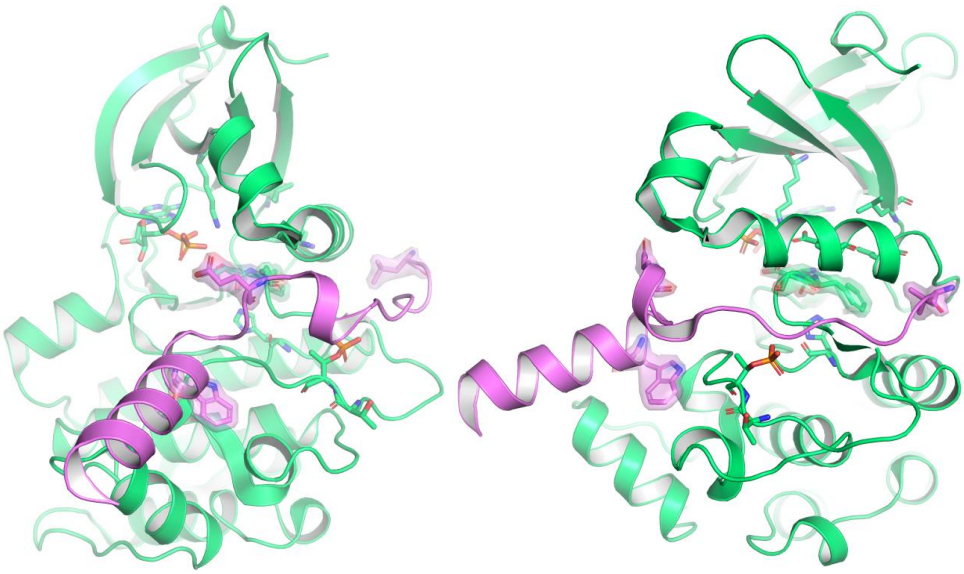


Figure 1-26 N-Myc TAD AurA structure

The sole crystal structure of AurA KD 122-403 and N-Myc 61-89. Left) The fragment of N-Myc binds of the activation loop, directly competeting for a site used in AurA-substrate interactions, and into the W-pocket whilst AurA is in an canonically active structure. Right) Side of the complex with key residues of the interaction highlighted.

1.4 Introduction to methodologies used in this thesis

1.4.1 Chemical crosslinking mass spectrometry (XL-MS) workflow to map protein interactions

1.4.1.1 Structural approaches investigations

As discussed in Section 1.3.3.3, many groups have used a wide variety of methods to characterize AurA. Exploring AurA conformational dynamics using label free mass spectrometry experiments with AurA crystallisation constructs will allow the investigation of AurA and its interactions with IDP activators whilst enabling comparison with crystallographic structures.[180]

1.4.1.2 Basic principles and workflow of XL-MS

1.4.1.2.1 Introduction to the principles of XL-MS

1.4.1.2.1.1 Basic principles of XL-MS in context

Protein structures, generated using X-ray crystallography[181] or cryo-electron microscopy (cryo-EM), provide the near-atomic or residue level positioning of protein polypeptide chains to give the distances, interactions, conformations and modification status of all structured amino acids and their accompanying ligands (drugs, inorganic, buffer), and they have served as the cornerstone of structural biology, especially in the context of drug development.

Proteins or protein complexes can resist crystallisation wholesale, not allowing their diffraction from which to generate a structure, or crystallisation can occur whereby parts of the protein pack heterogeneously in the crystal lattice, resulting in ambiguous and unresolvable structures.[182] Disordered proteins or proteins with disordered regions are prone to this effect.[183] As the energy of crystallisation is often more than the binding energies, crystal structures may have artificial binding interfaces or conformations.[184] Several strategies have been developed to combat proteins' resistance to crystallisation, and recombinant proteins are designed to minimize disordered regions or co-crystallisation with chaperones.[185]

Chemical crosslinking mass spectrometry (XL-MS), Figure 1-30, gives medium resolution structural information based on distance restraints, but has lower sample consumption and faster experiments, and is easier to adapt from already validated recombinant proteins.[186]

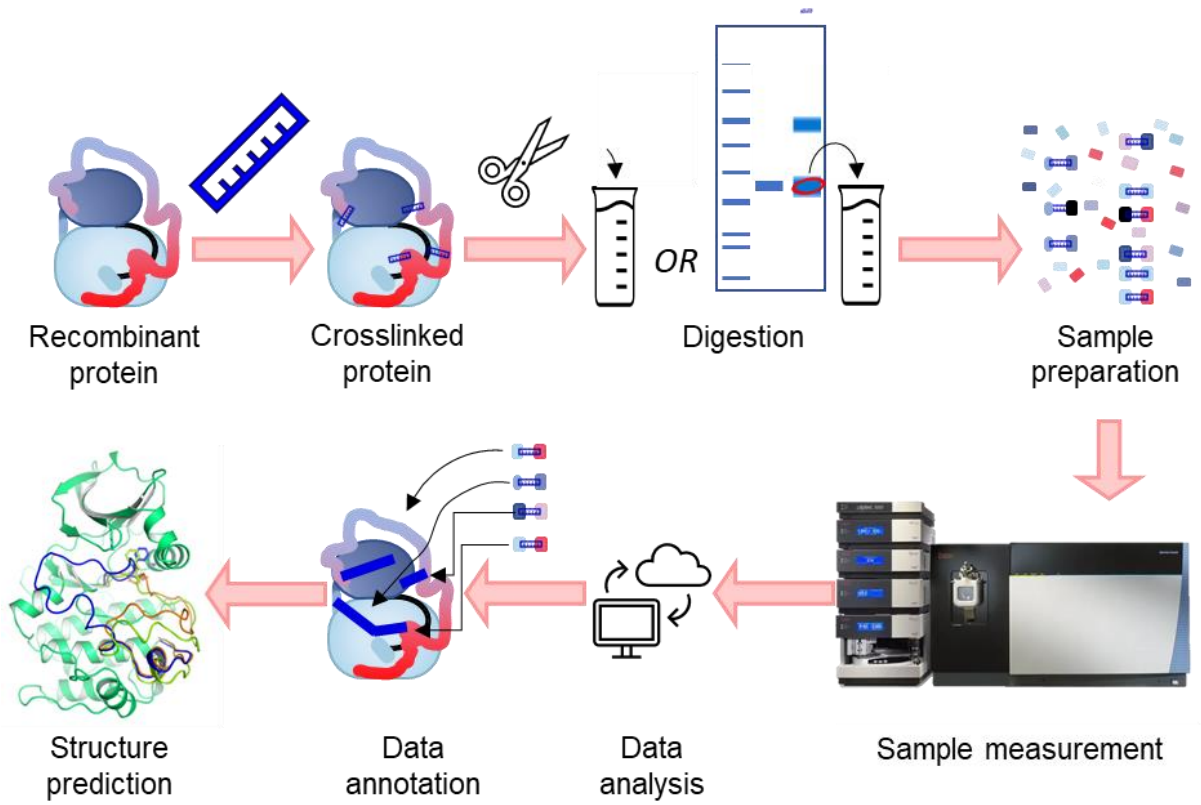


Figure 1-27 The basic process of XL-MS

The crosslinking of a recombinant protein samples requires stable protein sample to be reacted with a crosslinking reagent which is then either directly digested in solution or separated using SDS-PAGE gels and the target protein species is excised for digestion. After the measurement of the crosslinked peptides and processing of the data, crosslink residue pairs are used to generate distance restraints to help derive molecular models of the target protein.

XL-MS is an in-solution technique where a bifunctional-reagent acts as a 'molecular ruler', with the formation of covalent bonds between the reactive side chains of residues giving distance restraint information about a protein. The residues must be solvent accessible, reactive with respect to the chemistry of the crosslinker, and within the distance of the crosslinker for sufficient time for a reaction to occur. XL-MS is a 'bottom up' technique, where proteins are characterized after enzymatic digestion of the sample.[187]

Fundamental to all bottom up structural and proteomic MS approaches is the challenge of identifying modified peptides from a sample using database searches and locating the site of modification. The analysis and interpretation of the collected experimental data can be arduous and necessitate orthogonal data to eliminate differential hypothesises. For XL-MS, as well as optimising the crosslinking reaction and measurement, the challenge with using non-mass spectrometry cleavable cross-linkers is the n^2 ("n-square") problem.

The more complex the reacted protein system, the search space increases quadratically. This is due to the linear increase in the peptide database searched against; the search for two peptides covalently linked per spectrum increases quadratically.[188]

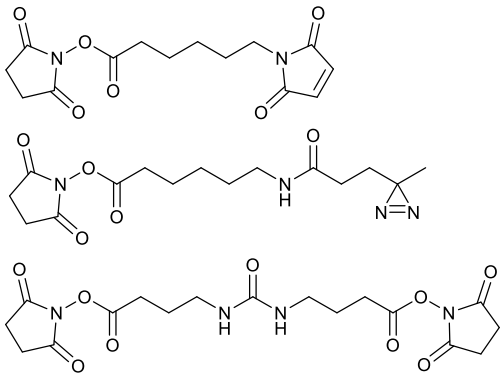


Figure 1-28 Parts of a crosslinker

The chemical structures of heterobifunctional and a homobifunctional crosslinkers. Upper) EMCS (N-ε-malemidocaproyl-oxysuccinimide ester) featuring NHS-esters and Maleimide reactive groups. Middle) NHS-LC-Diazirine (succinimidyl 6-(4,4'-azipentanamido)hexanoate) featuring NHS-esters and Diazirine reactive groups. Lower) DSBU (disuccinimidyl dibutyric urea) a trifunctional crosslinker, containing two NHS-ester groups, where a mass cleavable linker makes peptide identification easier than the alkyl chain equivalent.

1.4.1.2.1.2 Introduction to crosslink chemistry

A range of chemistries, lengths and functionalisations of crosslinkers have be developed to cater for a variety of biomolecular applications (Figure 1-31). Bifunctional crosslinkers can vary by the amino acid sides chain they react with, their chemistries, the specificities of the reactions, and the length of the linker joining these.[189 190] NHS esters, maleimides and photo-activatable diazirines are the common chemistries employed for biological structure determination. NHS-esters are primarily reactive towards lysine residues and secondarily reactive towards hydroxyl-containing side chains, maleimides are reactive towards cysteine residues and diazirines are only reactive after exposure to UV light and can react with any residue side chain with a preference towards acidic residues. [191] Trifunctional crosslinkers are where the linker is further functionalised to improve characteristics of the crosslinker. For example, crosslinkers can include an enrichment tag that allows for the removal of non-crosslinked peptides, to increase the sensitivity. This can be achieved by a click handle or directly through an inclusion of an affinity handle like biotin.

1.4.1.3 Development of chemical crosslinker technology

For this thesis, the crosslinker DSBU with crosslink search engine MeroX is used exclusively. It was selected as it fell between the middle ground of being ‘functionalised’ enough to be able to allow for timely crosslink identification on a personal computer, whilst still being off-the-shelf and cost effective.

1.4.1.3.1.1 Fragmentation

To circumvent the n^2 problem, linkers were developed that fragment during tandem MS measurement to give fingerprint ions that distinguish crosslinked peptides from non-crosslinked peptides. MS cleavable linkers, like DSBU, reduce the possible search space but increases the complexity of MS measurement.

1.4.1.3.2 Chemical crosslinking reaction

NHS ester reactive molecules have a long history of use in bio-conjugation, to decorate proteins and peptides with probes, dyes, enrichment handles, drugs etc.[196]

Chemical crosslinking is a type of bio-conjugation, and N-hydroxysuccinimide ester (NHS ester) reactive crosslinkers are the most commonly employed chemistry for the study of protein dynamics and PPIs. A wide array of NHS-ester crosslinkers have been developed, both bi-/tri-functional and hetero-/homo-functional, used with recombinantly expressed proteins as the reliability gives results across a diversity of protein types and reaction conditions.[197]

The covalent bond is formed between two amino acids by the nucleophilic attack of the ester carbonyl by a sufficiently reactive and proximal amino acid side chain, or the N-terminus of the protein, as shown in Figure 1-32. The primary amine of Lysine and the N-terminal residue, and hydroxyl groups of Serine, Threonine or Tyrosine, are normally considered reactive towards NHS-esters. However, the reactivities of these functional groups are dependent on the local environment of the residue side chain and the composition of the buffer. Mechanistically, the covalent amide bond is formed between the NHS-ester of the crosslinker and the side chain of lysine by the free electron pair of a neutral primary amine attacking the carbonyl of the activated ester. The rate limiting step is the formation of the tetrahedral intermediate prior to the ejection of the N-hydroxysuccinimide.[198] [199]

The half-life for the NHS ester in buffered water, due to hydrolysis, is hours at pH 7 and 0 °C, but minutes for pH > 7 at room temperature.[187]

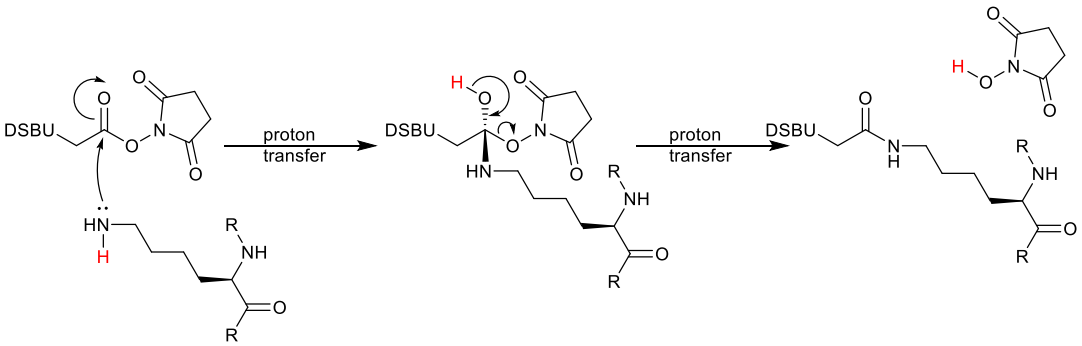


Figure 1-29 NHS ester reaction with transition steps

NHS-ester of DSBU reacting, nucleophilic addition, with zeta-nitrogen of lysine and generating N-hydroxysuccinimide. The reactive tetrahedral intermediate is the rate limiting step. Proton transfer steps removed for brevity.[198]

1.4.1.3.2.1 Crosslink types

The crosslinking molecule can react with a two-protein system to give potentially four types of crosslink, shown in Figure 1-33 and summarised in Table 1-5. For NHS-ester crosslinkers, these can be explained by progressing the reaction with hypothetical substrates to understand how and why certain crosslinks are generated for a protein system.

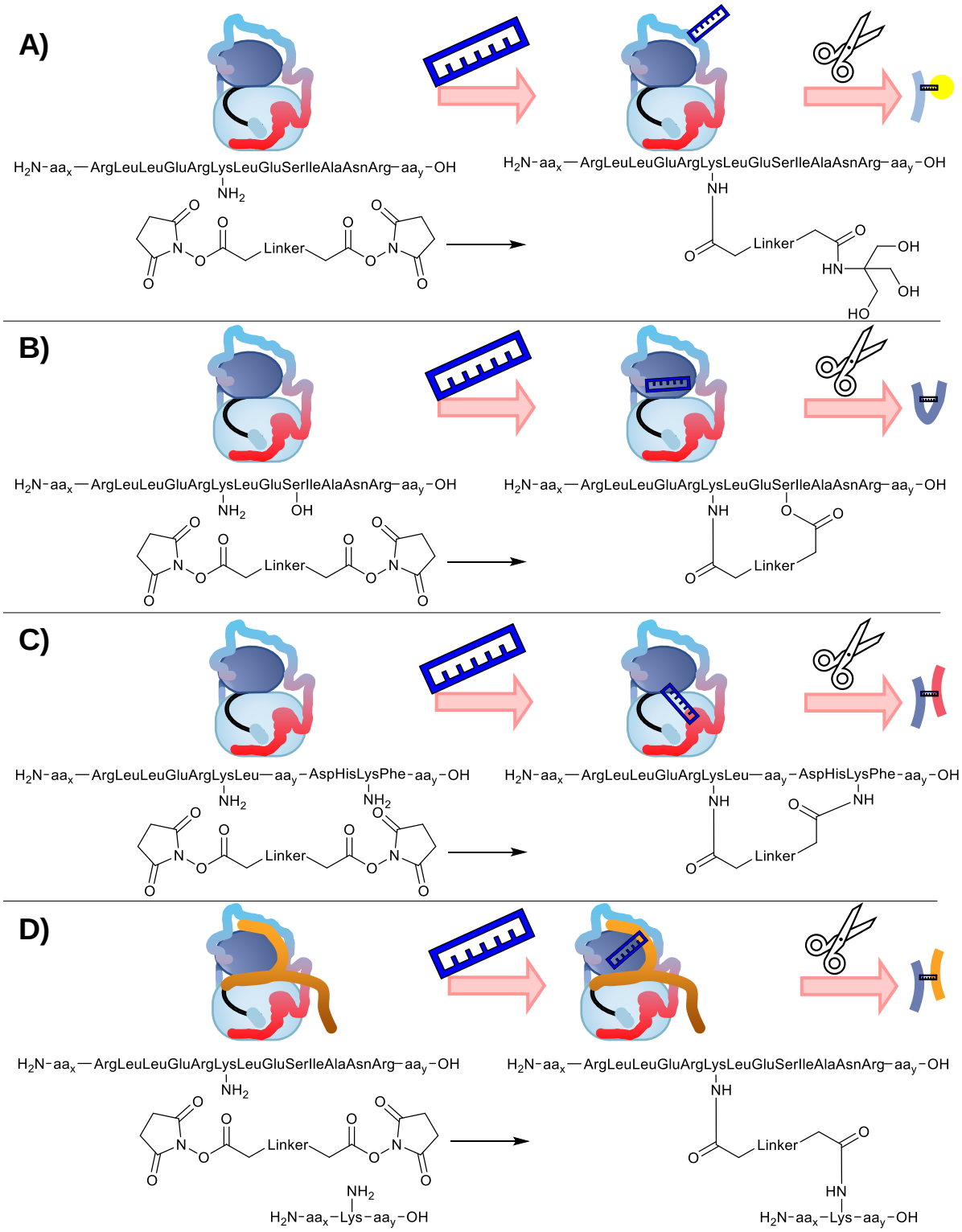


Figure 1-30 Different types of crosslinks

Different crosslinks generated from a monomeric and a hetero-dimeric protein complex. Panel A) shows a monolink generated by the NHS-ester crosslinker reacting with a quenching reagent, tris. Panel B) shows an Intra-peptide crosslink generated by at intra-protein crosslink lacking a cleavage site to generate the two peptides. Panel C) shows an Intra-protein crosslink. Panel D) Inter-protein crosslink.

Determining if and to what extent a residue is reactive with respect to NHS-ester is governed by the solvent accessibility and the local environment of the side chain. Muraoka *et al* divided this local environment into three factors: the number of hydrogen bonds the side chain was involved in, the estimated pKa value of the side chain, and the electrostatic energy.[200]

The primary discriminator for determining if a residue can generate crosslinks during the in-solution reaction is asking if the sidechain is solvent accessible. If a residue's side chain is never solvated for some or all of its lifetime across the ensemble of conformers, then it won't ever come into contact with a crosslinker molecule to react. Computational and experimental studies have shown that the rates of NHS-ester reactivity are governed by nucleophilicity of the side chain which is determined by hydrogen bonding, electrostatics and the local pKa.[201] The presence of neighbouring residues, like arginine under the correct pH, can catalyse the reaction of a lysine residue with NHS-esters.[202] In addition, proximal crosslinked residues can alter the reactivity of residues in an analogous manner.[193]

Table 1-5 Types of crosslinks generated

Crosslink type	AKA	Reactive moiety 1	Reactive moiety 2	Peptide
Monolink	Type-0; dead-end	Reacts with protein	Reacts with water from the buffer, buffer components or quenching reagent	Single peptide
Intra-peptide crosslink	Type-1	Reacts with protein	Reacts with protein within digestion sequence	Single peptide
Intra-protein crosslink	Type-2	Reacts with protein	Reacts with protein within digestion sequence	Two peptides
Homotypic		Reacts with protein	Reacts with separate protein at the same or similar site	Two symmetrical or semi-symmetrical peptides
Inter-protein crosslink	Type-2	Reacts with protein	Reacts with a different protein	Two peptides

The problem of homotypic crosslinks generated from protein oligomerisation and confusion in the analysis of crosslinks generated from these samples is solved by using isotopically labelled proteins.[203] This overcomes the limitation of not being able to deconvolute intra-protein monomer generated and inter-protein dimer generated crosslinks (Figure 1-34) as stable isotope-labelling of one of the two monomers generates 'light' and 'heavy' peptides.

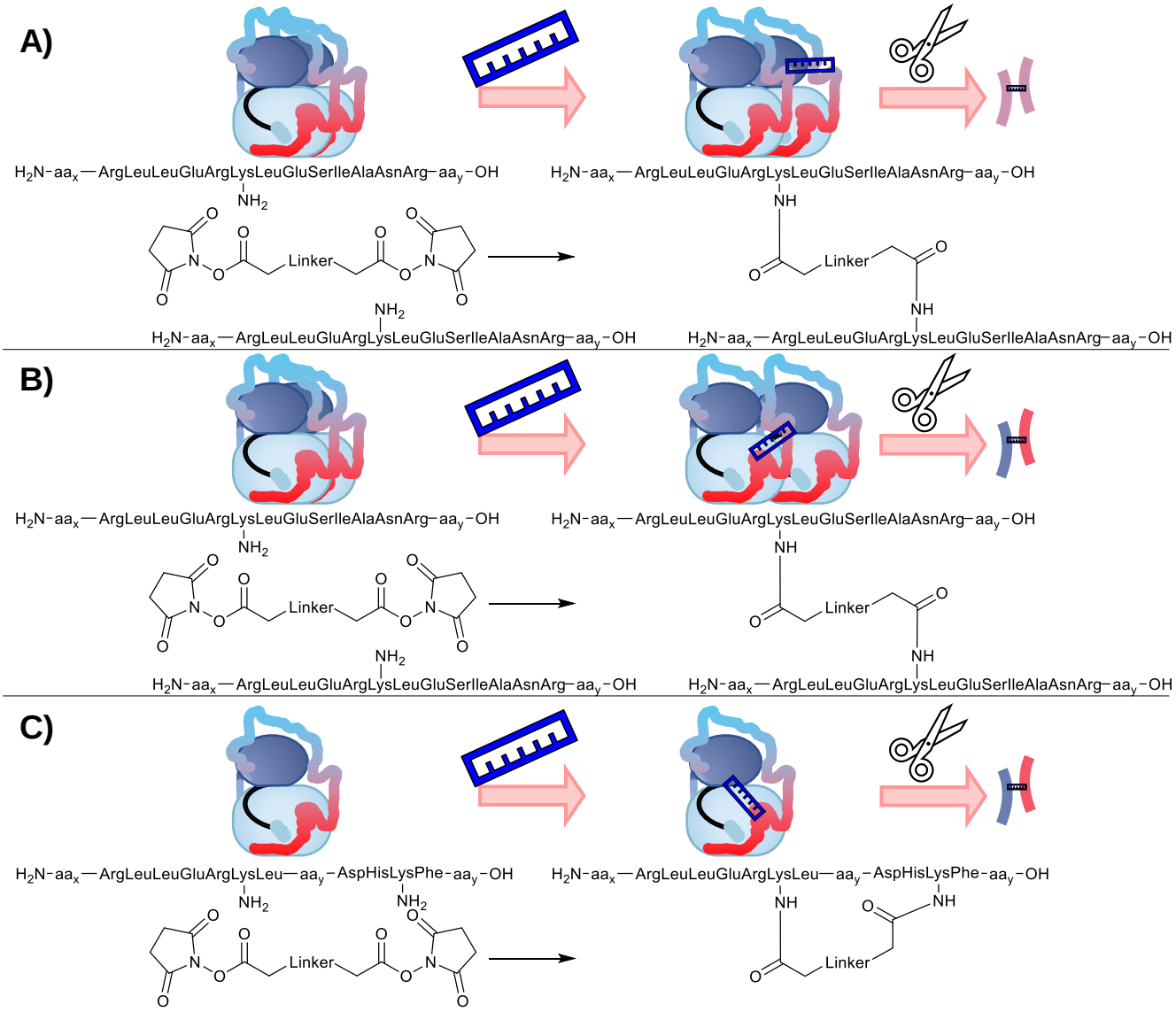


Figure 1-31 Different crosslinks generated from homo-dimeric proteins

Different crosslinks generated from homo-dimeric protein complex. This shows the range of crosslinks from a homo-dimer interaction, without isotope labelling, potentially some inter- and intra-protein crosslinks are identical. Panel A) where a symmetrical or homotypic crosslink is generated by an identical peptides crosslinked from a homo-dimeric protein complex. Panel B and C show the same crosslink can be generated from an inter-protein crosslink of a homo-dimeric protein complex and an intra-protein crosslink of a monomer.

1.4.1.3.2.2 Rationalising crosslinking reactivity.

Whether a crosslinker forms a monolink or crosslink is dependent on the distance to the nearest reactive side chain and thus the dynamics of the protein or complex. This is exemplified in Figure 1-35, summarising a theoretical pathway for a reaction of two protein residues with an NHS-ester bifunctional crosslinker. This highlights the challenges and complexity of XL-MS experiments, and the need for careful control of experimental conditions

to ensure the validity of the data obtained. In order to ensure the quality of XL-MS data, to allow for interpretation of the residue reactivity to crosslinking, protein digestion must first be optimised (near complete coverage is needed) to be able to observe all possible crosslinkable sites. Second, appropriate structures must be used for comparison with XL-MS data. When working with partial X-ray structures or homology models for one or both of the proteins, this requires the generation of hypothetical conformations/structures, calculation of the crosslink distances, and the elimination of hypothetical conformations. This results in an understanding of which residues should and could have reacted for the hypothetical conformations, and is often underpinned by MD and diverse structural methods. This complexity highlights why using XL-MS to validate generated structures of the larger complexes remains challenging.

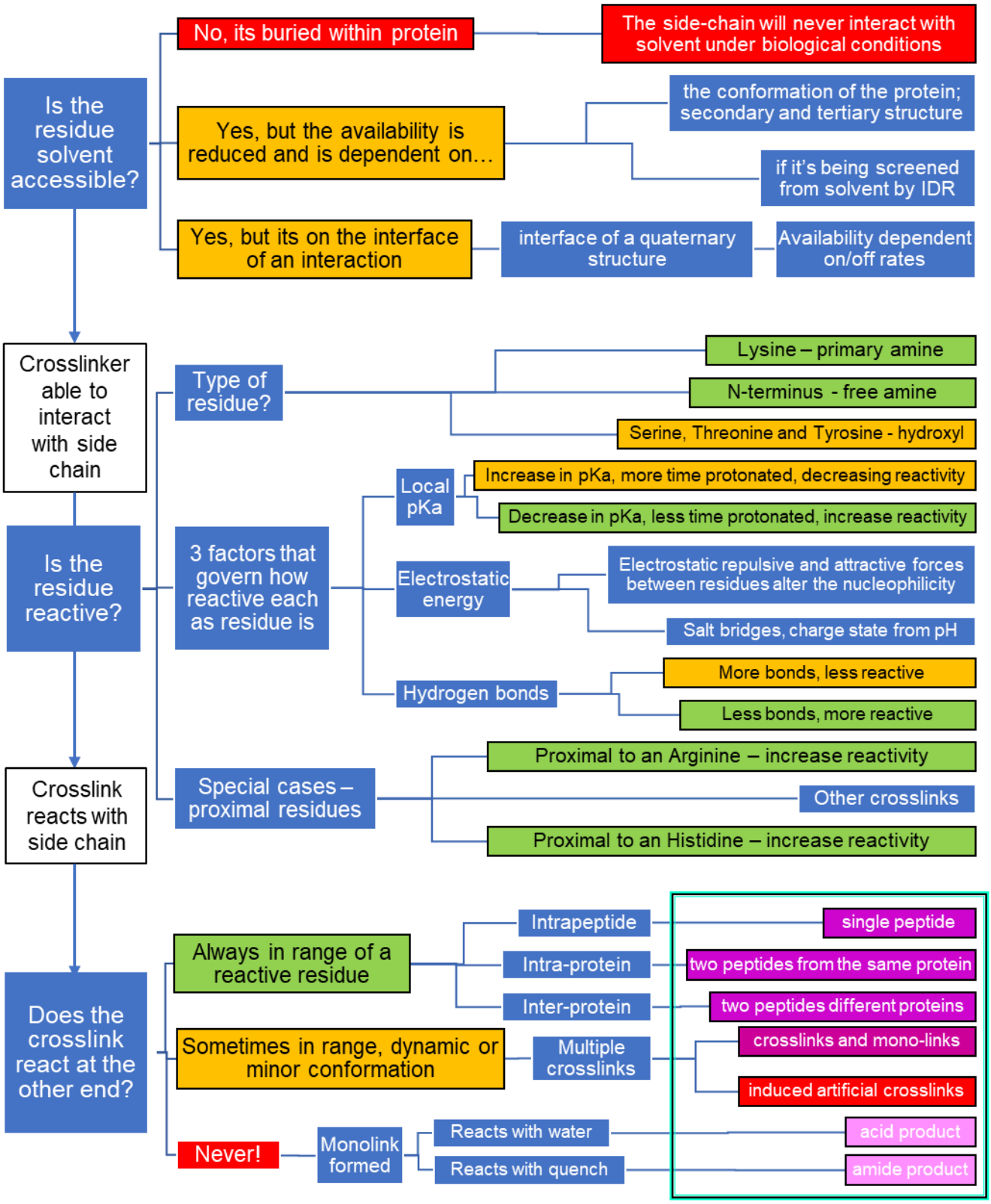


Figure 1-32 Flowchart to understand the path to reactivity of a crosslink

The flowchart summaries theoretical pathway an NHS-ester crosslinker takes to generate various types of crosslinks which starts with determining if first residue is accessible to the crosslinker. Blue boxes are questions and conditions, amber boxes crosslinks occur by with reduced probability, green boxes crosslinks occur by with increased probability and red no reaction.

Hogan observed the there was no correlation between the total ion intensity of a crosslinked peptide measured by MS and the distance between the resides crosslinked.[204] This in keeping with the theme that even for residues that are the ideal distance to react and which are both solvent accessible, factors such as local pKa and hydrogen bonds can prevent the formation of a crosslink or severely the reduce the rate of reaction. Bandyopadhyay and Kuntz have also shown that distance was not enough and that modelling pKa, electrostatic and surface area calculations into a rate of reaction model could help rationalise 'in range' residues not forming crosslinks. For structural proteomics of dynamic systems precisely where and how often a crosslink occurs can give insight into the system.[205] Quantitative XL-MS using isotope labelling, either of the crosslinker or protein, is required to be able observe changes with statistical analysis.

Hogan found that the relationship between residue-residue distance and number of crosslinks found (CSMs) was only proportional after normalisation of the reactivity of the residues crosslinked using mono-links to normalise the reactivity of the lysines.[204] As the recovery of crosslinks and mono-links from a reaction cannot be 'perfect' (i.e. coverage of the protein can be lost to poor digestion coverage, idiosyncrasies of peptide isolation, degradation or modification etc.) this data must be averaged over multiple reactions and measurements then both the information about which residues form crosslinks between themselves and the number of these crosslinks found for each unique residue-residue crosslink must be used; they called this the crosslink spectrum match or *CSM*.

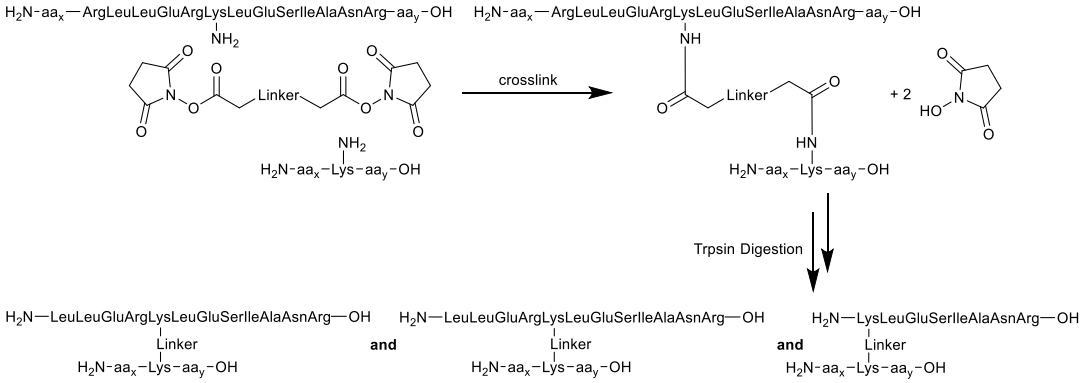


Figure 1-33 How one crosslink can be found across many peptide species

A single unique crosslink LysX-LysY, after trypsin digestion, found in 3 unique peptide configurations.

This is complicated by the fact that the enzymatic digestion of proteins can produce differing lengths of peptides so that crosslinks between LysX and LysY may be found within varying peptides. These must be grouped for analysis, (Figure 1-36). The same is true for a crosslink reaction that gives a mono-linked protein. Thus, after the alignment of crosslinks and mono-links, following the process of the crosslinking reaction, measurement by MS and data analysis must be used to locate the sites of mono-link labelling in the crosslinking reaction and the number of times that label is found, regardless of the peptide it is found within. The count of mono-links can be used to estimate a rate of reaction and thus a reactivity constant owing to the local environment of the residue from three properties that Hogan and Bandyopadhyay and Kuntz described that is averaged over the timescale of the reaction.[204 205] This is the mechanistic basis from which we can infer that a high yielding monolink and crosslink is reactive and accessible to infer solvent accessibility. The three properties can be estimated from X-ray, Cryo-EM or NMR structures or computational models of the proteins in the crosslinking reaction. Finding the all-unique crosslinks and mono-links and the count or number of times they occur is critical to understanding the reactivity, and thus solvent accessibility and conformation.

1.4.1.3.2.3 pH and temperature affect crosslinking

Crosslinking of a model protein at room temperature and at pH 7.4 for 45 minutes with 100-molar excess of BS3 crosslinker found that 20% of the measured crosslinked species resulted from reaction of crosslinker with the side chain of hydroxyl-containing amino acid and the remaining 80% came from lysine or the N-terminal amine. Serine accounted for 12.5%, tyrosine 4.3% and threonine 3%.[206] This 80/20 split for amine and hydroxyl reaction products can be rationalised by the reaction mechanism give in Figure 1-32.

The reaction of an NHS ester crosslinker with the ε-amine lysine is pH-dependent as the reaction requires the non-protonated amine. At low pH, the amine is more protonated but at a higher than than is optimal, the increase in the availability of the hydroxide ion increases the rate of crosslinking reagent hydrolysis. In the aqueous environment of protein buffers the hydroxyl and phenol side chains of serine, threonine, and tyrosine are also poorly nucleophilic. One would expect the higher reactivity of tyrosine, with its lower side chain pKa, to generate more crosslinked species; however, tyrosine is often found located within the hydrophobic core of folded proteins and is therefore rarely modified. As temperature also effects the rate of

reactions of hydrolysis and of lysine or hydroxyl crosslinking, the proportion of crosslinked products is dependent on temperature and pH.

1.4.1.3.3 Generation of measurable peptides

The crosslinking reaction is often optimised through monitoring the results of the reaction using one-dimensional gel electrophoresis (SDS-PAGE). A successful reaction shows the emergence of new protein bands at a weight range corresponding to the complex being reacted. If this new band doesn't occur or bands emerge that show oligomerization/non-specific or artefactual interactions, the reaction conditions can be adjusted.[186 207] Conditions such as temperature, time of the reaction, buffer composition of the reaction, buffer conditions especially pH, and ratio of protein concentration to crosslinker concentration are commonly adjusted.

1.4.1.3.3.1 Sample preparation for mass spectrometry

After the quenching of the crosslinking reaction, proteins must undergo the sample preparation protocol to generate peptides that can be measured by LC-MS/MS. This follows a generic workflow following established protocols for bottom-up MS structural proteomics for amino acid sequence resolution of protein modifications.[208]

1.4.1.3.3.2 Optimising reaction conditions

Table 1-6 highlights the variables available for modification whilst optimising the crosslinker reaction. As with all chemical reactions, the reaction conditions that crosslinking is performed in can have a substantial effect on the outcome of the reaction and quality of associated data. The ratio of crosslinker to protein of a reaction is given as a fold Molar excess (Figure 1-37).

$$\text{fold Molar excess} = \frac{[\text{Crosslinker}] - [\text{total protein}]}{[\text{total protein}]}$$

Figure 1-34 fold molar excess equation

Table 1-6 Crosslinking reaction variables

Condition	Predicted effect on NHS-ester crosslink	Protein/complex stability
Temperature of the reaction	Rate of reaction is slowed at lower temperatures	Lower temperatures less binding, higher more unstable?
Time	Longer the reaction more reactions can occur, potentially less monolinks as they react to form crosslinks.	Longer the reaction, more wider search space as minor conformers sampled
Buffer comp	Crosslinker reactive? Small amount maybe tolerated but buffer-mono links may disrupt analysis	Stability of protein and/or complex; more unstable more non-specific crosslinks.
pH	NHS-esters rate of reaction will be dependent on pH as the protonation side chains is pH dependent.	Stability of protein and/or complex
Ligands/ratio of proteins	-	Drug causing conformation change, full saturation, on/off rate
total concentration of protein	Less crosslinks generated from less reactive residues.	More concentrated more non-specific contacts?
concentration of crosslinker	More crosslinker, more reactions until non-specific reactions occur	Saturation of protein made alter conformation
Preparation of the protein prior to complex	More varied crosslinks seen if the protein is in an equilibrated state	Stability of protein and/or complex

1.4.1.3.3.3 Cleavage and capping of disulfide bridges

After quenching the crosslinking reaction and prior to preparation of protein samples for enzymatic digestion, they undergo a reduction and alkylation of cysteine residues, commonly using dithiothreitol and then Iodoacetamide**.** This is to break then cap disulfide bridges to prevent their reformation. Omission of this can lead to loss of cysteine containing peptides.[209] But additional processing and sample handling leads to the loss of peptides, e.g. to non-specific interactions of the peptides with sample tubes, and increases the potential for contamination.

1.4.1.3.3.4 Digestion by proteases

For bottom-up proteomics, protein samples are enzymatically digested into peptides using enzymes that cleave the protein at specific residues to give a predictable peptide that can be searched for using a peptide search engine software. Trypsin cleaves C-terminal to Arg or Lys and is commonly used, but the choice of enzyme will affect the sequence coverage seen.[210] If Arg or Lys are too common, peptides that are too short will be generated, but if Arg or Lys occur infrequently in a protein sequence, the peptides produced will be too long and won't ionise or they will be too large for the mass analyser

range. In silico, test digestions can be used to help find the optimal enzyme, and it should be noted that a number of crosslinked resides could alter the pattern of digestion. Optimising the use of digestion additives such as surfactants or chaotropic agents to solubilize and denature challenging proteins can be done to increase sequence coverage.

1.4.1.3.3.5 Multi-stage digestion

Digestion strategies have been developed that using multiple proteases with differing specificities in combination to reduce incomplete digestion.[211] This is in part a response to NHS crosslinking modifying Lys so that these residues are unable to serve as a site of proteolysis.

1.4.1.3.3.6 Enrichment of crosslinks

To increase the measurement of the desired low-abundant cross-linked peptides, various protocols have been developed to enrich the proportion of crosslinked peptides. A rule of thumb for recombinant protein complexes containing a few proteins, is that an enrichment step can be skipped if good separation of peptides by LC/MS/MS is achieved. Large complexes or protein complexes co-immunoprecipitated from lysate or tissue can benefit from crosslink enrichment despite the additional material required for the protocol. Commonly an ÄKTA protein chromatography system with a size exclusion chromatography (SEC) or Strong Cation Exchange (SCX) column suitable for peptides or small proteins is employed to enrich crosslinked peptides. These strategies work because crosslinked peptides will be larger and with a larger net charge than non-crosslinked peptides. Solid phase extraction pipette tips have also been shown to be effective at increasing the proportion of crosslinked peptides.[211 212 213]

1.4.1.3.3.7 Desalting

Protein buffer components, such as non-volatile salts like sodium phosphate or sodium chloride, or solubilizing agents and detergents, that are critical to protein/complex stability in solution can significantly impact MS sensitivity. The removal of non-volatile salts from digested peptides that have carried through from the buffers used to stabilise the protein is critical before MS measurement.

In addition to suppressing the ionisation of peptides thereby reducing the sensitivity of measurement, the failure to remove these from the peptide samples before measurement can lead to the peptides-salt adducts that disrupt the apparent mass to charge of a peptide and thus the data analysis

of the crosslink reaction. The polymeric nature of detergents causes spectral complexity by spreading signals over multiple m/z values.[214]

Desalting and peptide concentration can be performed either offline using solid phase extraction, dialysis or centrifugal filters or in-line with a desalting cartridge. In-line desalting allows minimally cleaned samples to be directly loaded onto LC-MS/MS with a desalting cartridge prior to LC.[215]

1.4.1.3.4 Measurement by LC-MS/MS

1.4.1.3.4.1 Introduction

The advent of modern orbitrap mass spectrometers, instruments capable of analysing complex and diverse target types, has meant that the measurement protocol for XL-MS is often straightforward.

1.4.1.3.4.2 Liquid chromatography

Prior to injection into the MS, peptides are separated using reversed-phase liquid chromatography (RPLC), which dramatically increases the identification of peptides because of the reduction in signal suppression from coeluting peptides. However crosslinked peptides will be on average larger than non-crosslinked peptides because two peptides are covalently linked. Longer peptides are hard to separate and poor separation by RPLC leads to a decrease in sensitivity so optimising the digestion protocol is critical to achieving good coverage of crosslinked peptides. Crosslinked protein samples can be digested either in-gel or in-solution. In-gel digestion allows for precise control of which species are digested, whether together or separately, and thus which are measured. In-solution digestion gives an “average” picture. There is a continued debate about the benefits and drawbacks of each approach. In-gel digestion is generally thought to give lower sensitivity because of the ‘loss’ of unrecoverable peptides. However this may just be because the increased amount of handling and thus loss to non-specific binding to plastic tubes.[216] The loss of sensitivity could be due to larger digestion enzymes that can’t penetrate gels effectively.[217] Novel protocols to mix the strengths of the both methods are being developed.[218] Newer in-solution methods, such as using digestion enzymes immobilized onto a solid support or filter-aided sample preparation have been developed to increase the speed and ease of digestion.[194 195] After digestion, the majority of peptides produced do not contain crosslinks with >80 % of peptides remaining un-labelled.[219] Mass spectrometry measures the mass to charge (m/z) ratio of ions. For conciseness, the mass spectrometric theory of

measurement of peptides by Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometric (LC/ESI-MS/MS) using quadrupole-orbitrap mass spectrometer will be covered. Tandem MS/MS is used to fully identify the amino acid composition, sequence and modifications of peptides. Its basic principle is that peptide ions are separated by their m/z in MS1, commonly a quadrupole, so that specific analyte ions can be fragmentated and the resulting peptide fragments can be analysed in MS2. The resulting fragmentation spectra yield insight into the amino acid sequences of the peptides and thus their parent proteins. This highlights why RPLC is used to prevent multiple peptides entering MS1 and co-fragmenting.[220] The manufacturer Thermo Fisher Scientific commercialised the use of Quadrupole-Orbitrap-mass analysers developed by Makarov.[221] Samples of ≈ µL volume containing ≈ ug peptides are injected onto an in-line desalting column where then they are separated using C18 RPLC to improve resolution prior to ionisation. Gradients are optimised for peptide separation balanced against diluting the sample and degrading the signal.

1.4.1.3.4.3 Ionisation

As separated peptides exit the column they are transferred to the mass analyser where the analyte solution is ionized using a nano-ElectroSpray Ionization (nano-ESI) source. Electrospray ionization is a *soft* ionization method as it does not impart enough energy into the sample undergoing ionization to cause fragmentation of the ionised molecules by the breaking of weak covalent bonds.[222] The analyte solution exits the capillary tip where it forms a plume of charged droplets. A nano-ESI source dispenses with the need for a flow of inert gas to aid droplet formation by using a finer capillary tip. A high voltage with a large potential difference is generated with the capillary acting as the anode, and the inlet of the mass spectrometer as the cathode. Thus, charged droplets are produced at atmospheric pressure which forms gas phase charged ions. The mechanism by which de-solvated charged peptides are produced from solubilised in droplets is contested.[223] How peptides are ionised does affect the composition of peptides observed by the mass analyser.[224]

1.4.1.3.4.4 MS^1

Quadrupole-Orbitrap mass spectrometers are sufficiently sensitive to support the resolution required for the measurement of the peptides produced by a crosslinking reaction.[225] [226] Newer mass spectrometers with higher resolutions are allowing the analysis of more complex samples, such

as live cell crosslinked proteins. Recent developments in Trapped Ion Mobility Quadrupole Time of Flight Mass Spectrometry bring another generation of breakthrough innovation that will support new crosslinking strategies.[227]

The Orbitrap Fusion series of Quadrupole-Orbitrap MS manufactured by Thermo Fisher Scientific is representative of this class of ESI-MS/MS instrumentation. The ionised peptides are transferred and separated from neutral particles using an “s-lens” and beam guide. The quadrupole in radio frequency mode can transfer the entire ion current for a full scan of ions by the Orbitrap.[228]

The quadrupole acts as a mass filter, to select ions with a specific m/z for CID fragmentation in the high-pressure cell of the linear ion trap and then detection in the Low-pressure cell of the linear ion trap, for speed of measurement, or in the orbitrap, for high resolution measurement.

1.4.1.3.4.5 MS^2

The fragmentation of peptide ions is paramount to obtaining adequate peptide sequence coverage, unambiguous cross-link identification and identification of any PTM present on residues. For Collision-induced dissociation (CID)-type fragmentation in the Linear ion trap, a potential is applied to the precursor ions, selected at MS1, to accelerate the analyte ions to collide with an inert gas, in this case typically nitrogen, which transfers energy to precursor ions and breaks the weakest bonds of the peptide ion. This is usually the backbone amide bond hence this process results in the generation of a series of these fragment ions that exposes the peptide sequence when detected. Depending which side of the peptide carries the charge the ions are termed either b or y ions.[229] Higher-energy C-trap dissociation (HCD, aka beam-type CID sometimes unhelpfully termed Higher energy collision-induced dissociation) uses increased voltages, that the ion routing multipole cell is capable of producing, to deliver increased acceleration of the analyte ions thus increasing the kinetic energy of the collision between the inert gas and the analyte ions. In a “tribrid” type mass spectrometer, with a quadrupole, an Orbitrap, and a linear ion trap, HCD takes place in the ion routing multipole, and the subsequent detection takes place in the Low-pressure cell of the linear ion trap or in the orbitrap. The curved Radio Frequency quadrupole ‘C-trap’ is required to bunch the ions prior to injection into the Orbitrap which then measures m/z values. The higher energy fragmentation can break stronger bonds primarily causing the production of a-/x-fragment ion series in addition to b-/y ions. Post-

translational modifications that bear weaker bonds such as those found in phosphorylation or glycosylation can also be removed by fragmentation.[230] [231] [232] After fragmentation the ions are transferred back to the C-trap and then injected into the Orbitrap mass analyser for detection or detected in the linear ion trap. Fragmentation of bonds other than the backbone can occur.[233] Explanation of the process of how crosslinked peptides are identified in LC/ESI-MS/MS explains why contaminating proteins and peptides can grievously reduce the sensitivity of XL-MS. For XL-MS of recombinant proteins it is essential that accidental co-purification of abundant host proteins is prevented along with the target protein. Any proteinaceous contaminants, for example antibodies from enrichment steps, keratins from dust or skin or enzymes used in digestions, will generate peptides that compete for instrument time and prevent the sequencing of peptides from target proteins and low-abundant crosslinked peptides.[234] [235]

1.4.1.3.5 Data analysis

After the measurement of peptides, the analysis of crosslinked peptides is performed using a peptide search engine software platform. There are many widely-used open source platforms for crosslinked peptide search and analysis that have been designed for use with specific crosslinks or crosslinker types.[236]

1.4.1.3.6 Data validation

Understanding the confidence in the data generated is important to correctly interpret the dataset's significance and limits when validating a structural hypothesis.[237]

The conventional strategy to assess confidence in a peptide and protein assignment from an MS/MS dataset has been to use false discovery rate generated from a target-decoy approach, where decoy sequences are generated by randomising the peptide sequences found in the target database. Any match to the decoy database is therefore a false positive and the search parameters are tuned to reduce the rate of false positives below a desired threshold. The quality of a peptide to spectrum match, PSM, can also be assessed and scored against number of parameters to help assess the probability that the determined peptide generated the spectrum in question.[238] On proteomic scale experiments, errors in peptide assignment and thus protein assignment can propagate.

Error estimation of proteomic XL-MS experiments has been shown to be underestimated.[239] Although still controversial, separation of crosslinks by type has been shown to give the best error estimation.[109]

For XL-MS studies that are focused on the structures of proteins and complexes and not just the interactions, PDB structures have been used to estimate the quality of crosslinks found in the study. The approximate ratio of cross-linked peptides that comply with the maximum Cα-Cα distance the crosslinker can span compared with the number of crosslinked peptides that exceed that distance can be used to assess the quality of the cross-links detected. This has also been applied to proteomic XL-MS experiments.[240] Applying error rate estimation to investigate small complexes and single proteins is complicated as fewer decoy hits are generated so the statistical power of the FDR estimation is reduced.

MeroX is a designed for the use with cleavable crosslinkers and allows for automatic FDR estimation using a target–decoy approach. It also calculates score cutoffs for different significance levels from the score distribution of target and decoy hits.[226] Cleavable crosslinkers do not necessarily improve error rates[107] but do improve coverage of crosslinks found by improving fragmentation.[241]

In addition to software platforms for crosslinked peptide search and analysis, peptide search engines designed for peptide sequencing, protein identification and quantification of proteomics mass spectrometry can search for peptides from crosslinked proteins to orthogonally search and sequence peptides.

1.4.1.3.7 Distance restraint software and further analysis

As the theoretical maximum distance a crosslinker can span is known it is possible to thus calculate the maximum distance between two amine groups (or other reactive side chains) for a reaction to occur, hence distance restraint data is generated from the crosslinker.

For lysines this would be the Nz-Nz distance, hence a protein structure can be evaluated: if the Nz-Nz distance falls within the theoretical maximum distance spanned by the crosslinker then a structure can be said to match the crosslinking data. If the distance is exceeded, then either the structure the crosslinks were modelled onto, or the reaction, may be compromised. This is also true for minimum distances too. As side chains can be dynamic or missing in X-Ray structures the Euclidean distance, between Cα of residues, is often used in place of side-chain distances with additional

distance added to account for the side chain. Further distance is also added to as a tolerance factor for dynamics of the backbone.

Merley et al found, using XL-MS experiments, MD simulation and database searches, that for disuccinimidylsuberate/bis(sulfosuccinimidyl)suberate, an eight carbon linker of 11.4 Å length, upper Cα-Cα distances of 26–30 Å were acceptable.[242] Software tools to visualise and measure the distances between crosslinked residues have been developed.[243] The ease of measuring the Cα-Cα Euclidean distances for each crosslink allows for the evaluation of larger datasets of crosslinks on larger structures. Discussion of non-Euclidean software tools will be covered in Chapter 3. Homology modelling, the generation of an atomic resolution of an all-atomic model of a protein from a template protein with homologous protein sequence and predicted structure, has commonly been used to generate models for a protein structure. A complete structure is required to annotate crosslinks on. Recently, the artificial intelligence structure prediction software AlphaFold[244] and the related ColabFold[245] have dramatically reduced the computational requirements to generate Homology all-atom models.

Evaluating protein structures or PPI interfaces using Cα-Cα Euclidean distances measured on previously generated Xray structures or models is less challenging than calculating solvent accessible surface Nz-Nz distance. For complex interfaces, that often present in PPIs of dynamic or disordered proteins the Euclidean distance will pass through the protein, erroneously presenting a crosslink of a shorter length than could occur.[246] The shortest distance between two residues a crosslinker could physically straddle is called the solvent accessible surface distance (SASD).[247] Crosslinks can also be used to guide molecular modelling of protein structures and complexes and structure prediction.[248 249] Webservers such as HADDOCK can even give basic binding conformations to aid hypothesis generation.[250] This generates a tension in scholarship; is XL-MS validating experimentally reported structures and homology models or are experimentally reported structures validating the distance restraints found by XL-MS?

All protein structural data is generated with an amount of artifice that is then validated by orthogonal methods. For example, protein-protein interfaces demonstrated by an X-ray structure are often the subject of binding assays where critical residues are mutated.[251] XL-MS is now routinely used to validate Cryo-EM structures.[252]

The MD simulation of complexes subject to crosslinking has shown that the alkyl chain crosslinkers have a distance preference and two, long and short,

stable conformers that disrupts the distance prediction and calculations.[253,254]

1.4.2 Molecular Dynamic (MD) simulations of kinases in dynamic systems

1.4.2.1 MD methods

1.4.2.1.1 Why is MD helpful to the structural biologist?

Molecular dynamics (MD) simulations are used for the characterization of the structure and dynamics of biological macromolecules using computational calculation. Molecular dynamics simulations for protein-protein interactions and protein dynamics help structural biologists hypothesise possible interactions that then can form the basis of experimentation. MD simulations serve as a 'bridge' between the ordered, static paradigm of crystallographic structures and the disordered, dynamic paradigm of in-solution biochemical techniques.[257 258 259]

High quality of examples of MD simulations being used to bridge XL-MS distance restraint data to Cryo-EM and X-ray structures are regularly being performed. Of interest for proteins with IDRs, MD simulations can be used to help explain dynamic phenomena observed in crosslinking data but not in the static structural methods.[256]

1.4.2.1.2 AMBER

AMBER is a suite of biomolecular simulation programs that have been employed across a diverse range of bio-molecular questions relating to the simulation of proteins.[260] AMBER has been used successfully to interpret the in-solution dynamics of kinase structural data, especially the modelling of activation loop mobilities in response to changes in phosphorylation.[261] Molecular Dynamics is a simulation technique in which Newton's laws of motion are solved for a set of atoms, interacting through a given potential under specific conditions over time. While in principle all the thermodynamic and kinetic properties of the system can be obtained from a trajectory (i.e., the coordinates and velocities as a function of time) in practice this is only possible for simple, rapidly converging systems (e.g., small molecules in gas or liquid phase). To reduce the complexity of all-atom simulations, water can be considered "implicitly", i.e., its effect on the protein, such as screening charges, are included by modifying the protein-protein interaction potential. An AMBER simulation requires the input of a protein's atomistic coordinates, usually taken from a crystallographic structure wherein the topological

information is defined. This is the atom connectivity, atom types, residue names, and charge of the atoms present. The topology for the common amino acids as well as N- and C-terminal charged amino acids, and bio-molecules such a DNA or sugars are pre-determined. The next requirement is the definition of the simulation's Force field. Parameters for all of the bonds, angles, dihedrals, and atom types in the system and critically the solvent water are thus defined. Once all these are in place the calculations and simulations can be performed.[262] Empirical energy functions describe the energy of the molecule as a function of the atomic coordinates.[263] Most experiments are carried out at a constant temperature, whereas the Verlet, based on Newton's laws of motion, represents a microcanonical ensemble where the number of particles, the volume and total energy are kept constant, not the temperature. All simulations that use Langevin dynamics to maintain the temperature, 'constant-temperature, constant-volume' (NVT) are the preferred approach to conformational focussed simulations whereas ensemble 'constant-energy, constant-volume ensemble' (NVE) uses an additional of a friction coefficient to maintain a desired temperature.[246 265] Hydrogen mass repartitioning allows the increase in simulation time by redistributing some of the mass from heavy atoms connected to hydrogen into the bonded hydrogens. This reduces instability-related errors produced from high-frequency hydrogen motion, thereby accurately representing a longer time frame.[266] An additional method to increase the efficiency of the MD simulation calculations is parallelisation. This allows the MD simulations, that incur high computational cost, to be performed efficiently, when processes are run in parallel. The MPI communication protocol allows the division of the MD simulation across processing Graphical Processing Units.[267]

2 Can Chemical crosslinking mass spectrometry elucidate the binding interactions of kinases with intrinsically disordered regions and IDP interactors?

2.1 Abstract

Kinases function as signallers within the cell; using phosphorylation of the proteins with which they interact as molecular switches to perform functions within the cell. Outside of their core catalytic domain, kinases can contain intrinsically disordered regions (IDRs) as part of their structure[75] or can interact with intrinsically disordered proteins (IDPs).[268] The disorder, or lack of stable structure, of IDRs and IDPs can result in a protein or region being without any structure, or dynamically processing between stable structures, giving rise to an ensemble of conformations.[269] Both kinds of disordered concepts are found in kinases and their protein-protein interactions (PPI) partners and disease pathways.[32]

As structural biologists, we are charged with generating models of biological phenomena that attempt to explain the structure of proteins and PPIs. However, IDRs resist the mapping of their structure and interfaces using crystallographic methods. Chemical crosslinking mass spectrometry (XL-MS) is an in-solution technique that uses a bi-functional reagent to generate distance restraints between the reactive side chains of residues that can be mapped to previously generated structures or models.[190] For larger, less dynamic complexes, crosslinked residues can be directly interpreted as a stable contact; however, in dynamic systems, the reactivity of residues can vary to the extent that the number of crosslinks seen is not proportional to spatial proximity and contact time.[248] A monolink-labelled residue occurs when it has reacted with a crosslinker but hydrolysed at the other reactive site, giving contextual information. We built on the work of Hogan, who previously demonstrated with a two-state complex that the mono-linking reactivity of individual residues can be normalised to assess the relative cross-linking propensity of pairs of residues,[204] to show that surface accessibility information could be extracted post-hoc when comparing the protein of interest with and without an interactor.

Here we take this further, investigating whether XL-MS can give insights into dynamic, kinase PPIs by using reactivity normalised XL-MS on the extensively studied testbed interaction between the oncogenic mitotic kinase AurA and an IDP interactor TPX2.[270] Our question is to what extent intrinsic disorder limits the application of XL-MS in this context. We then applied this

approach to the high value oncogenic and complex interaction between AurA and N-Myc TAD.

2.2 Introduction

Dynamic kinase proteins with disordered regions are complicated and resist the structural biologist's methods to explore them, and this is a challenge in the context of developing therapies for cancers that feature the mis-regulation of kinases.[268 271] In this chapter, we aim to explore whether chemical crosslinking mass spectrometry (XL-MS) can provide structural information about the protein-protein interactions (PPIs) involving intrinsically disordered regions (IDRs) or intrinsically disordered proteins (IDPs). This work will be focused on testing and generating hypotheses regarding the structure of kinases, work that is adjacent and foundational to the structural insights required for the development of small molecule and peptidomimetic modulators of oncogenic kinases. XL-MS is an averaging technique that provides information about the distances between residues; unlike crystallographic techniques, it can report in-solution conformational information and species' composition of in vivo and in vitro protein complexes. The complexity of the sample will impact the length of time taken to reconstruct the protein structure and conformation and will also affect the accuracy of the reconstruction. Direct investigations of disordered proteins using XL-MS have been limited as the disentangling and analysis of such proteins has proven complicated, without sufficient orthogonal information.

Protein kinases with IDRs and that interact with IDPs lend themselves to this type of analysis thanks to the wealth of analogous crystallographic and biochemical data, and the array of orthogonal in-solution methods available to build testable models from.

AurA±TPX2 has been extensively studied using crystallographic, FRET, and molecular dynamics (MD) methods, for example in the study of cancers such as lymphoma and breast cancers.[177, 272] Fortuitously, many crystallographic structures of the interface of the IDP TPX2 1-43 have been produced, alongside extensive orthogonal in-solution methods to characterise the molecular basis of the interaction. This complex will serve as our testbed to investigate if functional data can be produced using XL-MS. This validation will be performed in concert with an exploration of AurA+N-Myc. Only a partial structure of AurA+N-Myc has been produced and this seems a sensible first target to explore.

This work is divided between chapters 2 and 3. Chapter 2 is centred on optimising and validating the methodology manually, while chapter 3 takes the crosslinked datasets and explores the use of MD simulations in tandem with orthogonal data and semi-automatic structure evaluation. This will allow us to validate the use of XL-MS to investigate proteins with disordered regions that interact with IDPs in the context of drug discovery.

2.2.1 Using the AurA+TPX2 complex as a test model for XL-MS in kinase IDRs

2.2.1.1 Introduction to our research questions

2.2.1.1.1 Step 1 – ascertain whether we can we use XL-MS to generate structural data with AurA+TPX2

Our research question is divided into two steps. Firstly, we need to ensure that reliable crosslinks can be generated through the reaction of the crosslinker with AurA+TPX2, or whether the dynamic nature of the complex makes it too reactive to study in this way. In essence, this step is concerned with ascertaining if a certain interaction is of high enough affinity to be measured by this particular technique. This is necessary because the disordered and dynamic regions of AurA and TPX2 may react in a way that generates a majority of non-specific crosslinks. This first requires the optimisation of reaction conditions.

Once we have shown that AurA+TPX2 can generate crosslinks that are specific and form biologically relevant in-solution species, we then need to examine if these crosslinks produce residue-to-residue distances that conform to the current understanding of the complex. We will also need to ascertain whether the resolution of the structural data obtained is sufficient to generate useful insights into the complex, or if there is the potential that crosslinks from the disordered regions cloud the quality of the data.

2.2.1.1.2 Step 2 – collect structural data for the high value target AurA+N-Myc

Once we have shown that our crosslinks are specific and of sufficient resolution to perform the analysis, we will attempt to map the interactions of the unstructured regions of AurA KD and N-Myc TAD. We will apply the protocol developed using AurA+TPX2 to investigate the regions of N-Myc that bind to AurA that are unstructured in the existing crystal structure.

2.2.1.2 How are we testing our research question?

2.2.1.2.1 Theoretical basis of our approach

Due to the amount of previous scholarship investigating the structure of AurA+TPX2, supported by biochemical assays such as kinase assays and peptide array pulldowns, the protein constructs have already been optimised. The optimised minimal domain construct and binding conditions of AurA+TPX2 have been shown by a higher throughput fragment screening, performed by Bayliss lab members,[86 273] which will be directly adapted for our needs. This construct, commonly named the crystallographic construct or double cys mutant of AurA, has been optimised by the removal of the majority of the NTE of AurA (residues 1-118), and residues outside the minimal binding domain of TPX2. It has also been shown that two Cysteine residues in AurA need to be mutated to prevent non-specific oligomerisation.[274] Hence for this study will used the crystallographic construct of AurA in combination with the minimum binding domains of TPX2 and N-Myc. Where possible and available constructs without purification tags were used.

2.2.1.2.2 Selection of the reagents and methods for XL-MS of kinase-IDR

2.2.1.2.2.1 The chemical crosslinker and MS

Chemical crosslinkers are bifunctional labelling reagents that are sometimes known as a molecular rulers because their fixed length gives a distance between the residues they react with. The most commonly used is an NHS-ester, which primarily forms crosslinks by reacting with the primary amine of a lysine side chain to form an amide bond. These also react with hydroxyl side chains of serine, threonine and tyrosine.

The slow rate of reaction of NHS-esters is thought to allow for the generation of crosslinks between residues that, on average, would be out of distance. It has also been theorised that slower reacting crosslinkers may generate crosslinks from transient or rare conformations by a process known as kinetic trapping.[275]

Photoreactive crosslinkers only react after they have been irradiated, and may provide a better snapshot of the protein and complex conformer. However, their use is analytically challenging. Here we sought to use NHS-esters first, with a view to moving to photo-reactive crosslinkers or similar if we encountered challenges. In this case, we will substitute the selected NHS-ester crosslink for a photoreactive crosslinker, if the reaction conditions

cannot be adjusted to generate stable, specific, explainable species, as analysed by SDS-PAGE gels and Western blots. These may be caused by crosslinks from regions that are buried, or IDR residues that are generating an outsized number of crosslinks.

The 12.5 Å spacer arm mass cleavable crosslinker DSBU has been used in a wide range of target and experiment types with a number of published protocols. Mass cleavable crosslinkers were developed to help solve the N^2 search problem. This is briefly defined as the complexity of searching for crosslinked peptides increasing quadratically with the complexity of the sample (see section 1.4.1.3.1.1).

The crosslink peptide search software MeroX was developed to directly search for the fingerprint ions found for the fragmentation of DSBU crosslinked samples. This allows for the straightforward identification of crosslinked peptides.

Our initial protocol for the crosslink reaction conditions and sample processing comes from Sinz *et al*. It advises adjusting the temperature, and using an excess of cross-linker to manipulate the crosslinked species generated, as shown by SDS-PAGE gel and Western blots.[197]

The sample processing, mass spectrometer and measurement conditions will all affect the quality and reliability of the data produced. The crosslinked protein reaction mixture is commonly processed using protocols originally developed for protein sequence and PTM identification.[276] In these, after the quenching of the reaction, the mixture can be processed and digested using an in-solution protocol, meaning that all protein species will be subject to enzymatic digestion and then measured. The alternative is to excise the desired bands after separation by SDS-PAGE gel. Although this latter method gives precise control over the species measured, it is prone to contamination from the increase in sample handling and protein coverage, and the fraction of residues observed can be arbitrarily suppressed, as the gel can retain peptides.[277]

For the crosslinking of larger complexes such as RNA polymerase II or the ribosome,[256 278] it is also common to enrich the crosslinked peptides, as non-crosslinked peptides will make up most of the digested peptides from a crosslink reaction.[216] As well as increasing the sample requirements, AurA+TPX2 is likely a small enough molecule that this step will be unnecessary, so long as LC separation is sufficient, and the resolution of the mass spectrometer is reasonable. Modern Orbitrap instruments have the

resolution to measure a mixture of crosslinked and non-crosslinked peptides. This approach (i.e. not enriching the crosslinked peptides) has already used to obtain structural insights into the IDP tumour suppressor p53.[279]

2.3 Methods

2.3.1 General

All reagents were purchased from Sigma Aldrich, VWR, Fisher Scientific, Thermo Fisher Scientific or Promega UK, and were graded suitable for molecular biology. Iodoacetamide, dithiothreitol, trifluoroacetic acid, ammonium bicarbonate and the solvents used for mass spectrometry were graded suitable for mass spectrometry.

Disuccinimidyl Dibutyric Urea was purchased from Thermo Fisher Scientific (DSBU; CAS number: 1240387-33-0; Catalogue number: A35459).

Proteases were purchased from Promega. The proteases used were Trypsin Gold (Mass Spectrometry Grade; Catalogue number: V5280), Chymotrypsin (Sequencing Grade; Catalogue number: V1061), and Glu-C (Sequencing Grade; Catalogue number: V1651).

2.3.2 Buffers

Buffers were produced as 10X stocks without β-mercaptoethanol, glycerol or Tween-20, diluted to 1X, additives included and used within 24 hours.

Buffer	Formulation
Washing	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.02% (v/v) Tween-20
Aurora-A	5 mM Tris (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, 2 mM β-mercaptoethanol,
Myc	20 mM HEPES (pH 7.5), 100 mM NaCl
PBST	137 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 2.7 mM KCl, (pH 7.4), 0.1 % (v/v) Tween-20
Blocking	Milk powder (5%, 0.05 g/mL), 37 mM NaCl, 12 mM Phosphate, 2.7 mM KCl, pH 7.4, 0.1 % (v/v) Tween-20
Transfer	25 mM Bicine, 25 mM Bis-Tris, 1.025 EDTA free acid, 20% (v/v) methanol

2.3.3 Antibodies

Name	Supplier	Catalogue number	Host	Dilution ratio
Anti-N-Myc	Santa-Cruz Biotechnology	sc-53993	Mouse	1:1000
Anti-Aurora A	Sigma Aldrich	A1231 SIGMA	Mouse	1:4000
Anti-6xHis	Takada	631212	Mouse	1:5000
Anti-Mouse ECL	GE Healthcare	GENA931	Sheep	1:2000

2.3.4 SDS-PAGE

Protein samples were resolved using polyacrylamide gels, using 4-12 % gradient NuPAGE Bis-Tris Pre-Cast gels (Life Technologies, 1.0 mm, 15-well) run at 160 volts at constant voltage using 1x MES running buffer (Life Technologies) in an XCellSureLock Mini Cell (Life Technologies). SeeBlue

Plus2 Pre-stained Protein Standard (Invitrogen) was run alongside all protein samples as a molecular weight marker. Samples were diluted 1:1 with loading buffer (62.5 mM Tris-HCl, pH 6.8, 25 % glycerol, 2 % SDS, 2 % (v/v) β-mercaptoethanol, 0.01 % bromophenol blue). Prior to loading, samples for analysis were mixed with loading buffer in a 1 mL eppendorf tube, heated at 98 ^{o}C for three min, and centrifuged at 25 °C at 2,000 rpm for 1 min (Eppendorf microcentrifuge 5425; rotor fa-24x2). Gels were visualised by Coomassie Blue staining with InstantBlue Protein Stain (Expedon); 60 min stain and 120 min de-stain in Milli-Q water.

2.3.5 Western blotting

Resolved proteins (see section 2.3.4) were transferred to nitrocellulose blotting membrane (Amersham protran 0.2 µm nc) using a NuPAGE XCell II Blot Module (Life Technologies) according to the manufacturer's instructions, with a 20 % methanol transfer buffer (25 mM Bicine, 25 mM Bis-Tris (free base), 1 mM EDTA, 20 % (v/v) methanol, pH 7.2) using 30 V for 60 min. Following the transfer, the membrane was removed and rolled into a 50 mL falcon tube with 5 mL PBS with 0.1 % (v/v) Tween-20, twice for two minutes and then stained with Ponceau stain solution for 2 minutes (5 mL; Sigma; 0.1 % (w/v) in 5 % acetic acid). After imaging in an A4 plastic wallet the membranes were transferred to a fresh 50 mL falcon tube and incubated with blocking buffer (three times over, 5 mL of buffer was added and incubated for 10 minutes at room temperature). Following this, the blocking buffer was removed, and the membranes were incubated overnight at 4 ^{o}C with the specified primary antibody added at the recommended dilution with blocking buffer. Before the addition of the secondary antibody the membranes were rinsed with PBST (three times over, 5 mL of buffer was added and incubated for 10 min at room temperature). The membranes were then incubated with the specified horseradish peroxidase conjugated secondary antibody (2.5 µL in 5 mL of blocking buffer) for 60 min at room temperature. The membranes were then rinsed with PBST (three times over, 5 mL of buffer was added and incubated for 10 min at room temperature), transferred to an A4 plastic wallet and visualised with a ChemiDoc XRS+ Imaging System (Biorad) using ECL Western Blotting Substrate (Pierce) according to the manufacturer's instructions.

2.3.6 Protein constructs and expression

For MeroX searching the Protein FASTA sequences used are given in Appendix 5.2.1.

2.3.6.1 N-Myc

N-Myc proteins were supplied by Dr Selena Burgess and Dr Eoin Leen. FLAG-N-MYC 1-137 was expressed in *E. coli* BL21(DE3)RIL and purified as described previously with a TEV cleavable His-NusA fusions.[123] Tagless N-Myc 22-89 and Tagless N-Myc 1-137 were was expressed in E. coli BL21(DE3)RIL and purified as described previously and expressed as a TEV cleavable N-terminal His-tag fusion.[171] His tagged N-Myc 1-137 was produced using the same protocol as Tagless N-Myc 1-137 with the omission of the TEV cleavage removal of the N-terminal tag. Sequences of N-Myc constructs are shown below and are aligned to the human gene sequence (Uniprot). Blue highlighted residues indicate tags/vestigial residues, and yellow residues indicate additional residues.

```
sp|P04198|MYCN_HUMAN      ------------------------------------MPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHGRGPPTAG   144
3xFLAG_N-myc_1-137        ----------MDYKDHDGDYKDHDIDYKDDDDKAAMVPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG-------   163
His_N-myc_1-137           MHHHHHHSSGLVPRGSGMKETAAAKFEENLYFQGAMMPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG-------   173
Tagless_N-Myc_1-137       ---------------------------------GAMMPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG-------   140
Taqless N-Myc 22-89       ------------------------------------------------GAMDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWG----------------------------------------------------------   71
```

2.3.6.2 TPX2

TPX2 1-43 was expressed in E. coli BL21(DE3)RIL and purified as described previously Dr Selena Burgess with a TEV cleavable His-NusA fusions with the plasmid pETM6T1.[123]The sequence of the construct is shown below, aligned to a portion of the human sequence. Blue indicates vestigial residues.

```
TPX2_HUMAN-1-43           GAMAQVKSSYSYDAPSDFINFSSLDDEGDTQNIDSWFEEKANLEN---------------    45
sp|Q9ULW0|TPX2_HUMAN      --MSQVKSSYSYDAPSDFINFSSLDDEGDTQNIDSWFEEKANLENKLLGKNGTGGLFQGK    58
```

2.3.6.3 AurA

6xHis-N-terminally tagged human AurA 122–403 C290A C393A wild-type expressed from a pET30TEV vector in E. coli BL21(DE3)RIL as described previously.[274] The sequence of the AurA construct used is shown below, aligned against the human

sequence. The tag region is shown in pink, TEV protease site is shown in green, Cys point mutations are highlighted in yellow and vestigial residues from cloning are shown in blue.

```
>AurA_C290A_C393A-122-403  ------------------------------------------------------------   0
sp|O14965|AURKA_HUMAN      MDRSKENCISGPVKATAPVGGPKRVLVTQQFPCQNPLPVNSGQAQRVLCPSNSSQRIPLQ   60

>AurA_C290A_C393A-122-403  ------------------------------MHHHHHHSSGLVPRGSGMKETAAAKFEENL   30
sp|O14965|AURKA_HUMAN      AQKLVSSHKPVQNQKQKQLQATSVPHPVSRPLNNTQKSKQPLPSAP--E---NNPEEELA   115
                                                  :: ::*.  :* .   :       **

>AurA_C290A_C393A-122-403  YFNGAMESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKAQLEKAGVE   90
sp|O14965|AURKA_HUMAN      SKQKNEESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKAQLEKAGVE   175
                             :   ************************************************************

>AurA_C290A_C393A-122-403  HQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTAT   150
sp|O14965|AURKA_HUMAN      HQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTAT   235
                           ************************************************************

>AurA_C290A_C393A-122-403  YITELANALSYCHSKRVIHRDIKPENLLLGSAGELKIADFGWSVHAPSSRRTTLAGTLDY   210
sp|O14965|AURKA_HUMAN      YITELANALSYCHSKRVIHRDIKPENLLLGSAGELKIADFGWSVHAPSSRRTTLCGTLDY   295
                           ************************************************.*****

>AurA_C290A_C393A-122-403  LPPEMIEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPDFVTEG   270
sp|O14965|AURKA_HUMAN      LPPEMIEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPDFVTEG   355
                           ************************************************************

>AurA_C290A_C393A-122-403  ARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPSNAQNKESASKQS     318
sp|O14965|AURKA_HUMAN      ARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESASKQS     403
                           ************************************.**********
```

2.3.7 XL-MS

2.3.7.1 Testing of chemical crosslinking reactions

2.3.7.1.1 His N-Myc 1-137 with Aurora-A kinase 122-403 C290A C393A

2 nmol of His-N-Myc 1-137 (388 µM; 100 mM NaCl, 25 mM Tris pH 7.5 4 mM BME) was bound to washed Ni-NTA resin (25 µL suspended in 200 µL of wash buffer, 200 mM NaCl, 25 mM HEPES pH 8, 5 mM $MgCl_2$, 5 % (v/v) glycerol, 0.05 % (v/v) Tween-20; QIAGEN) then incubated with 3 nmol of Aurora-A kinase 122-403 C290A C393A (738 µM; 200 mM NaCl, 25 mM Tris pH 7.5, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol, 10 % glycerol), washed twice with buffer (200 mM NaCl, 25 mM HEPES pH 8, 5 mM $MgCl_2$, 5 % (v/v) glycerol, 0.05 % (v/v) Tween-20) and suspended in 200 µL XL buffer (200 mM NaCl, 25 mM HEPES pH 7.5, 5 mM $MgCl_2$, 5 % (v/v) glycerol) giving 10 µM of 1+1 complex. Four reactions were performed with the addition of DSBU from a 50 mM stock solution for 120 min at 4 ^oC (final concentrations of DSBU at 0.75 mM, 0.5 mM, 0.25 mM and 0 mM) and a further for 30 min at 4 ^oC. The series of reactions had final concentrations of DSBU at 0.25 mM, 0.175 mM, 0.025 mM and 0 mM. The reactions were quenched with addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration 20 mM. The samples were resolved using SDS-PAGE gels and western blotted using anti-Aurora A and anti-His antibodies.

2.3.7.1.2 3xFLAG-N-Myc 1-137 with Aurora-A kinase 122-403 C290A C393A

1 nmol of 3xFLAG-N-Myc 1-137 (90.176 µM ; 100 mM NaCl, 25 mM HEPES pH 7.5, 2 mM β-mercaptoethanol) was incubated with Aurora-A kinase 122-403 C290A:C393A (350 µM ; 200 mM NaCl, 25 mM Tris pH 7.5, 5 mM MgCl2, 2 mM BME, 10 % glycerol) and then diluted to a final protein concentration of 10 µM with buffer (200 mM NaCl, 25 mM HEPES pH 7.5, 5 mM MgCl2, 5 % (v/v) glycerol). The proteins were then crosslinked with the addition of DSBU (final concentration of 0.025 mM) at 4 ^oC for 30 min. The reactions were quenched with addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration 20 mM. The samples were resolved using SDS-PAGE gels.

2.3.7.2 Crosslinked peptide preparation for measurement

2.3.7.2.1 3xFLAG-N-Myc 1-137 with Aurora-A kinase 122-403 C290A C393A using in-solution digestion

FLAG-N-Myc 1-137 (1 nmol; 17 µg; 20 mM HEPES pH 7.2, 150 mM NaCl, 2 mM β-mercaptoethanol) was incubated with Aurora-A 122-403 C290A

C393A (1 nmol; 32.9 µg; 20 mM TRIS pH 7.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10 % glycerol) and was diluted to a final protein concentration of 10.0 µM with buffer (25 mM HEPES pH 7.2, 200 mM NaCl, 5 % glycerol, 2 mM $MgCl_2$) on ice. The sample was incubated with end over end rotation for 60 minutes in the cold room. The sample was divided into 0.4 nmol/10 ng aliquots. The crosslinking reaction was performed with the addition of DSBU (final concentration of 0.3 mM) at 4 ^oC for 30 min with shaking. The reactions were quenched with addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration of 20 mM, and allowed to come to room temperature.

A Smart digestion kit (Thermo) was used and fresh stocks of dithiothreitol (400 mM), iodoacetamide (800 mM) were prepared. The quenched reaction was diluted with Smart Digestion Buffer (proprietary buffer included with the kit; three volumes of the reaction volume). Trypsin smart digestion beads (5 µg) were added, and the mixture was shaken for 30 minutes at 70 ^oC. The beads were removed by centrifugation (3 minutes, 1400 g). Dithiothreitol was added to a final concentration of 4 mM, and the sample was incubated 30 minutes at 56 ^oC. Then iodoacetamide was added to a final concentration of 8 mM and the sample was incubated for 30 min in the dark at room temperature. The alkylation was quenched with the addition of dithiothreitol (480 nmol). A second proteolysis step was performed with the addition of GluC (0.5 µg), overnight at 32 ^oC. The protein digestion was stopped with 10 % (v/v) trifluoroacetic acid solution. The volume was reduced to 40 µL using a vacuum concentrator.

2.3.7.2.2 3xFLAG-N-Myc 1-137 with Aurora-A kinase 122-403 C290A C393A using in-gel digestion

As with (section 2.3.7.2.) however, the sample was not aliquoted. After the quenching of the cross-linking reaction, the samples were resolved using SDS-PAGE gels and visualised as previously stated. Each lane was loaded with at least 1.36 µg of total protein. The band corresponding to AurA+N-Myc complex was excised and cut into cubes. Taking precautions to reduce contamination, and freshly preparing solutions of dithiothreitol (10 mM), iodoacetamide (55 mM) and acetonitrile (neat), the excised band was incubated in 500 µL acetonitrile for 10 minutes at room temperature. The supernatant was removed, and the excised band was then incubated with dithiothreitol (50 µL) for 30 minutes at 56° C. The supernatant was removed and replaced with 500 µL acetonitrile for 10 minutes. After the acetonitrile was removed, the excised band was treated with iodoacetamide (50 µL) for

20 minutes in the dark. The supernatant was removed and replaced with 500 µL acetonitrile for 10 minutes. After removal of the acetonitrile, the gel pieces were rehydrated with ammonium bicarbonate (20 mM) for 60 min with shaking at room temperature and then, for enzymatic digestion, trypsin (0.5 µg) was added, and the sample was incubated with shaking overnight at 37 °C. A second proteolysis step was performed with the addition of GluC (0.5 µg) and after two hours the protein digestion was stopped with 10 % (v/v) trifluoroacetic acid solution. To extract the peptides from the gel slice digests, ammonium bicarbonate:acetonitrile (100 µL, 1:1) was added and the gel pieces were shaken for 15 minutes at 37 ºC. The supernatant, containing the peptides, was removed to a clean sample tube, and a second extraction was performed on the gel pieces. The pooled supernatant was reduced to a volume of 40 µL using a vacuum concentrator.

2.3.7.2.3 Aurora-A kinase 122-403 C290A C393A using in-gel digestion

Aurora-A 122-403 C290A C393A (0.35 nmol; 11.6 µg; 20 mM TRIS pH 7.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10 % glycerol) was diluted to a final protein concentration of 10.5 µM with buffer (25 mM HEPES pH 7.2, 150 mM NaCl, 5 % glycerol, 2 mM $MgCl_2$) on ice. The crosslinking reaction was performed with the addition of DSBU (final concentration of 1.0 mM) at 4 ºC for 30 minutes with shaking. The reactions were quenched with the addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration of 20 mM. The samples were resolved using SDS-PAGE gels and visualised as previously stated. The band corresponding to AurA monomer was excised and cut into cubes. Taking precautions to reduce contamination, and freshly preparing solutions of dithiothreitol (10 mM), iodoacetamide (55 mM) and acetonitrile (neat), the excised band was incubated in 500 µL acetonitrile for 10 min at room temperature. The supernatant was removed, and the excised band was then incubated with dithiothreitol (50 µL) for 30 minutes at 56° C. The supernatant was removed and replaced with 500 µL acetonitrile for 10 minutes. After the acetonitrile was removed, the excised band was treated with iodoacetamide (50 µL) for 20 min in the dark. The supernatant was removed and replaced with 500 µL acetonitrile for 10 minutes. After removal of the acetonitrile, the gel pieces were rehydrated with ammonium bicarbonate (20 mM) for 60 min with shaking at room temperature and then, for enzymatic digestion, trypsin (1:20 enzyme/protein ratio w/w) was added, and the sample was incubated with shaking overnight at 37 °C. A second proteolysis step was performed with the addition of chymotrypsin (1:20 enzyme/protein ratio w/w) and after two

hours the protein digestion was stopped with 10 % (v/v) trifluoroacetic acid solution. To extract the peptides from the gel slice digests, ammonium bicarbonate:acetonitrile (100 µL, 1:1) was added, and the gel pieces were shaken for 15 minutes at 37 ^{o}C. The supernatant, containing the peptides, was removed to a clean sample tube, and a second extraction was performed on the gel pieces. The pooled supernatant was reduced to a volume of 40 µL using a vacuum concentrator.

2.3.7.2.4 TPX2 1-43 with Aurora-A kinase 122-403 C290A C393A using in-gel digestion

TXP2 1-43 (0.36 nmol; 2.4 µg; 20 mM KPO_4 pH 7.0, 50 mM NaCl, 1 mM DTT) was incubated with Aurora-A 122-403 C290A C393A (0.35 nmol; 11.6 µg; 20 mM TRIS pH 7.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10% glycerol) and was diluted to a final protein concentration of 10.5 µM with buffer (25 mM HEPES pH 7.2, 150 mM NaCl, 5% glycerol, 2 mM $MgCl_2$) on ice. The sample was incubated with end over end rotation for 20 min in the cold room. The crosslinking reaction was performed with the addition of DSBU (final concentration of 1.0 mM) at 4 ^{o}C for 30 min with shaking. The reactions were quenched with addition of ammonium bicarbonate (0.5 uL, 250 nmol) to a final concentration of 20 mM. The samples were resolved using SDS-PAGE gels and visualised as previously stated. The band corresponding to AurA+TPX2 complex was excised and cut into cubes.

2.3.7.2.5 Tagless N-Myc 1-137 with Aurora-A kinase 122-403 C290A C393A using in-gel digestion

Tagless N-Myc 1-137 (0.3 nmol; 4.7 µg; 20 mM HEPES pH 7.2, 150 mM NaCl, 2 mM β-mercaptoethanol) was incubated with Aurora-A 122-403 C290A C393A (0.35 nmol; 11.6 µg; 20 mM TRIS pH 7.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10 % glycerol) and was diluted to a final protein concentration of 10.5 µM with buffer (25 mM HEPES pH 7.2, 150 mM NaCl, 5 % glycerol, 2 mM $MgCl_2$) on ice. The sample was incubated with end over end rotation for 20 min in the cold room. The crosslinking reaction was performed with the addition of DSBU (final concentration of 1.0 mM) at 4 ^{o}C for 30 min with shaking. The reactions were quenched with the addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration of 20 mM. The samples were resolved using SDS-PAGE gels and visualised as previously stated. The band corresponding to AurA+TPX2

complex was excised and cut into cubes. The preparation of the peptides was performed in section 2.3.7.2.4.

2.3.7.2.6 Tagless N-Myc 22-89 with Aurora-A kinase 122-403 C290A C393A using in-gel digestion

Tagless N-Myc 22-89 (0.34 nmol; 2.71 µg; 20 mM HEPES pH 7.2, 150 mM NaCl, 2 mM β-mercaptoethanol) was incubated with Aurora-A 122-403 C290A C393A (0.35 nmol; 11.6 µg; 20 mM TRIS pH 7.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10 % glycerol) and was diluted to a final protein concentration of 10.5 µM with buffer (25 mM HEPES pH 7.2, 150 mM NaCl , 5 % glycerol , 2 mM $MgCl_2$) on ice. The sample was incubated with end over end rotation for 20 min in the cold room. The crosslinking reaction was performed with the addition of DSBU (final concentration of 1.0 mM) at 4 °C for 30 min with shaking. The reactions were quenched with the addition of ammonium bicarbonate (0.5 µL, 250 nmol) to a final concentration of 20 mM. The samples were resolved using SDS-PAGE gels and visualised as previously stated. The band corresponding to AurA+TPX2 complex was excised and cut into cubes. The preparation of the peptides was performed as in section 2.3.7.2.4.

2.3.7.3 LC/MS/MS measurement of crosslinked peptides

Mass spectrometry performed in the Sinz lab, Halle, DE

For measurement of the peptide mixtures from the crosslinking reactions, approximately > 2 µg were loaded onto a pre-column for desalting (C8 reversed phase, Acclaim PepMap, 300 µm × 5 mm, 100 Å) and washed with 0.1 % TFA for 15 minutes before separation on a main column (C18 reversed-phase, Acclaim PepMap, 75 µm × 250 mm, 2 µm, 100 Å). Separation was performed using a 90 min gradient of 99 % solvent A and 1 % solvent B to 65 % A and 35 % B, followed by 5 minutes of 35 % solvent A and 85 % solvent B. Solvent A is 0.1 % formic acid in water and solvent B 0.08 % formic acid in acetonitrile. The nano-HPLC system was directly coupled to a Fusion Tribrid Orbitrap mass spectrometer equipped with a Nanospray Flex Ion Source.

Data acquisition was controlled with Xcalibur v3.0.63 (Thermo Fisher Scientific) and the measurement was acquired in data-dependent MS/MS mode with stepped higher-energy collision-induced dissociation (normalized collision energies 27 %, 30 % and 33 %). Each high-resolution full scan (m/z 300 to 1700 with a resolution of 120,000 at m/z 200) in the Orbitrap was followed by high-resolution product ion scans (resolution of 17,500) of the

ten most intense signals in the full-scan mass spectrum (isolation window 2 Th); the target value of the automated gain control was set to 3,000,000 (MS) and 200,000 (MS/MS), while maximum accumulation times were set to 100 ms (MS) and 250 ms (MS/MS) and the maximum cycle time was 10 s. Dynamic exclusion (exclusion duration 60 s) was enabled.

For measurement of all samples, except the samples including His-/Flag-tagged N-Myc, inclusion lists comprising all theoretical masses and charge states of potential cross-linked products of the proteins used in the reactions were specified, generated using MeroX, to allow for targeted mass analysis.

2.3.7.4 Identification of crosslinked peptides

Data was analysed using the MeroX (v2.0.1.4) software tool with the following search settings: no additional amino acids added, 5-30 peptide length, 5 missed cleavages, 800-8000 Da, semi-spec, 5 ppm precursor, 10 ppm fragment precision, no re-calculation, M to m (2) and C to B (2) as variable modifications, KSYT and N-term to KSTY and N-term crosslinks searched for, dead-end with “H2O,NH3”, maximum of 3 isotope shifts, minimum number of fragments as 3, RISEUP mode with 2 missing ions max. 10 % pre-score, 1 % FDR or score of 50 cut-off.

Each run was searched separately then combined for each condition.

2.3.7.5 Identification of peptides from crosslinked samples

To ensure that suitable coverage of the crosslinked peptides was achieved, the peptide search engine software, PEAKS 10.4, was used.

2.3.7.6 Model Evaluation using Crosslinked peptides

2.3.7.6.1 Generation of crosslink and mono-link lists

For each condition, the files containing data from the three runs were exported to Excel and separated by crosslink type. For each crosslink type, the crosslinks were then sorted by residue number, and then score and crosslink spectral matches (CSMs) were tallied, so that each crosslink found in the MeroX search of the triplicates had the CSMs associated with it. From this, crosslink lists were collated by removing crosslinks that were found to have only one CSM or only one CSM and less than the specified normalised factor, as described in the text (Section 2.4.2.2.1.2)

2.3.7.6.2 Annotation of structures

In PyMOL, the PyXlinkViewer plugin with the specified structural model was used to annotate the generated crosslink lists. A cut-off of 35 Å was used to denotate overlong crosslinks.

2.3.8 Model and structure generation

2.3.8.1 AurA

2.3.8.1.1 Homology model built from 1OL7 of AurA 119-403

In PyMOL, PDB structure 1OL7 was loaded, with the solvent and ligands removed. Aurora A AlphaFold2 structure (https://alphafold.ebi.ac.uk/entry/O14965) was loaded. Residues 119-129 from the AlphaFold2 structure were extracted and aligned to chain_A of 1OL7, and then 119-126 of chain_A of 1OL7 were replaced with the AlphaFold2 structure. The mutagenesis tool was used to correct the sequence of the replaced residues, so that the sequence modelled matched that of the construct crosslinked, including tags. The structure was then relaxed to a stable pose using the sculpt tool. The command *print(cmd.get_fastastr('all'))* was used to ensure the sequence modelled matched the sequence crosslinked.

2.3.8.1.2 Homology model built from 1OL7 with 7O2V of AurA 119-403

In PyMOL, PDB structure 7O2V was loaded, solvent and ligands removed, and aligned to residues 121-196 of the homology model built from 4CEG of AurA 119-403. Residues 119-126 of chain_A of 4CEG were replaced with those residues from 7O2V. Residues 390-403 of chain_A of 4J8M were replaced with those from AlphaFold2 structure.

2.3.8.1.3 Homology model built from 4J8M

In PyMOL, PDB structure 4J8M was loaded, solvent and ligands removed, and aligned to residues 121-196 of the homology model built from 4CEG of AurA 119-403. Residues 390-403 of chain_A of 4J8M were replaced with those from AlphaFold2 structure.

2.3.8.2 AurA + TPX2

2.3.8.2.1 Homology model, built from 1OL5

The starting structure, Aurora A 122-403 C290A C393A pT288 bound to TPX2 1-43, modelled from PDB structure 1OL5, was a gift from Mathew Bachelor.The structure was loaded into PyMOL with the Aurora A AlphaFold2 structure (https://alphafold.ebi.ac.uk/entry/O14965). Residues

119-129 from the AlphaFold2 structure were extracted and aligned to the Aurora A chain and then 119-126 of chain_A of 5G1X were replaced with those from the AlphaFold2 structure. The mutagenesis tool was used to correct the sequence of the replaced residues. The structure was then relaxed to a stable pose using the sculpt tool. The command *print(cmd.get_fastastr('all'))* was used to ensure the sequence modelled matched the sequence crosslinked.

2.3.8.2.2 Homology model, built from 4C3P

The starting structure, Aurora A 122-403 C290A C393A bound to TPX2 1-43, modelled from PDB structure 4C3P, was a gift from Mathew Bachelor. The structure was loaded into PyMOL with Aurora A AlphaFold2 structure (https://alphafold.ebi.ac.uk/entry/O14965). Residues 119-129 from the AlphaFold2 structure were extracted and aligned to the Aurora A chain and then 119-126 of the structure were replaced with the AlphaFold2 structure. The mutagenesis tool was used to correct the sequence of the replaced residues and phosphorylation. The structure was then relaxed to a stable pose using the sculpt tool. The command *print(cmd.get_fastastr('all'))* was used to ensure the sequence modelled matched the sequence crosslinked.

2.3.8.3 AurA + N-Myc

2.3.8.3.1 Homology model built from 5G1X

In PyMOL, PDB structure 5G1X was loaded, and solvent and ligands removed. The Aurora A AlphaFold2 structure (https://alphafold.ebi.ac.uk/entry/O14965) was loaded. Residues 119-129 from the AlphaFold2 structure were extracted and aligned to chain_A of 5g1x, and then 119-126 of chain_A of 5g1x were replaced with the AlphaFold2 structure. The mutagenesis tool was used to correct the sequence of the replaced residues. This process was repeated for 390-403 of chain_A. N-Myc AlphaFold2 structure (https://alphafold.ebi.ac.uk/entry/O14965) was loaded, and residues 1-62 from the AlphaFold2 structure were extracted and aligned to chain_B of 5g1x, and then 1-60 of chain_B of 5g1x were replaced with the AlphaFold2 structure. This process was repeated for 89-137 of chain_B. The structure was then relaxed to a stable pose using the sculpt tool. The command *print(cmd.get_fastastr('all'))* was used to ensure the sequence modelled matched the sequence crosslinked.

2.4 Results

2.4.1 Gel-based validation of the cross-linking method for kinases with IDRs using AurA+N-Myc 1-137

The Bayliss lab has previously characterized the interaction of the N-Myc TAD with the Aurora A KD using coprecipitation on affinity resin.[171] An on-resin coprecipitation protocol was adapted as the starting point for crosslinking reactions, as binding to resin was predicted to allow for a more controlled stoichiometry and prevent aggregation. The aim of the optimisation and validation we attempted here was to produce a DSBU chemical crosslinking reaction that balances the yield of crosslinking, maximising the generation of the 1+1 crosslinked complex, with minimising the generation of artificial, over-crosslinked/oligomeric complexes where the crosslinks have 'warped' the in-solution structure. This was predicted to maximise the crosslinks that would help elucidate the in-solution structure under investigation.

We chose to optimise the chemical crosslinking reaction conditions using AurA+N-Myc 1-137 by SDS-PAGE gel analysis before moving to mass spectrometry. We hypothesised that as N-Myc 1-137 lacks any significantly ordered structure and includes residues outside that of the minimum binding domain (28–89) of the AurA+N-Myc TAD, AurA+N-Myc 1-137 could generate over-crosslinked products, and thus functionally useless distance restraints. Over-crosslinked is commonly defined as protein crosslinks generated from the addition of excessive crosslinker, so that the proteins' structure is distorted. Once crosslinking conditions were established, we would apply the protocol to AurA±TPX2 to generate distance restraints for AurA Apo and AurA+TPX2 that could be annotated to crystallographic structures. Once the approach was validated, we would explore the AurA±N-Myc complex using AurA+N-Myc 1-137 (the full length TAD), and AurA+N-Myc 22-89 (the previously validated minimum binding domain of N-Myc).

2.4.1.1 Experiment design of the initial validation of XL-MS for kinases with IDRs and interacting with IDPs

2.4.1.1.1 Process to optimise and validate XL-MS of kinase with IDRs

In our initial evaluation of the literature regarding the XL-MS of disordered or dynamic proteins, we found two separate but related aspects of the protocol that needed evaluation before we could investigate the structural dynamics

of the AurA+N-Myc interaction. Firstly, we needed to optimise the reaction conditions to generate a homogeneous or near homogeneous mixture of crosslinked species, suitable for in-solution digestion and then measurement. To do this, we decided to use AurA+N-Myc 1-137, as the presence of N-Myc as an IDR could generate artificial or over-crosslinking. We used SDS-PAGE and Western blot analysis of the reaction mixtures to evaluate the suitability of the crosslinked species generated for MS measurement.

Secondly, we needed to validate that the crosslinks generated by this protocol produced distances that are reasonable and comparable to previous structures. As there is no crystal structure, or similar model, of the full-length AurA+N-Myc 1-137 interaction, we used AurA+TPX2 for this task. There are many crystal structures of this interaction, with strong orthogonal biochemical validation of the in-solution structure. Once we had determined that the crosslinked protocol had passed these steps, we would use it to investigate the AurA+N-Myc interaction.

2.4.1.2 Gel-based validation of AurA crosslinking reactions

At the time these experiments were designed (2018), we were unable to find example protocols that closely matched the planned experiment of crosslinking a ≈14 kDa disordered protein that transiently interacted with a ≈30 kDa globular protein. We proceeded to find conditions that produced a clean crosslinked 1+1 band at 38 kDa by SDS-PAGE (Figure 2-1)**,** based on protocols from Sinz *et al*.[233] The MS cleavable crosslinker DSBU was chosen for our study as it was reported in the XL-MS characterization of the intrinsically disordered tetrameric tumour suppressor protein p53, with data analysis supported by the free, fully automated crosslink identification software MeroX. To begin evaluating if XL-MS is a viable tool in understanding the structural basis of an interaction between a kinase and an IDP, literature protocol from Sinz *et al.*, was adapted and reaction conditions were examined using SDS-PAGE gels and Western blots.[197] In this protocol, the authors crosslinked p53, a 46 kDa IDP that forms a tetramer, using 5-fold to 50-fold molar excess of crosslinker, at 4 ^{o}C for 2 hours.[279]

Previously, Richards *et al* had coprecipitated Flag-tagged N-Myc 1-137 with Aurora A 122-403 C290A C393A, and visualised the interaction by Western blot, using antibodies directed against Aurora-A and the Flag-tag on N-Myc.[171] These conditions were adapted, using His-tagged N-Myc 1-137, with

the Sinz protocol, to perform a titration of crosslinker to assess what reaction conditions should be used to measure crosslinked peptides. Titration of 25-75-fold molar excess of crosslinker, and an incubation period of 2 hours, gave a band corresponding to the molecular weight of AurA+N-Myc (≈49 kDa) that blotted for the His-tag of His-tagged N-Myc 1-137 (Figure 2-1). However, the cross-linked sample was shown to contain near-continuous bands of oligomerization. Herzog *et al* found a 30 minute reaction reduced artefactual oligomerization.[280] The change in mobility of His-tagged N-Myc in the presence and absence of the crosslinker is also noticeable, suggesting that this protein is being modified by the crosslinker. DSBU will react with lysine residues: this will change the charge of the protein and thus potentially its mobility during SDS-PAGE separation. Figure 2-1 demonstrates that a titration of 2.5-25-fold molar excess reacted for 30 minutes gave a crosslinker-protein complex with no smearing, but less N-Myc was incorporated into the complex. On the Ponceau stained blot we see an AurA band in the no DSBU lane. In the Anti-His blot we see an over-crosslinked smear at > 62 kDa. With the reduced reaction time defined crosslinked species are appear. However the Anti-AurA blot has a band consistent with dimeric AurA. The reaction when performed with or without affinity resin showed no noticeable difference when analysed by SDS-PAGE gel (data not shown).

In an attempt to further reduce the formation of crosslinked oligomeric species, the total protein concentration was reduced from ≈20 μM to ≈10 μM (Figure 2-2). However, at 1.5-fold molar excess of DSBU, there was a reduction in the 1+1 complex being produced. At 30-fold molar excess (Figure 2-2), a satisfactory amount of 1+1 complex formation was produced for analysis of the sample by in-solution digest and MS.

We were concerned that the positive charge of the His purification tag used on N-Myc could cause artefactual interaction with Aurora A, which was Flag-tagged (and hence negatively charged), under the optimised affinity XL-MS conditions. However, the reaction when performed with or without affinity resin showed no noticeable difference when analysed by SDS-PAGE gel, so preincubation of the tagged N-Myc to the affinity resin step of the protocol was removed (data not shown).

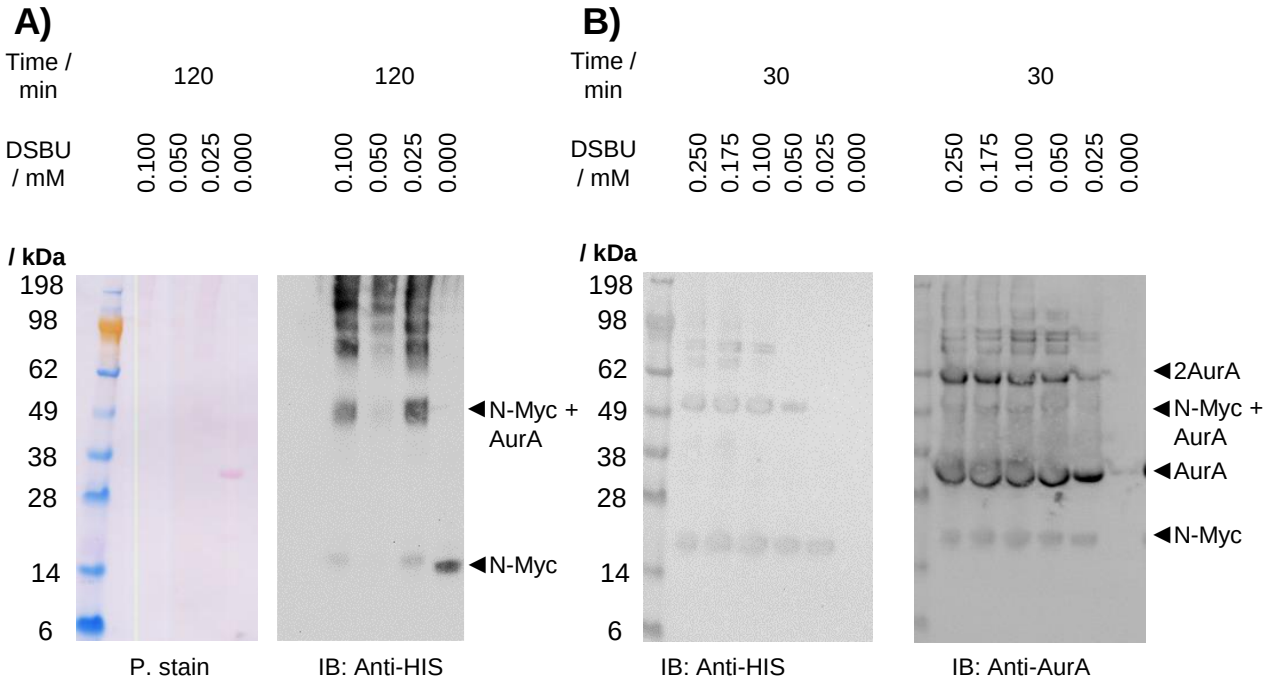


Figure 2-1 Characterization of AurA HIS-N-Myc

Western blot analysis of test DSBU cross-linking reactions between 10 µM AurA KD and 10 µM N-Myc TAD. Precipitated proteins were detected by Western blotting with the indicated antibodies or Ponceau stain where indicated. A) Titration of DSBU with fixed concentrations of His-tagged N-Myc 1-137 and Aurora-A kinase 122-403 C290A:C393A at 4°C. The His-tagged N-Myc was bound to a nickel-charged affinity resin then incubated with AurA. After a washing step, the crosslinker was added and, after the given reaction time, quenched by the addition of ammonium bicarbonate. B) Titration of DSBU concentration using the same reaction conditions but with a reduced reaction time. At x4-fold molar excess of DSBU (0.10 mM), a 1+1 dimer forms. Note the ineffective stripping of the HIS primary antibody from the membrane, resulting in the AurA image containing a signal from Anti-His tag AB.

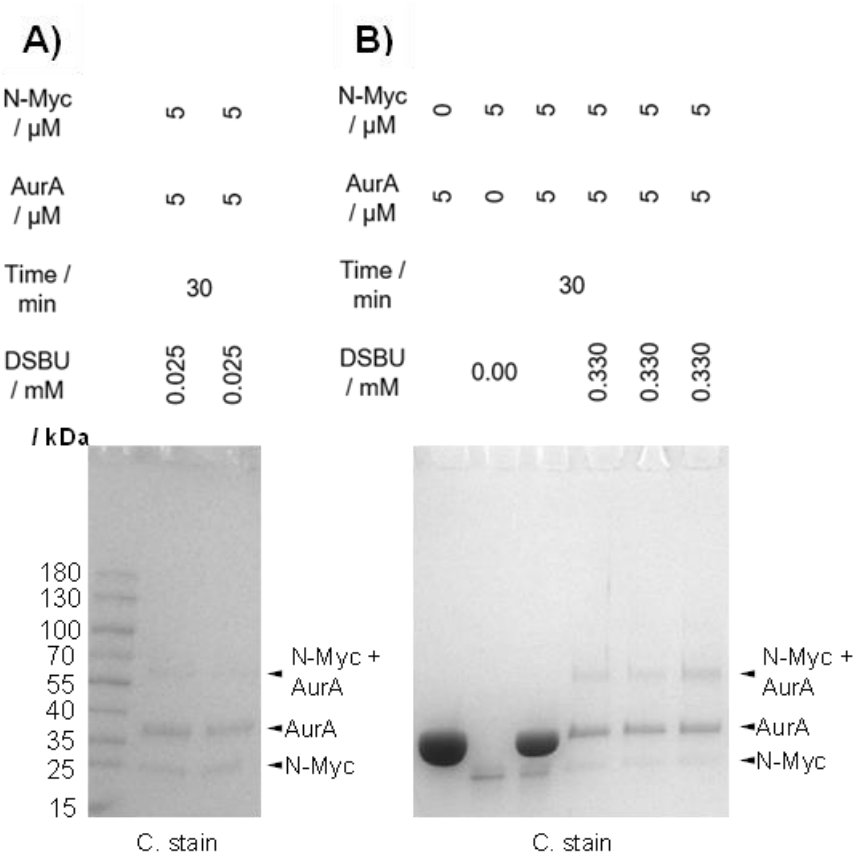


Figure 2-2 Characterisation of optimised Aurora-A/N-Myc cross-linking reactions

SDS-PAGE analysis of the optimised Aurora-A/N-Myc cross-linking reaction conditions visualised with Coomassie brilliant blue stain. A) Replicates of the reaction of DSBU, in 1.5-fold molar excess, to AurA KD incubated with Ni-NTA resin bound His-N-Myc 1-137. B) Replicates of reaction of DSBU, in 32-fold molar excess, to AurA KD incubated with FLAG-N-Myc 1-137. Lanes 1-3 analysed in the absence of DSBU and presence of DSBU.

2.4.1.3 Optimization of MS measurement of AurA/N-Myc

Having optimised the reaction conditions that gave a single 1+1 AurA/N-Myc band without significant higher mass complexes, by adjusting the crosslinker concentration, reaction time and temperature, attention turned to optimising the preparation of peptides for measurement in collaboration with the Sinz group. We first evaluated in-gel digestion compared to in-solution digestion to generate the samples for MS: in-gel has the drawback of increased processing time and potential reduction in peptide coverage, but has the benefit of being able to remove unreacted and over-reacted, oligomerised species.

2.4.1.3.1 In-solution digestion measurement of FLAG-N-Myc 1-137 and Aurora-A kinase 122-403

2.4.1.3.1.1 Crosslinking reaction of AurA+N-Myc for measurement

The crosslinking reactions of FLAG-N Myc 1-137 and AurA KD 119-403 C290A C393A were performed and digested using in-solution and in-gel protocols with a ≈30-fold molar excess of DSBU. Trypsin in combination with GluC were selected as the digestion proteases. This was due to Trypsin alone being expected to give poor coverage of N-Myc 1-137, since it lacks consistent distribution of arginine or lysine residues. To determine reaction size, we followed the rule of thumb of injecting into the instrument a volume of sample equivalent to ≈2 µg crosslinked protein, prior to digestion, per injection. As we wanted to measure each sample in triplicate a reaction size of 10 µg of total protein was selected.

For the reaction, a final total protein concentration of 10.5 µM was incubated with DSBU at a final concentration of 0.025 mM, for 30 minutes of end over end rotation in a cold room held at ≈4 ^{o}C. The reaction was quenched with the addition of ammonium bicarbonate, with a final concentration of 25 mM. After in-solution digestion and concentration, the mixture of peptides was separated in the UltiMate™ 3000 RSLCnano HPLC system using online desalting and separation, then measured on an Orbitrap Fusion Tribrid mass spectrometer fitted with a nano-ESI source. As the instrument used for measurement was in regular use for DSBU-crosslinked samples, there were validated separation and acquisition protocols ready for use.[281 282] The only parameter that needed adjustment was the length of time used for LC separation. We found that a separation time of 90-120 minutes produced an acceptable chromatogram during measurement; any longer and the sample was diluted, producing adverse outcomes.

2.4.1.3.1.2 Optimising MeroX search parameters

MeroX was used to search the mzml converted file formats data, which yielded 128 CSMs of 58 crosslinks and monolinks (Table 2-1).

Crosslink spectrum matches, CSMs (sometimes referred to as cross-link peptide-spectra matching), are where an MS^2 spectrum has been identified as being generated from the measurement of a pair of peptides crosslinked together. For the sake of this thesis, CSMs refers to both crosslink spectrum matches and monolink spectrum matches. Specific discussion of monolinks will use MSMs. Peptide spectrum matches, PSMs, refer to non-crosslinked peptides.

An estimation of the error rate of these searches is critical to ensure reliable data. MeroX, and PEAKS used for the searching of non-crosslink peptides,

use a target–decoy approach, where the sequences of the proteins to be searched are randomised and analysed under the same conditions. The rate of the matching of a spectrum to a decoy peptide is used to estimate the false discovery rate. MeroX also uses positive identification of fragment ions from DSBU, peptide intensity cut-offs, and scoring to assess the quality of CSMs. Although Beveridge *et al* demonstrated that the use of cleavable crosslinkers in proteomics does not inherently improve the FDR, this approach does improve assignment confidence generally.[235]

Here we used MeroX, in RISEUP mode, using generally accepted scoring methods of 10 % intensity prescore with 1 % FDR and CSM score of >50 cut-offs. This was in addition to slightly smaller mass tolerances than usual, and the requirement of positive identification of two of the four DSBU fingerprint ions.[283]

Balancing the inclusion of false or mis-identified peptide crosslinks with the exclusion of crosslinked peptides that degraded, due to processes like deamidation or oxidation, is crucial to producing reliable and repeatable structural data.[284] [285] If data controls are too stringent, we risk the omission of crosslinked peptides that are beneficial in structure determination.

Using the “off the shelf” MeroX search parameters, with wider mass tolerances and FDR, and stricter digestion parameters, gave a portion of the generated crosslinks. However, when including monolinks generated from the reaction of DSBU with NH_3 (from the ammonium bicarbonate quench, which gives an amide terminus to the free end of the crosslinker), Tris (a component on the N-Myc buffer) and hydrolysis (giving acid terminus), in concert with a semi-specific digestion, this gave a significant increase in CSMs (Table 2-1): from 77 to 126 total CSMs. The mass tolerances were made stricter as a precaution due to the wider search parameters, but the average MeroX score for CSMs increased to 126 from 110.

Another important control performed was the searching of non-crosslinked peptides (data not shown). As crosslinked peptides are a small proportion of the total peptides,[286] confirming that the peptide coverage is improved, and that the protein sequence and PTM status is as expected, is important.

Table 2-1 Breakdown of crosslinks and monolinks found for in-solution AurA KD crosslinked to Flag-N-Myc 1-137

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM. The standard search conditions are those suggested for MeroX crosslink peptide searching conditions given by Sinz *et al* 2018. Improved method is discussed in section 2.4.1.3.1.2 of the text.

Search conditions	Cut-offs	AurA				AurA-N Myc protein		NMyc			
		intra-protein		Mono-link				intra-protein		Mono-link	
		Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
Standard	n≥1	24	43	9	17	8	8	0	0	3	9
	n>1	9	28	4	12	0	0	0	0	2	8
Improved	n≥1	24	44	13	40	13	20	4	6	4	17
	n>1	10	30	10	37	3	8	2	4	3	16

2.4.1.3.1.3 Alignment of crosslinks and monolinks is necessary for ease of quantification

To ensure adequate coverage of the crosslinked AurA+N-Myc sample, a two-stage Trypsin-GluC digestion protocol was employed. This necessitated using a semi-specific digestion search protocol with MeroX. For this, one cleavage site of the in-silco generated peptide to be searched against needs to be from the specified proteases, and other cleavage can occur non-specifically in accordance with the specified peptide length parameter. This is due to the desired incomplete GluC digestion generating a complex mixture of peptides. We found for ease and simplicity of the crosslink and mono-link quantification, crosslinks and monolinks needed to be aligned to their respective residues for analysis.

A single, complete protease digestion protocol will generate a homogeneous mixture of peptides where crosslinks are found on a single pair of peptide species and monolinks on a single peptide species for any given stretch of sequence. During LC separation prior to injection into the mass spectrometer, this peptide should elute as a single peak. Hence the total number of CSMs for a given crosslink is directly proportional to the total ion count of a specific single peptide species. Figure 2-3 shows different peptide combinations found for a single crosslink.

Figure 2-3 Example of the need to align crosslinks of the same residues but differing peptides

The measurement of AurA KD + N-Myc 1-137 returned six CSMs for the crosslink Lys143-Lys171, however these were found on three different peptide species which needed to be combined for analysis. The AurA residue numbers are given by 'Peptide 1 and 2'; retention times of the crosslinked peptide during LC separation are given by 'Retention time'.

Crosslink	Peptide 1	Peptide Sequence 1	Peptide 2	Peptide Sequence 2	Retention Time / s
Lys143-Lys171	142-152	G**K**FGNVYLAR	171-180	**K**AGVEHQLR	4130
Lys143-Lys171	142-152	G**K**FGNVYLAR	171-180	**K**AGVEHQLR	4135
Lys143-Lys171	142-152	G**K**FGNVYLAR	171-180	**K**AGVEHQLR	4131
Lys143-Lys171	142-152	G**K**FGNVYLAR	171-180	**K**AGVEHQLR	4130
Lys143-Lys171	138-144	PLGKG**K**	171-180	**K**AGVEHQLR	3018
Lys143-Lys171	135-144	IGRPLGKG**K**	171-181	**K**AGVEHQLRR	2965

We found that the two-protease stepped digestion protocol generated crosslinks and mono-links on diverse peptide species which could elute at differing points of separation.[211] This is an example of CSM alignment, showing that the total number of CSMs for a specific crosslink, or MSMs for a specific monolink, is often the sum of the spectrum matches, for a handful of species. Thus, for the quantification we will use MSMs and CSMs to calculate the relative reactivities and so calibrate the importance of crosslinks to our structural understanding.

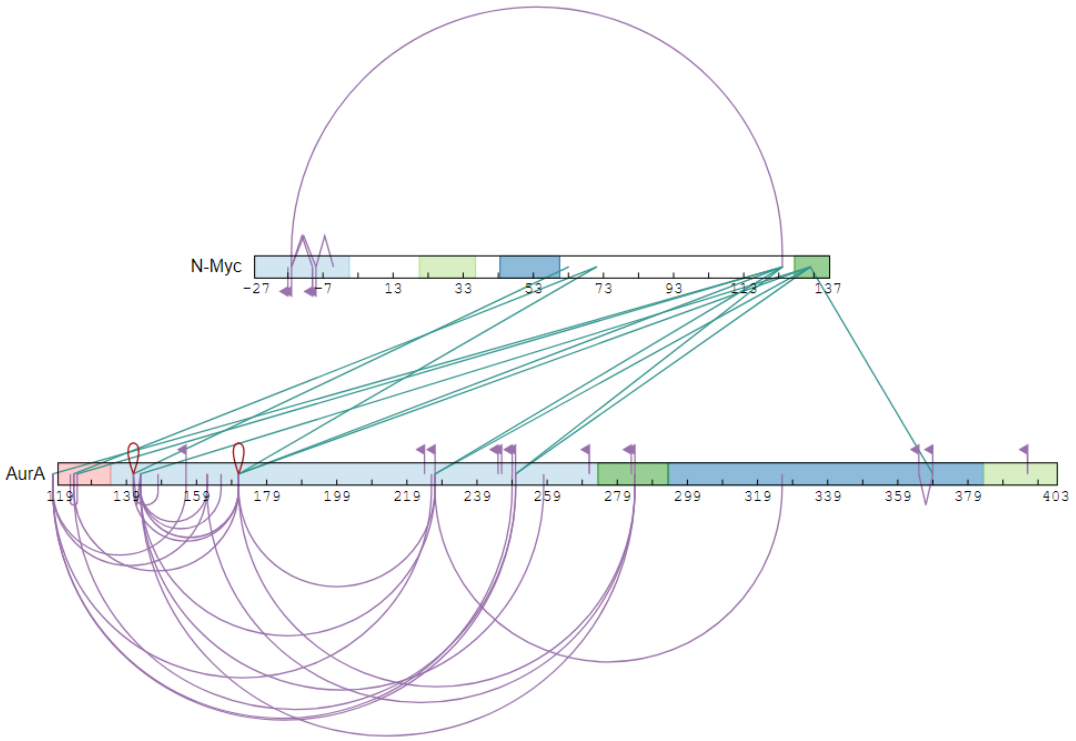
2.4.1.3.1.4 Results of the aligned crosslinks

Prior to annotating the aligned crosslinks on crystal structures or an appropriate model, the web server XiNet was used to generate crosslink plots. Crosslinks identified are shown in Table 2-3.

In the test dataset, we saw a plurality of crosslinks with a single CSM and an expected exponential distribution of CSMs. Therefore, as part of data processing to generate crosslink and monolink lists to be annotated on crystal structures or appropriate models to generate crosslink distances, we chose to exclude crosslinks and monolinks with a single CSM, as an additional safeguard against the inclusion of false positive crosslink identification, as well as crosslinks generated from transient or non-equilibrium conditions. However, this highlighted the need to produce final datasets which contained measurement replicates in the future, since we found that CSM based cut-offs (i.e. ignoring all crosslinks with a single CSM) risked eliminating our ability to observe an entire region's reactivity. For example, N-Myc Lys52 generated two or more crosslinks but all with only 1-2 CSMs.

Despite observing intra-protein crosslinks from MB0 and the C-terminus of N-Myc to AurA, homotypic crosslinks showing the presence of dimeric species were found. These included the homotypic crosslink AurA Lys171-Lys171, which was found with four CSMs. We were therefore concerned that the in-solution digestion was including a significant number of oligomeric species, hence further work using in-solution digestion was abandoned. The significant presence of homotypic crosslinks indicates that crosslinks from dimeric species are being measured and so we cannot determine which AurA crosslinks are intra- or interprotein. Figure 2-4 shows an XiNet diagram of all found crosslinks from the in-solution digested sample.

A) All identified crosslinks



B) Crosslinks identified with more than a single CSM

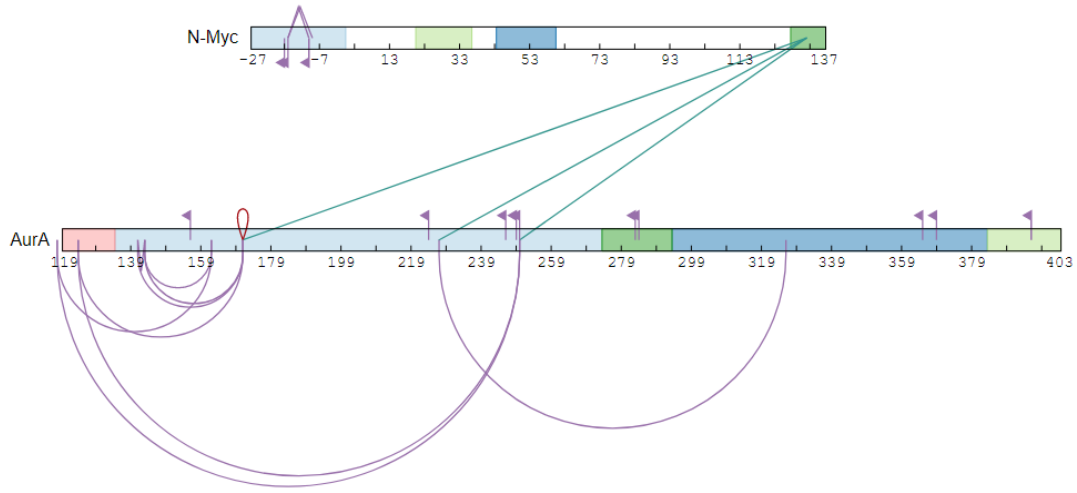


Figure 2-4 All crosslinks identified from in-solution digest of AurA KD crosslinked with DSBU to Flag-tagged N-Myc 1-137

Overview of the identified crosslinks between AurA KD 122-403 and FLAG N-Myc 1-137. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For N-Myc (L to R), domains are the N-terminal affinity tag, Myc Box 0 (20-36), Myc Box 1 (44-63), Myc Box 2 (128-143). For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label; straight-sided link represents an intra-peptide crosslink; curved link is an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.1.3.1.5 Composition of monolink products

Curiously, we found that ≈35 % of the monolinks reported were amidated, the product of the DSBU reacting with ammonium from the quenching ammonium bicarbonate shown in Table 2-2. The remaining monolinks were generated from the reaction with the buffer component Tris or hydrolysis, corresponding to ≈15 % of the total CSMs found. This trend was found in all other datasets where Tris or ammonium bicarbonate was used to quench the crosslinking reaction.

This suggests that an appreciable portion of the DSBU that reacts with the protein during the experiment remains unreacted until the end of the reaction.

Table 2-2 Composition of monolink products from in-solution digested AurA+FLAG N-Myc 1-137

	Protein	
Monolink Product	**AurA KD**	**Flag N-Myc**
Amide	15	5
Acid	17	9
tris	8	3

Table 2-3 Crosslinks identified from AurA+Flag N-Myc 1-137

Overview of the identified crosslinks from the DSBU reaction AurA+Flag N-Myc 1-137, in-solution digested. Where 'Type' is the type of crosslink (table 1-5), P1 is Protein 1, R1 is the residue number of the protein 1 that bears the crosslink, P2 is Protein 2. 'Top score' is the highest MeroX score of the CSMs found for that crosslink, where the minimum score cut-off is >50 and Total CSM. Residues of AurA (green) and NMyc (orange) are coloured using gradient light to dark. Negative numbers of the resiudes indicate crosslinks from an N-terminal tag.

Type	P1	R1	P2	R2	Top Score	Total CSMs
Intrapeptidal	AurA	141	AurA	143	157	3
Intrapeptidal	AurA	365	AurA	369	88	1
Intraprotein	AurA	118	AurA	156	100	1
Intraprotein	AurA	118	AurA	162	74	3
Intraprotein	AurA	118	AurA	226	69	1
Intraprotein	AurA	118	AurA	249	140	1
Intraprotein	AurA	118	AurA	250	80	2
Intraprotein	AurA	123	AurA	125	81	1
Intraprotein	AurA	124	AurA	171	110	3
Intraprotein	AurA	124	AurA	250	138	3
Homotypic	AurA	141	AurA	141	201	1
Intraprotein	AurA	141	AurA	148	66	1
Intraprotein	AurA	141	AurA	171	84	2
Intraprotein	AurA	143	AurA	162	86	2
Intraprotein	AurA	143	AurA	166	55	1
Intraprotein	AurA	143	AurA	171	185	6
Intraprotein	AurA	143	AurA	227	63	1
Intraprotein	AurA	143	AurA	258	87	1
Intraprotein	AurA	143	AurA	284	77	1
Intraprotein	AurA	162	AurA	284	73	1
Homotypic	AurA	171	AurA	171	234	4
Intraprotein	AurA	171	AurA	227	175	1
Intraprotein	AurA	171	AurA	284	179	1
Intraprotein	AurA	227	AurA	326	89	2
Intrapeptidal	NMyc	-15	NMyc	-9	128	2
Intrapeptidal	NMyc	-15	NMyc	-8	173	2
Intraprotein	NMyc	-15	NMyc	125	78	1
Intrapeptidal	NMyc	-8	NMyc	-3	92	1
Intraprotein	AurA	118	NMyc	125	74	1
Intraprotein	AurA	124	NMyc	72	51	1
Intraprotein	AurA	125	NMyc	125	109	1
Intraprotein	AurA	141	NMyc	64	88	1
Intraprotein	AurA	143	NMyc	133	81	1
Intraprotein	AurA	171	NMyc	72	51	1
Intraprotein	AurA	171	NMyc	125	74	1
Intraprotein	AurA	171	NMyc	133	180	4
Intraprotein	AurA	227	NMyc	125	57	1
Intraprotein	AurA	227	NMyc	133	152	4
Intraprotein	AurA	250	NMyc	125	70	1
Intraprotein	AurA	250	NMyc	133	176	2
Intraprotein	AurA	369	NMyc	133	67	1

Type	P1	R1	Top Score	Total CSMs
Monolink	AurA	156	163	3
Monolink	AurA	224	168	3
Monolink	AurA	246	146	2
Monolink	AurA	249	218	4
Monolink	AurA	250	208	2
Monolink	AurA	283	148	2
Monolink	AurA	284	200	3
Monolink	AurA	365	176	10
Monolink	AurA	369	131	3
Monolink	AurA	396	165	5
Monolink	NMyc	-16	165	7
Monolink	NMyc	-15	163	3
Monolink	NMyc	-9	185	6
Monolink	NMyc	-8	179	1

2.4.1.3.2 In-gel digestion measurement of FLAG-N-Myc 1-137 and Aurora-A kinase 122-403

2.4.1.3.2.1 Crosslinking reaction of AurA+N-Myc for measurement

The crosslinking reaction of FLAG-N-Myc 1-137 and AurA KD 119-403 was repeated, with the same reaction conditions as above, and the resulting mixture was resolved using SDS-PAGE gel. The 1:1 complex band was excised from the gel and measured in duplicate.

2.4.1.3.2.2 Crosslinks found

The combined measurements were searched, which returned 139 CSMs of 54 crosslinks and monolinks that were within the tolerance of MeroX search parameters. Search and cut-off parameters were as previous, and the output is shown in Table 2-4. The homotypic crosslink AurA Lys171-Lys171 was not observed, consistent with oligomers being removed before digestion. A table of crosslinks is shown in Appendix 5.1.1.1.

Table 2-4 Breakdown of crosslinks and monolinks found for in-gel AurA KD crosslinked to Flag-N-Myc 1-137

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	AurA				AurA-N Myc protein		NMyc			
	intra-protein		Mono-link				intra-protein		Mono-link	
	Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
n≥1	19	34	13	60	19	43	1	1	1	1
n>1	7	22	10	58	9	33	0	0	0	0

2.4.1.3.2.3 Results of the aligned crosslinks

Figure 2-5 shows the XiNet diagram of all crosslinks with more than a single CSM found from the duplicate in-gel digested Flag N-Myc 1-137 to AurA 119-403. As expected with the digestion of a single species, there is a relative reduction in crosslinks and monolinks found and the CSMs identified. In-gel digestion will affect coverage to a certain extent,[216] but the loss of N-Myc interprotein crosslinks and monolinks suggests that these result from species other than AurA+N-Myc. In addition, the increase in AurA-AurA crosslinks, especially involving the NTE and C-lobe, and lack of AurA homotypic crosslinks, again suggests the in-solution digestion includes AurA monomer and dimer species that are altering the pattern of crosslinks. Analysis of non-crosslinked peptides using the the PEAKS search engine

showed coverage comparable with the in-solution measurement (data not shown). The promiscuity of the MBII residue Lys125 is concerning, because it may be masking/impairing reactions at other crosslinkable sites, as is the lack of reactivity of the N-terminus of N-Myc, as this region should be binding to AurA. Furthermore, we see a noticeable reduction in monolinks in AurA and lack of any monolinks from N-Myc. The lack of homotypic crosslinks shows that the excised band measured was not from a dimeric species. The lack of AurA-MB0 crosslinks, seen in the in-solution measurement, may be accounted for by the in-gel digestion removing oligomeric AurA+N-Myc species. An improved in-gel digestion protocol may improve the coverage of data, assuming that triplicate measurement of the sample does not bring such improvements.

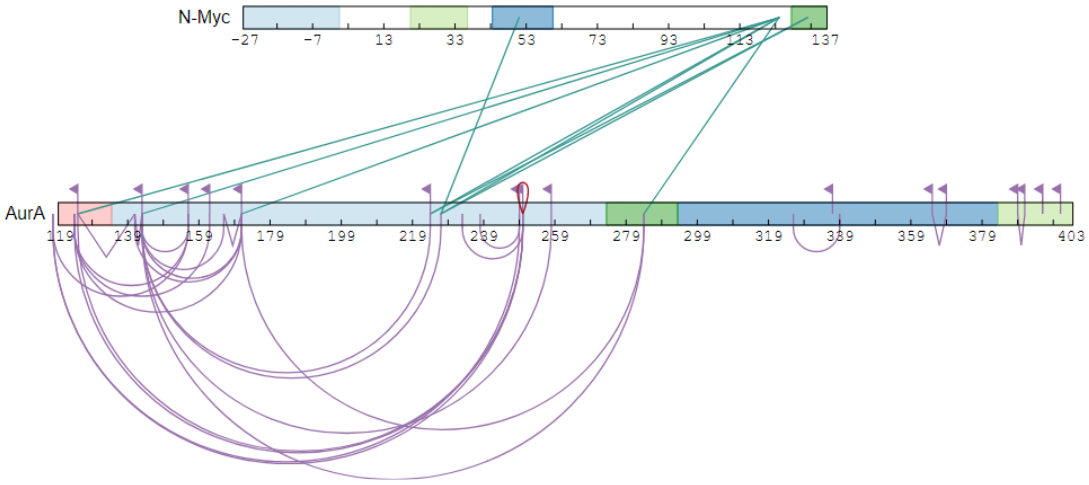


Figure 2-5 Combined breakdown of crosslinks and monolinks found from AurA KD crosslinked to Flag-N-Myc 1-137 from in-gel digestion with more than a single CSM

Overview of the combined identified crosslinks between AurA KD 122-403 and FLAG N-Myc 1-137 and AurA KD 122-403 from their respective SDS-PAGE bands. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For N-Myc (L to R), domains are the N-terminal affinity tag, Myc Box 0 (N-Myc 20-36), Myc Box 1 (N-Myc 44-63), Myc Box 2 (N-Myc 128-143). For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.1.3.3 In-gel digestion measurement of Aurora-A kinase 122-403 from FLAG-N-Myc 1-137 and Aurora-A kinase 122-403 reaction

The band corresponding to the weight of AurA in the above reaction was also excised and measured to examine the differences in crosslinking and monolinking found with the in-solution measurement (Table 2-5). Again, the

duplicate measurements were combined and searched, here returning 237 CSMs of 68 crosslinks and monolinks, Examination of the monolinks found that ≈39 % were amidated.

2.4.1.3.3.1 Crosslinks found

It is striking that we found a ×1.7-fold increase in CSMs and a ×1.25-fold increase in crosslinks identified in-solution compared to the in-gel AurA+N-Myc measurement. The AurA+N-Myc and AurA bands had a similar percentage reduction in crosslinks compared to the in-solution data.

Table 2-5 Breakdown of crosslinks and monolinks found for AurA monomer, found in AurA KD crosslinked to Flag-N-Myc 1-137 sample by in-gel digestion

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	AurA				AurA-N Myc protein		NMyc			
	intra-protein		Mono-link				intra-protein		Mono-link	
	Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
n≥1	45	126	23	111	0	0	0	0	0	0
n>1	25	106	14	102	0	0	0	0	0	0

2.4.1.3.3.1 Results of the aligned crosslinks of the AurA band

After searching and aligning the crosslinks of the duplicate in-gel digested AurA 119-403 band, the crosslinks were visualised in an XiNet diagram, (Figure 2-6). This highlights the pronounced preference for N-lobe reactivity over the C-lobe. Whilst the reactivity of AurA in this sample in the absence of Flag-N-Myc 1-137 was similar to the previous sample, it differed in several key aspects. We see that the activation loop is reactive and gives Lys143-Ser284 and Lys171-Ser284. This is critical to understanding the mobility of the activation loop, as inactive AurA is understood to have a dynamic activation loop. Additionally, the NTE of AurA increases its reactivity to the N-lobe. There is still consistent monolink coverage.

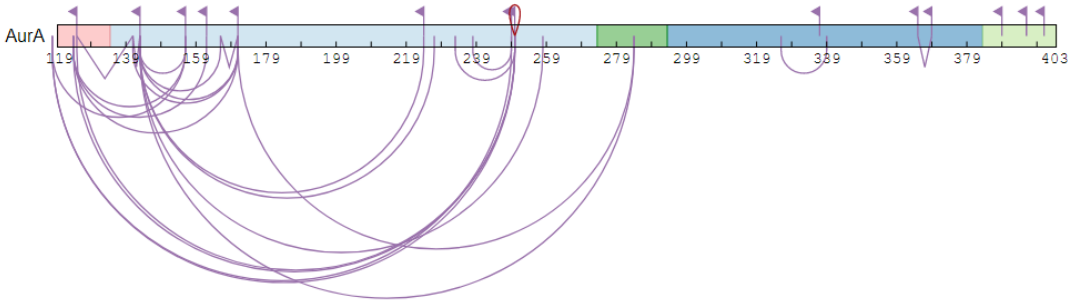


Figure 2-6 Breakdown of crosslinks and monolinks found for AurA band from AurA KD crosslinked to Flag-N-Myc 1-137 in-gel digestion with more than a single CSM

Overview of the identified crosslinks of AurA KD 122-403. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For AurA (L to R), domains are NTE (including cloning artefact residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.1.3.3.2 Assessment of and changes to the crosslinking reaction protocol

To summarise the data presented in the section above: the changes to the initial crosslinking reaction protocol conditions, and the reduction in crosslinker concentration and time, produced a protocol that generated a sufficient band of AurA+N-Myc that could be processed, and crosslinked peptides of the interaction measured. Initially, an in-solution digestion method was chosen due to the implicit benefits of this methodology. After minor adjustments to the MeroX crosslink peptide searching parameters, this generated crosslinks from multiple species with an abundance of homotypic crosslinks. In-gel digestion and measurement of the AurA+N-Myc and AurA bands of the SDS-PAGE separated reaction mixture do not include all the crosslinks found from the in-solution digestion. This, along with the lack of homotypic crosslinks, suggests that the in-gel protocol specifically excludes crosslinks for an oligomeric AurA species. Nevertheless, the in-gel protocol brings a reduction in N-Myc crosslink coverage. To be cautious, we decided to continue with the in-gel digestion protocol, to remove misleading AurA crosslinks originating from the AurA dimer. The reduction in N-Myc coverage we see when comparing the in-solution and in-gel AurA+N-Myc crosslinks we attempted to counteract through changes to our crosslinking and sample processing protocol. An increase of the crosslinker concentration ought to improve the yield of the desired bands, in addition to improving the digestion

protocol to increase the coverage of N-Myc crosslinks from the AurA+N-Myc band. We then decided to do this with AurA+TPX2.

2.4.2 AurA KD±TPX2 1-43 as a testbed for XL-MS of kinases with IDRs

2.4.2.1 In preparation for XL-MS of AurA+TPX2

We have shown that XL-MS of AurA and its complexes can generate rationalizable crosslinks and monolinks ready for analysis using generated model structures of the complex. As a full-length structure has not been generated for N-Myc- AurA, which is a high-value anti-cancer target, it is not suitable to validate whether the crosslinks are within tolerance for DSBU. Given the challenges of studying the N-Myc-AurA interaction, we decided to turn next to a more defined and well-studied complex. TPX2, like the N-Myc TAD, is another disordered interactor of AurA. However, crystal structures have been produced of the entire minimal binding domain of TPX2 for AurA. Thus, we employed AurA+TPX2 to validate the crosslinks that this protocol generates, by assessing whether the residue-to-residue crosslink distances were reasonable and comparable to previous structures, notwithstanding minor changes to the protocol. We hoped it would provide additional validation of the XL-MS protocol, and potentially a second structure to contrast with AurA Apo and AurA+N-Myc AurA Interacting Region (AIR). TPX2 has a higher affinity to AurA than N- Myc and has also been shown to stabilise AurA dynamics by pinning the activation loop.[86] This higher affinity is likely to generate more homogenous species prior to crosslinking.

2.4.2.1.1 The approach to XL-MS of restraints on AurA+TPX2

Using the protocol developed for the crosslinking of AurA+FLAG N-Myc 1-137, which was shown to be a viable strategy for generating XL-MS data between a kinase with disordered NTE and an IDP, we turned to the crosslinking of AurA KD±TPX2 1-43. This complex has several crystal structures of the minimum binding domain of TPX2 in complex with AurA with, orthogonal, in-solution biochemical data.[144] Our aim was to test whether our crosslinking protocol could adequately generate successful distance restraints. If this was the case, we should observe a majority of crosslinks within tolerance of DSBU, with no crosslinks involving fully solvent inaccessible residues, and changes to monolink reactivity that reflect the binding site and conformational changes that result from TPX2 interaction with AurA. With this in hand we could then return to generating structures of AurA+N-Myc that could be annotated with the distance restraints found.

2.4.2.1.2 Outcome of the crosslinking reaction of AurA+TPX2

The crosslinking reaction of AurA Apo and AurA+TPX2 performed as expected with the Apo sample generating monomeric, dimeric and oligomeric bands (Figure 2-7). Unlike the AurA+N-Myc crosslinking reactions, for the AurA+TPX2 reaction we only see an AurA monomer, AurA+TPX2 and a faint dimeric AurA+TPX2 (2:2 complex) band, and no higher-order AurA species. This unexpected result – of crosslinking with TPX2 producing far fewer crosslinked species, relative to the crosslinking reaction with N-Myc – will be explored further in the discussion.

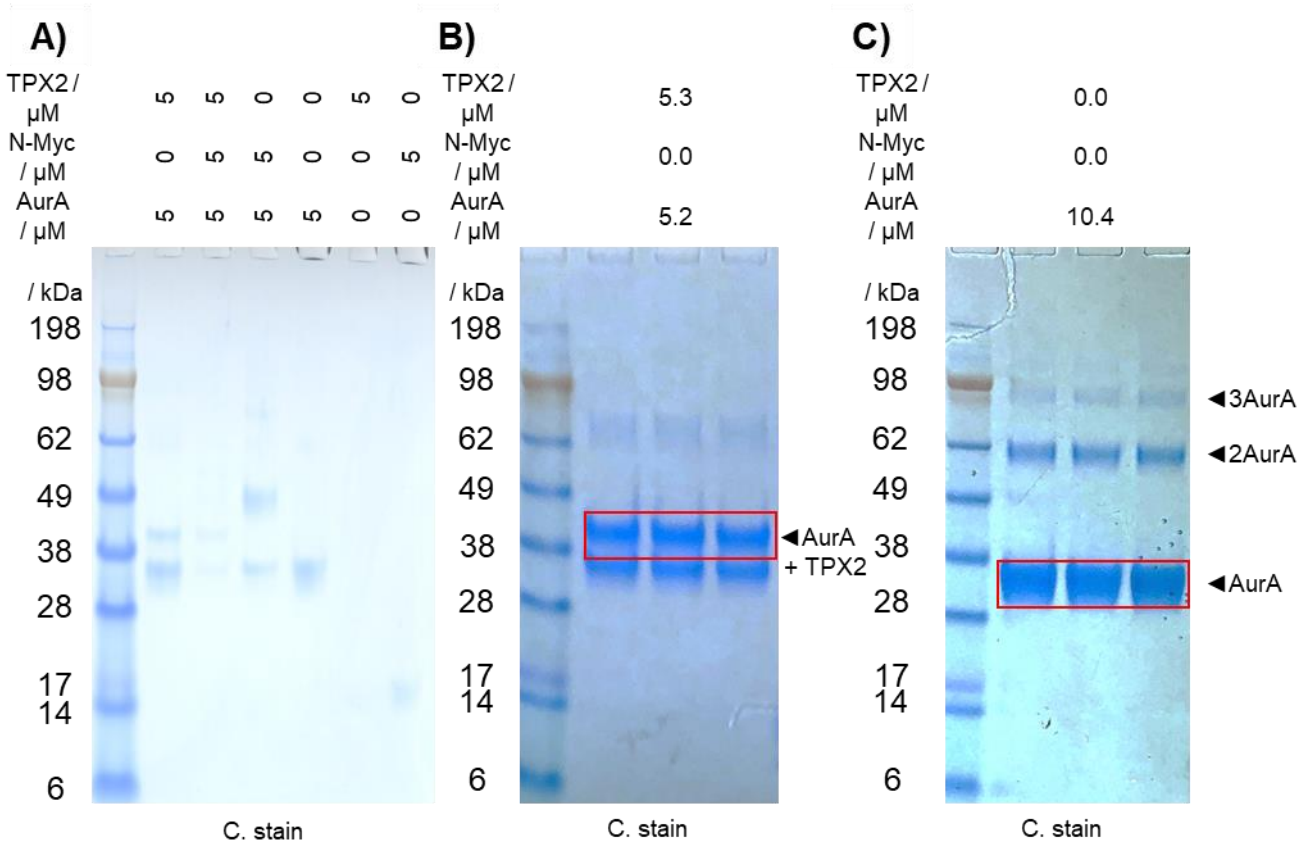


Figure 2-7 In-gel AurA Apo and AurA+TPX2 reactions

SDS-PAGE analysis of an example test crosslinking reaction of AurA+Tagless N-Myc 1-137 and AurA+TPX2 1-43, and reactions used for measurement of AurA Apo and AurA+TPX2. Visualised with Coomassie brilliant blue stain. All lanes contained 1.0 mM DSBU and wer crosslinked for 30 minutes. A) Test crosslinking reaction of AurA, TPX2 1-43 and tagless N-Myc 1-137 at ≈90-fold molar excess. TPX2 1-43 is known to not stain well C) Replicates of reaction of DSBU, in ≈90-fold molar excess, to AurA KD. B) Crosslinked AurA KD with TPX2 1-43. Indicated bands (red box) excised and processed for MS measurement.

2.4.2.2 XL-MS of AurA KD and analysis of the distance restraints generated

As crosslinking measurements of Aurora A 122-403 C290A C393A with N-Myc 1-137 successfully showed that the disordered regions did not dominate the distance restraints found to the extent of making them unusable, we used the optimised reaction conditions to crosslink AurA KD±TPX2 1-43. The reaction of DSBU with AurA KD alone and AurA KD+TPX2 1-43 was performed using the same conditions as in the previous section and analysed in triplicate on SDS-PAGE. However, the fold molar excess of the crosslinker was increased from ≈30 to ≈90 as the digest was performed in-gel, meaning that oligomerized species were not such an issue, to increase the yield of 1+1 complex. A sequential digestion of the crosslinked proteins, with trypsin and chymotrypsin, was used in place of the previous two-step Trypsin-GluC digestion, in order to improve coverage of N-Myc and the activation loop of AurA. An inclusion mass list was also used to increase the proportion of functional crosslinks and monolinks.[291] In-silico digestion of the protein sequences was performed using MeroX, which decorated peptides with potential monolinks and crosslinks. This was used in the MS/MS instrument setup so that after a precursor scan, ions matching the desired charge and m/z were selected for fragmentation, before then selecting ions by intensity.[225]

The resulting crosslinking reaction, alongside improved sample processing and measurement, revealed that higher mass AurA species generated by crosslinking were different in the presence and absence of TPX2. The crosslinked active AurA Apo, 32.83 kDa, showed an extended smeary monomer, with bands corresponding to a dimeric and trimeric species. However, crosslinked AurA KD±TPX2 1-43 presented as monomeric AurA and a band corresponding to the 1+1 complex and a faint band of dimeric 2:2 complex.

2.4.2.2.1 Analysis of crosslinked peptides of AurA APO by MeroX

The crosslinking reaction of AurA KD using ≈95-fold molar excess of DSBU (1.0 mM) generated an extended band at ≈35 kDa, figure 2-7. This was excised and once processed and measured, the results from each measurement were combined and searched using MeroX. The combined MeroX searches found 615 CSMs for 48 crosslinks and monolinks. A single homotypic crosslink was removed. A sequential digestion procedure with Trypsin-Chymotrypsin was used.[192] The sequential digestion used the non-

or low-specific protease, Chymotrypsin, incubated with the target proteins for a reduced length of time, following a digestion with the more specific protease, trypsin, to reduce the number of over-long peptides unsuitable for measurement. The coverage of non-crosslinked peaks was found to be > 90 %. Six residues have crosslinks but no monolinks (118, 124, 125, 156, 401 and 403). It is noticeable that all the residues that lack monolinks come from peptides that are neighbouring or are within a few residues of another crosslinkable residue. Table 2-6 shows crosslink type count during refinement. Note that crosslinks with more than 1 CSM were used .

Table 2-6 Breakdown of crosslinks and monolinks found for in-gel digested AurA KD Apo

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment, and after the removal of crosslinks with a single CSM.

	AurA			
Cut-offs	Intra-protein		Mono-link	
	Number	CSM	Number	CSM
n≥1	25	104	24	509
n>1	18	97	21	506

2.4.2.2.1.1 Results of the aligned crosslinks

Figure 2-9 shows the XiNet diagram of crosslinks found with more than one CSM from the triplicate in-gel digested AurA 119-403 band**.** It shows previously unseen crosslinks in the C-lobe not found in the AurA band of the AurA+N-Myc experiment. We observed that the NTE and N-lobe of AurA were more reactive than the rest of the protein. It is unfortunate that we find no crosslinks involving the activation loop, because this would allow us to positively discriminate the conformation of the activation loop of AurA Apo.

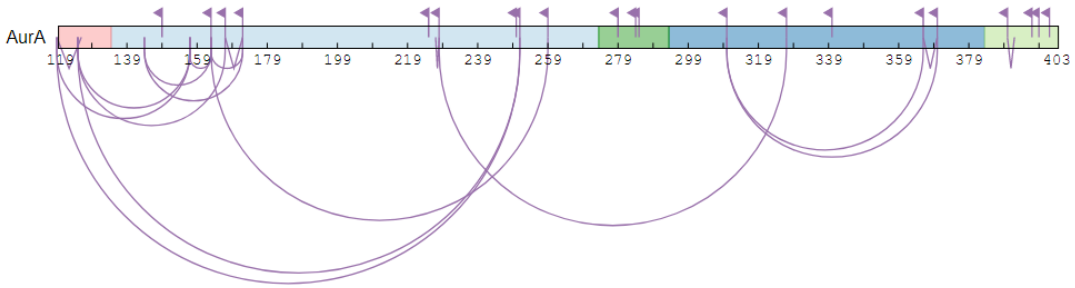


Figure 2-8 Crosslinking diagram of AurA 119-403 Apo from in-gel digestion

Overview of the identified crosslinks of Apo AurA KD 122-403 found more than a single CSM. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.2.2.1.2 Approach to understanding the structurally important crosslinks: normalisation

Analysis of the CSM distributions of the in-gel and in-solution digested AurA+Flag-tagged N-Myc reactions, showed that the number of CSMs of the observed crosslinks and monolinks presents as an exponential distribution, Figure 2-8. Where a few crosslinks and monolinks are extremely reactive and generate a plurality of the CSMs observed. Although this phenomenon is observed across a diverse set of target and system crosslinking experiments, it has never been investigated directly.[288]

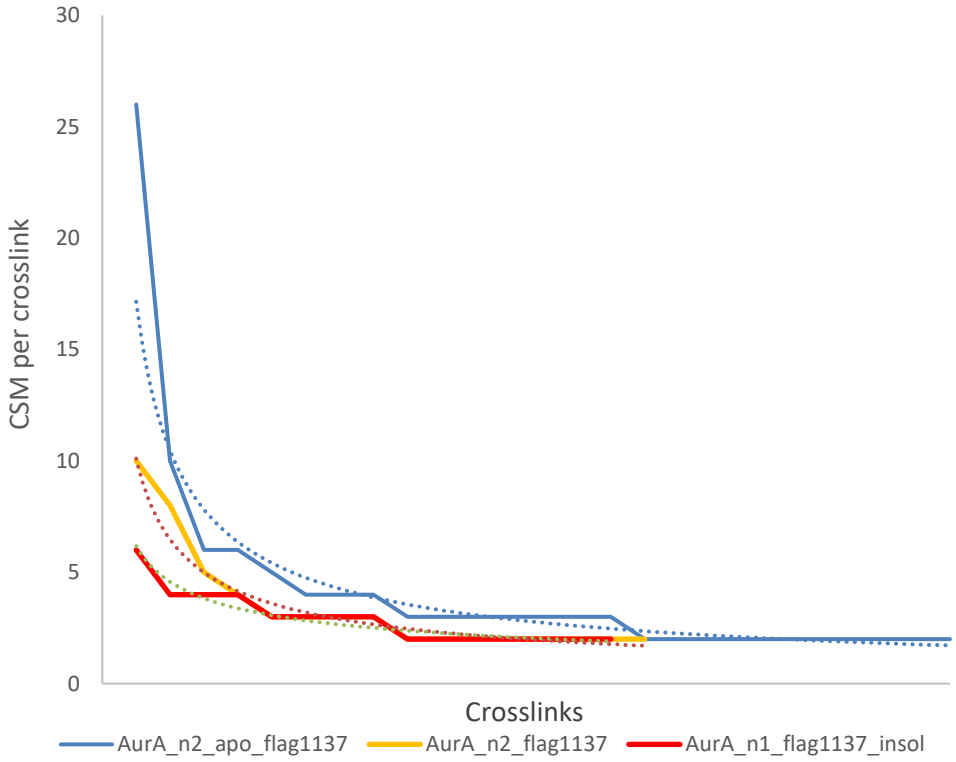


Figure 2-9 Plotting of the CSMs for each crosslink for the optimisation reactions

The CSM per crosslink for the measurements AurA+FLAG N-Myc 1-137 in gel, AurA band AurA+FLAG N-Myc 1-137 in gel AurA+FLAG N-Myc 1-137 in solution were extracted and plotted from crosslink with the most number of CSMs to least. Each condition has an average trendline plotted, dotted line.

In static proteins, it has been shown that highly reactive residues can effectively screen structural information from the domain of which they are a part by out-competing less reactive residues for crosslinker.[289 290] This is typically not a problem, since the protein-protein interfaces of static proteins can more easily be rationalised by fewer crosslinks. In dynamic proteins and complexes, such as kinase-IDR targets, we would expect a greater range of

reactivity. Therefore, it is not always obvious whether highly reactive residues are in fact screening important structural information. We examined literature for a method of data processing that would help us understand the inherent reactivities of our protein system, so that we could discriminate between crosslinks from residues that provide structural information, and those that are inherently reactive and do not generate structurally important information. Given the known structural dynamics of AurA Apo and Aura+TPX2, and the wealth of previously validated structural information, any process selected to deal with the problem outlined above could confidently be tested on our system.

Hogan *et al* reported using cross-linked peptide ion intensity to normalise the reactivity of crosslinked residues, thus allowing for the direct comparison of the abundance, as measured by total ion count, between crosslinks, accounting for their relative reactivity.[204] The number of CSMs for a crosslink is normalised, or calibrated, using the reactivity of the residue, as delineated by the MSM of the parent residues. Accordingly, a crosslink with a greater than average number of CSMs, resulting from the crosslinking of a pair of highly reactive residues, would be normalised to be of low importance. This is because if this pair of residues are within distance, one would expect them to generate a large number of crosslinks. However, a pair of residues of lower than average reactivity generating a large number of CSMs demonstrates that this pairing is within crosslinking distance, and is thus significant for the protein structure.

2.4.2.2.1.3 Changes to the Hogan normalisation approach

For the analysis of TPX2-AurA, we decided to apply a similar approach, but to use monolink CSMs for each residue, without using TIC data to quantify the reactivity of the observed crosslinks and monolinks. The Hogan protocol uses MS peak heights, as determined by the TIC, of the elution time window for each crosslink and mono-link. Additionally, Hogan used isotope-labelled crosslinkers as an internal control for the normalisation process; as isotope-labelled DSBU is unavailable commercially, and our digestion protocol increased the complexity of this approach, we hypothesised that an MSM and CSM count could be a post-hoc stand-in for TIC quantification.

As previously explained, the digestion protocols generated a heterogeneous composition of peptides for each crosslink and monolink, requiring their alignment for quantification. This negated our ability to use TIC for each crosslink or monolink, as this would require the summation of aggregated TICs for quantification. Hence, we instead explored if calibrating the reactivity of crosslinks could be performed using the normalised MSM count

for the residue pairs. We decided to compare the scoring of structures using all crosslinks with more than one CSM, and separately, all crosslinks with more than a single CSM and >50 % normalised reactivity score.

2.4.2.2.1.4 Normalisation of crosslink reactivity – theory

For monolinks, the two residues (Lys365 and Lys171) with the most CSMs account for ≈58 % of the total CSMs. The crosslink with the most CSMs accounts for ≈16 % of the total CSMs. As previously discussed in section 2.4.2.1.3, the crosslinks and monolinks match an exponential distribution. Given this large dynamic range of reactivity, and given also the potential for multiple conformations present in the sample, weighting the importance of crosslinks as either equal or proportional to their CSMs appears as though it would give misleading indications about which crosslinks must be considered important to the distance restraint analysis, and hence we need to normalise these values. After combining the MeroX search files of the triplicate measurement, the CSMs and MSMs were aligned so that each crosslink and monolink had a single CSM/MSM value associated with it.

Alignment gives a total monolink MSM value for each residue, and from this Z-scores were generated to allow cautious comparisons between conditions later, and for ease of normalisation. The same process was repeated for the crosslinks, combining intra-peptide and inter-peptide crosslinks.

$$Z_n = \frac{x - \mu}{\sigma}$$

Table 2-7 Variables of the normalisation equation

Symbol	Variable
Z_n	normalised MSM score for the residue
Zr	normalised reactivity for a crosslink
x	total number of aligned MSMs for the residue being normalised
μ	mean number of MSMs found for the condition
σ	standard deviation of MSMs found for the condition

To normalise reported crosslinks by reactivity, the Z-score for each crosslink was divided by the product of the two-cognate normalised monolinks, as shown in Table 2-7. If no monolinks were identified for a crosslinked residue, then the Z-score of CSM value of 0.1 was used. This 0.1 was not included in the average or standard deviation calculation for the Z-score.

The Hogan normalisation protocol does not explain its process for crosslinks involving residues lacking monolinks, hence our reason for using the ‘0.1’ MSM workaround; this value is functionally 0, but since 0 would not work mathematically in the equation, 0.1 was selected as the value to use.

$$ZR_{a\leftrightarrow b} = \frac{Z_{a\leftrightarrow b}}{Z_{n_a} \times Z_{n_b}}$$

Here $ZR_{a\text{-}b}$ is the normalised reactivity for a crosslink between residue a and residue b, from the product of the normalised monolinks ($^{Z}m_a$) and ($^{Z}m_b$). This process of normalisation was applied to the AurA Apo crosslinks.

2.4.2.2.1.5 Normalisation of crosslink reactivity – AurA Apo

Table 2-8 shows that ten crosslinks that were found to have more than one CSM and > 50 % normalised reactivity value. The process of normalising the reactivity of the crosslinks for AurA Apo generated a list of ten crosslinks with more than one CSM with above average normalisation value. Seven residues that were crosslinked but lacked monolinks used the 0.1 MSM equivalent. The lack of monolinks to these regions is persistent across the previous measurements of AurA. The crosslink Lys143-Lys171 was the most abundant crosslink, and so was ranked first by CSMs, however it moved to fifteenth place out of eighteen after normalisation. For example, despite the lack of Lys143 monolinks, as Lys171 is second for monolinks, the relative reactivity of the crosslink was dramatically reduced. Lys162-Lys258, Lys143-Lys162 and Lys156-Lys162 all have below average normalised monolinks reactivity paired with average crosslink reactivities.

Table 2-8 Normalised AurA Apo crosslinks

'Type' is the type of crosslink, R1 is the residue that bears the crosslink, R2 is the residue of the other crosslinked peptide, Z_r is the normalised cross reactivity, Z_n is the normalised MSM and >50 % is to identify crosslinks in the upper half of the normalised reactivity scores. The Z_r values are coloured positive (red) to negative (blue)

Type	R1	R2	Total CSMs	Z_r	>50 %	Type	R1	Total CSMs	Z_n	>50 %
Intraprotein	162	258	5	7.4	1	Mono	365	215	4.44	1
Intraprotein	143	162	6	5.54	1	Mono	171	81	1.37	1
Intraprotein	156	162	6	5.54	1	Mono	309	34	0.29	1
Intraprotein	227	326	4	1.6	1	Mono	227	26	0.11	1
Intraprotein	124	156	5	1.1	1	Mono	166	19	-0.05	1
Intraprotein	309	365	7	0.67	1	Mono	224	19	-0.05	1
Intraprotein	118	250	4	-0.31	1	Mono	283	17	-0.1	1
Intraprotein	124	250	4	-0.31	1	Mono	258	14	-0.17	1
intrapep	365	369	7	-0.84	1	Mono	162	12	-0.21	1
intrapep	118	125	3	-1.52	1	Mono	278	11	-0.23	1
Intraprotein	118	156	2	-2.84	0	Mono	369	11	-0.23	1
Intraprotein	162	171	7	-3.02	0	Mono	249	9	-0.28	0
intrapep	389	391	2	-3.25	0	Mono	326	9	-0.28	0
intrapep	226	227	5	-4.86	0	Mono	401	8	-0.3	0
Intraprotein	143	171	16	-5.49	0	Mono	250	7	-0.33	0
Intraprotein	124	166	2	-6.26	0	Mono	398	4	-0.39	0
intrapep	166	171	6	-8.15	0	Mono	148	2	-0.44	0
Intraprotein	309	369	6	-8.24	0	Mono	284	2	-0.44	0
						Mono	339	2	-0.44	0
						Mono	389	2	-0.44	0
						Mono	396	2	-0.44	0

2.4.2.2.2 Crosslink distance and models to evaluate crosslinking of Kinase-IDR

Assessing how far a DSBU molecule can span on a protein surface is critical to understanding if the crosslinks found are consistent with the distances found in the relevant crystallographic structures. This is asking the same question twice, relative to the perspective of the data being evaluated: firstly, is the model generated for the crosslinked protein or complex representative of the in-solution conformer, and, when annotated with the crosslinks found, are the residue side chain distances reported all within tolerance? Or, secondly, are the crosslinks generated the result of a poorly designed experiment that is producing out-of-tolerance crosslinks? So far, we have assumed for ease of analysis that AurA Apo and AurA+interactor exist as a homogeneous ensemble of a single conformer. This simplification is antithetical to the proteins and complexes that contain disordered regions or proteins. As with reactivity normalisation – an attempt to remove crosslinks that are redundant in determining the structure – here we must determine if over-long crosslinks are present but are explainable in the context of the expected dynamics of a domain or region, or whether the model generated provides an unrepresentative or incomplete picture of the in-solution structure. The side-chains of ordered and stable proteins undergo dynamic molecular motions that can bring two residues within this backbone distance even if their average position (or position in a crystal structure) is out of range. We expect that a dynamic region of a protein will be able to generate long crosslinks due to increased flexibility, but if tranches of crosslinks are over-long we must allow for the possibility that a second conformer or new model is required. Determining if a pair of residues that have been crosslinked are in or out of the tolerance of the crosslinker used is critical in showing that a structure is representative of the in-solution structure, whether generated from solely crystallographic methods, a molecular dynamics simulation, a homology model approach, or a combination of these. The crosslinkers DSS/BS3 have a backbone distance of 11.4 Å, whereas DSBU has a backbone distance of 12.5 Å. Merkley *et al* used crosslinking datasets from repositories and molecular dynamics simulations to determine distance restraints of 26-30 Å between Cα carbon of crosslinked residues for DSS/BS3.[242] Theoretically, 27.1-31.1 Å would be an appropriate maximal distance for DSBU. However, Ihling *et al* found that on average Cα-Cα distance of DSBU crosslinked BSA was 17.7 Å, with very few crosslinks under <10 or >35 Å.[227]

Herein we will consider inactive and active AurA conformers for annotation; to determine if DSBU crosslinks annotated to crystal structure models are

representative of the in-solution structure, and we will use 35 Å as the absolute cut-off. If the average Cα-Cα distance found deviated from a ≈18±10 Å this would be cause for revaluating the models used to produce the distances generated. The quality of measurement data, FDRs and spectrum scores, will not be validated using distance restraint data, as this has previously been shown to be inappropriate.[240] Instead, we have used the previously described FDR, MeroX peptide scores and CSM cut offs to ensure the quality of our data.

2.4.2.2.3 Generation and annotation of homology AurA Apo models

2.4.2.2.3.1 Generation of the models

The PDB contains many AurA kinase domain structures that fall into two broad categories: structures that are monomeric and fit into an active/inactive continuum, or structures outside this continuum (discussed in Section 1.3.3.1.2).

Structures that are monomeric and fit into an active/inactive continuum differ in where the activation loop sits, but the C-lobes remain relatively stable, only opening to allow activation loop movement. Three such structures were generated from crystallographic structures. PyXlinkViewer was used to annotate the models using crosslinks with two or more CSMs. This required all residues to be modelled to calculate a hypothetical Cα-Cα distance from which this model could be evaluated; hence the missing residues were replaced from the AlphaFold model. The structure of doubly phosphorylated AurA (PDB code: 1OL7), a canonically active structure, was generated using the same AurA construct as that used in our crosslinking experiment, and hence was adapted for annotation. The structure of AurA in complex with CD532 (PDB code:4J8M), an MLN-like inactive conformer, was used as an example inactive conformer. Minimal modifications were made to the 1OL7 structure to generate the model used to generate Cα-Cα distances. The 1OL7 structure was further modified by replacing residues 119-125 with those of PDB code 7O2V, which models the N-terminal extension binding into the F-pocket. This generated a third model, termed “NewCap.”

2.4.2.2.3.2 Annotation of the models

Models were annotated, using the PyXlinkViewer plugin of PyMOL[243], giving the Cα-Cα crosslinks in Table 2-9 and histograms of those distances in Figure 2-10. The histograms presented have an asymmetric distribution with

an extended tail shape, common to optimised samples across a broad spectrum of target types.[292]

Table 2-9 Distances found with AurA Apo

Overview of the identified crosslinks from the DSBU reaction of AurA Apo. Where 'Type' is the type of crosslink, R1 is the residue that bears the crosslink, and R2 is the residue of the other crosslinked peptide, Z_r is the normalised cross reactivity, Z_n is the normalised MSM and >50 % is to identify crosslinks in the upper half of the normalised reactivity scores.

Type	R1	R2	Total CSMs	Z_r	>50 %	1OL7 / Å	Newcap / Å	4J8M / Å
Intraprotein	162	258	5	7.4	1	18.1	18	20.1
Intraprotein	143	162	6	5.54	1	10.7	10.7	12.7
Intraprotein	156	162	6	5.54	1	17.9	18.2	17.4
Intraprotein	227	326	4	1.6	1	8.6	8.7	8.5
Intraprotein	124	156	5	1.1	1	13.6	17.6	21.2
Intraprotein	309	365	7	0.67	1	10.2	10.5	10.2
Intraprotein	118	250	4	-0.31	1	42.4	20.5	17.9
Intraprotein	124	250	4	-0.31	1	34.5	18.8	25.1
intrapep	365	369	7	-0.84	1	8.8	9.2	8.4
intrapep	118	125	3	-1.52	1	13.7	16.6	13.6
Intraprotein	118	156	2	-2.84	0	19.1	15.5	30.8
Intraprotein	162	171	7	-3.02	0	18.6	18.7	18.6
intrapep	389	391	2	-3.25	0	6.7	6.8	6.5
intrapep	226	227	5	-4.86	0	3.8	3.8	3.8
Intraprotein	143	171	16	-5.49	0	12.8	12.9	11.4
Intraprotein	124	166	2	-6.26	0	25.1	21.1	27.2
intrapep	166	171	6	-8.15	0	8.5	8.5	8.5
Intraprotein	309	369	6	-8.24	0	7.8	7.8	7.5

This distribution, shown by MD and experience, should have a maximum number of crosslinks falling on the average Cα-Cα distance, with more being longer than average and fewer shorter than average. This asymmetry is a consequence of crosslinker conformation dis-favouring shorter than average crosslinker distances. Overlong-crosslinks, or crosslinks out of tolerance for the crosslinker, form the extended tail. It is understood the tail, formed of a grouping of overlong crosslinks, is discontinuous with the curve of in-tolerance crosslinks, suggesting that a domain or region is misplaced or that a dynamically fluctuating region is in a secondary conformer.[280] The cross-links were annotated onto the three models of AurA Apo (PDB: 1OL7, 4J8M and the 1OL7/1OL2 NewCap). The histogram of Cα-Cα distances (Figure 2-10) shows that most of the cross-links are within the maximum distance constraint of 35 Å.

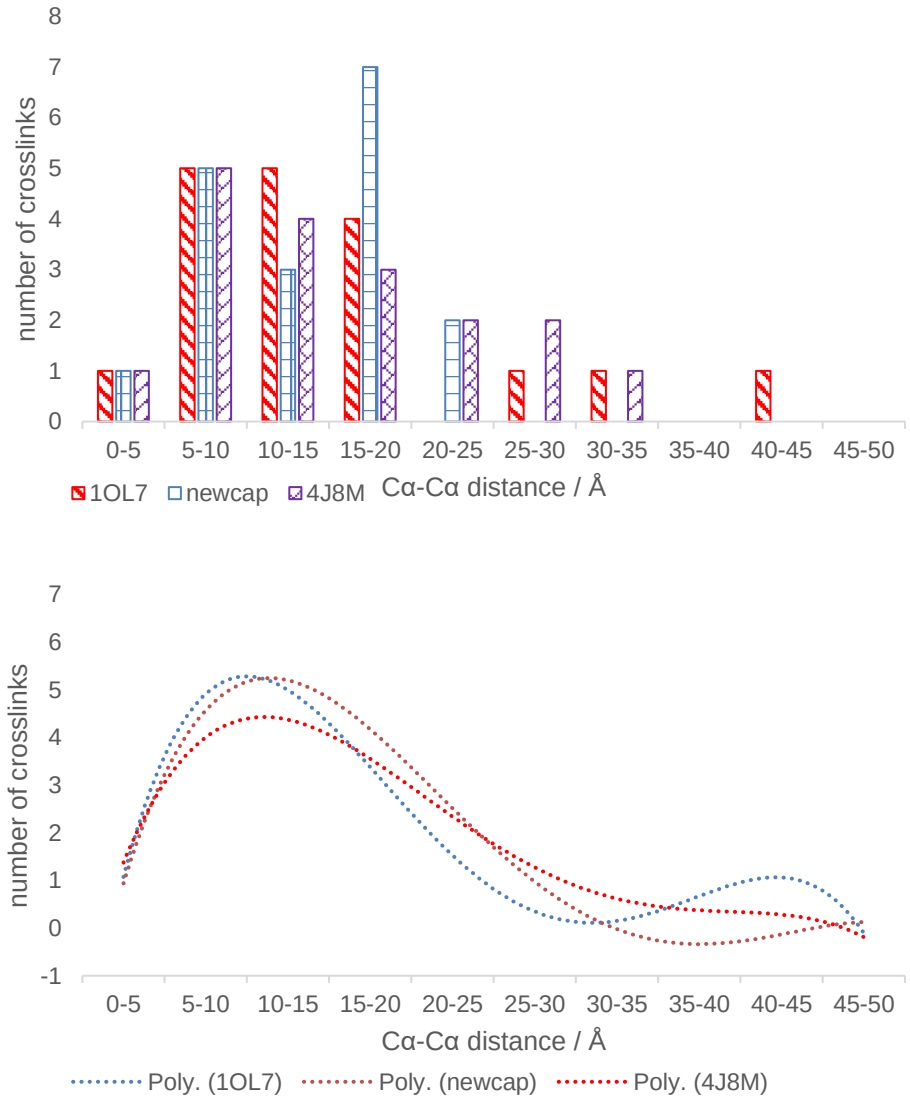


Figure 2-10 Histogram of Cα-Cα distances found for AurA Apo crosslinks annotated on model structures

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM.

2.4.2.2.3.3 Structure analysis of Crosslinking of AurA APO from 1OL7

The canonically active structure for AurA annotated with crosslinks showed that only a single crosslink exceeded the Cα-Cα distance tolerance of DSBU: the free amino group of Gly119-Lys250 (Figure 2-11). It is interesting that no crosslinks were found involving the CTE, despite it containing Lysine residues, (residues 392-403) as this is assumed to be as dynamic as the NTE (residues 119-130).

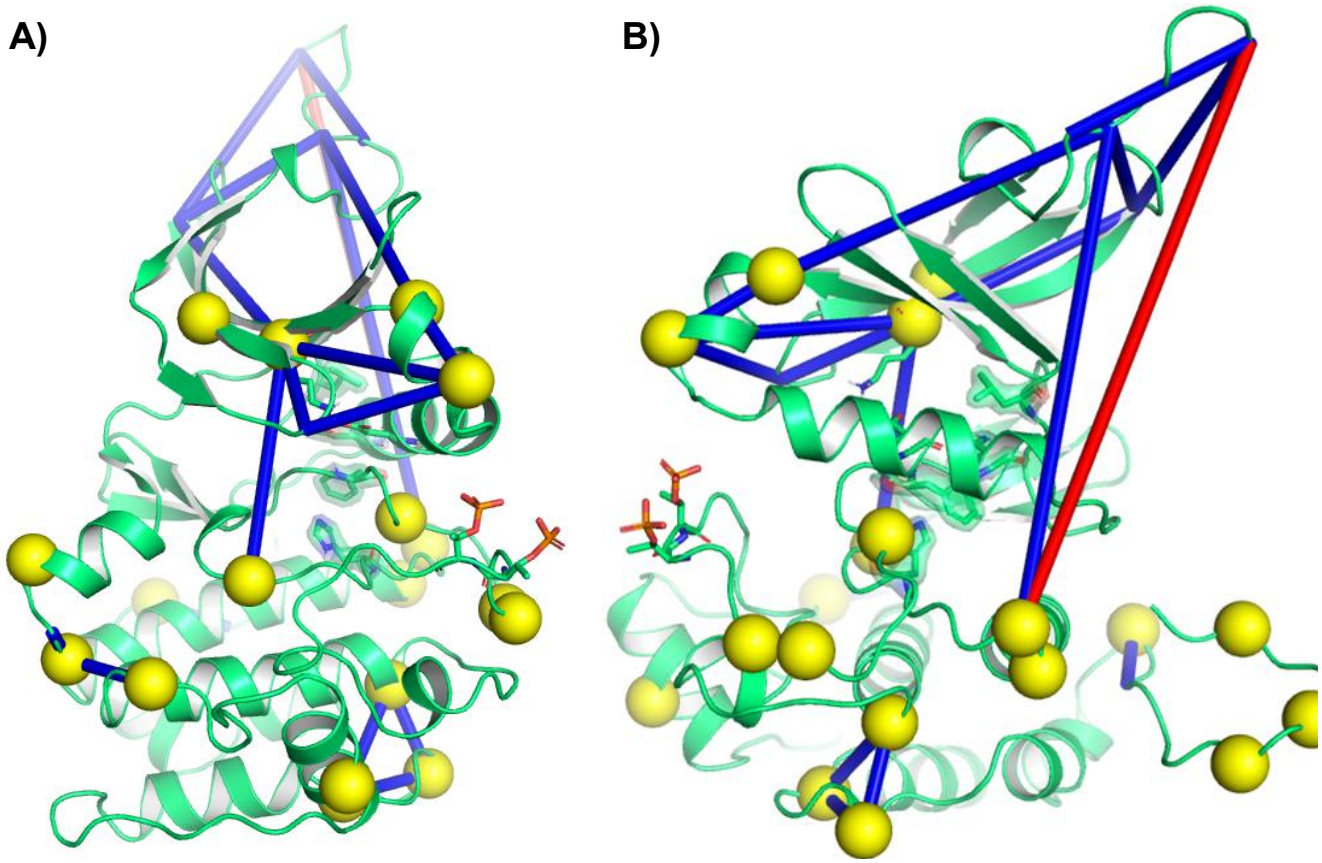


Figure 2-11 Aura Apo crosslinks on 1OL7

AurA Apo n>1 crosslinks annotated to a model built from 1OL7. Cα-Cα distances <35 Å shown in blue and >35 Å in red. Monolinked residues shown with yellow spheres. A) Single crosslink from N- to C-lobe, Lys162-Lys258, found with five CSMs and the highest normalised reactivity. No crosslinks were found on the activation loop but Ser278, Ser283 and Ser284 were monolinked. B) The Helix B and glycine-rich loop and the NTE were the locus of reactivity. The N-term-Lys250 crosslink was the only one found overlong.

2.4.2.2.3.4 Structural analysis of the crosslinking of AurA Apo with 7O2V

Previous structural studies have not found any evidence of a stable static structure of the NTE.[49] It is expected that the NTE of AurA could be dynamically sampling the surface of AurA, and the crosslinks suggest it could be bound to or weakly interacting with the N-lobe. Thus, we searched the PDB for a crystallographic structure that contained a structured NTE. In

2021, Rapposelli *et al* published a His-TEV tagged, wild type AurA 125-391 construct structure with a 2-oxindole-based derivative type II kinase inhibitor. This yielded an X-ray structure with N-terminal region,119-126, which is missing from 4CEG, structured and placed over the F-pocket. The structure shows 116-390 but with activation loop 280-289 and the C-terminal 391-403 missing. The construct was produced in *E. coli* with Lambda phosphatase co-expression system, meaning that the AurA structure was measured without any phosphorylation. Crosslinks were mapped onto this structure (Figure 2-12). The average Cα-Cα in this structure is reduced from 17.5 to 14.8 Å. The model was created by grafting NTE AurA from 7O2V onto the 1OL7 model, as 7O2V contains an unstructured activation loop and is crystallised with an inhibitor. The mixed model was used as we wanted to directly compare the difference between the Cα-Cα distances only. When compared to the AlphaFold predicted model, the 119-250 crosslink in the new model was within the Cα-Cα distance for DSBU and the Lys124-Lys156/Lys250 crosslinks, previously out of tolerance, shorten to within ≈1 Å of the median crosslink, previously out of tolerance.

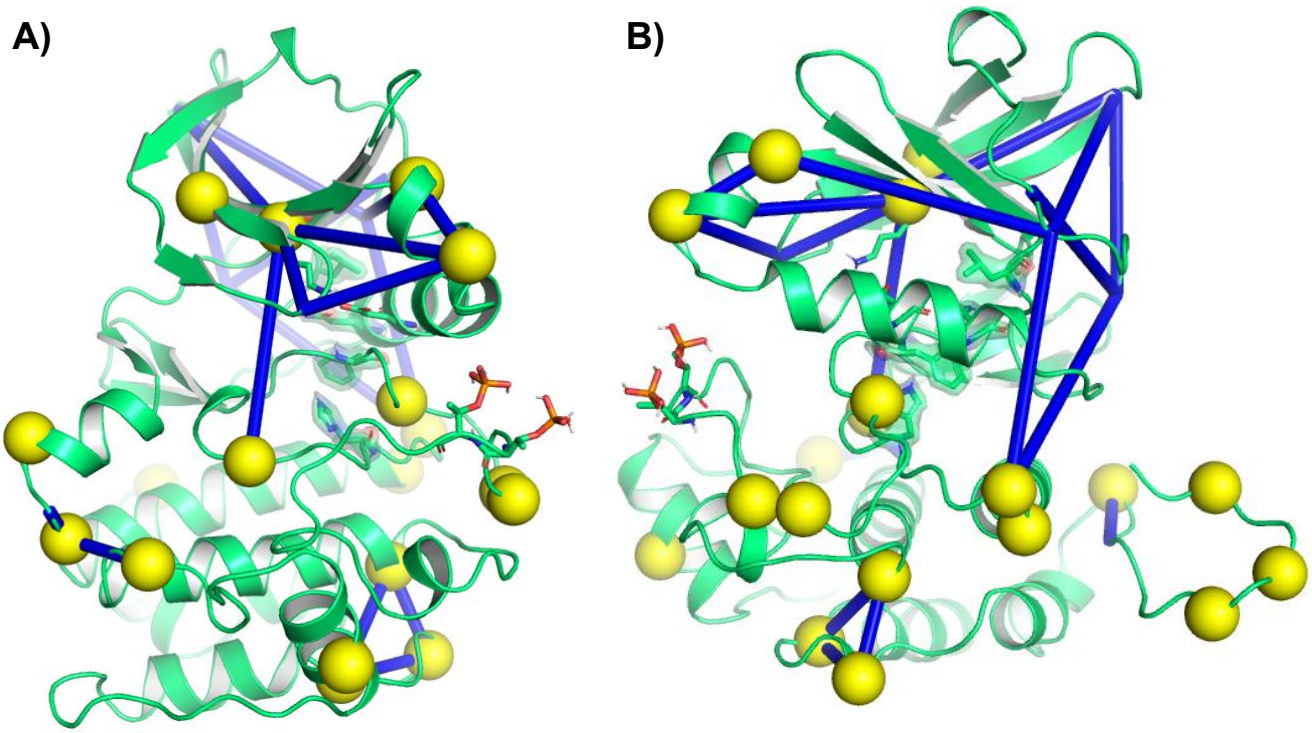


Figure 2-12 Aura Apo crosslinks on 7O2V

'NewCap' model AurA 119-403, adapted from PBD 7O2V. Blue crosslinks are under 35 Å Cα- Cα, red are over-long. Yellow spheres show the Cα carbon of monolink residues. A) Front face of the model. The NTE bound behind the N-lobe into the F-pocket. B) Side face of the model showing the NTE crosslinks within an acceptable tolerance for DSBU.

2.4.2.2.3.5 Structural analysis of the crosslinking of AurA Apo with 4J8M

An MSP1-like structure was selected, and the missing residues replaced as previously described and annotated using the AurA Apo crosslinks (Figure 2-13). The crystal structure has CD532, a type II inhibitor, bound (omitted from the figure), and features the phosphorylated subdomain of the activation loop bound to Helix D (PDB code: 4J8M). This places the NTE binding over the W-pocket and away from the F-pocket, with possible contacts of the NTE to His187 and Arg189. For both the NewCap and 4J8M models, the distances of crosslinks involving the N-terminus improve relative to the 1OL7 model, with some changing from out of tolerance to in; in addition, they produce histogram curves without the dis-continuous tail.

However, as we observe no crosslinking involving the activation loop, we cannot positively discriminate between the canonical active activation loop position or the MSP1-like position. However, as we found Ser283 to be monolinked at a low rate, we can infer that the reactivity of the activation loop matches the active position, since we would expect the MLN-like position to generate crosslinks with the highly reactive Helix D.

2.4.2.2.4 Effect of using normalised reactivity cut-off

Although there is some minor improvement to the shape of the histogram curves with the reduced crosslink lists (Figure 2 -14), as the number of crosslinks is small it is hard to draw strong conclusions from this. Specifically, the NewCap histogram maximum correctly is placed over the average DSBU distance. Normalisation removed the over-long crosslinks N-terminus119-Lys250 and Lys124-Lys250, and shorter than average intra-peptide Ser226-Lys227distance. Noticeably, crosslinks with Lys171 and Lys166 were removed. Both of these residues are found on Helix B, and thus are both very solvent accessible.

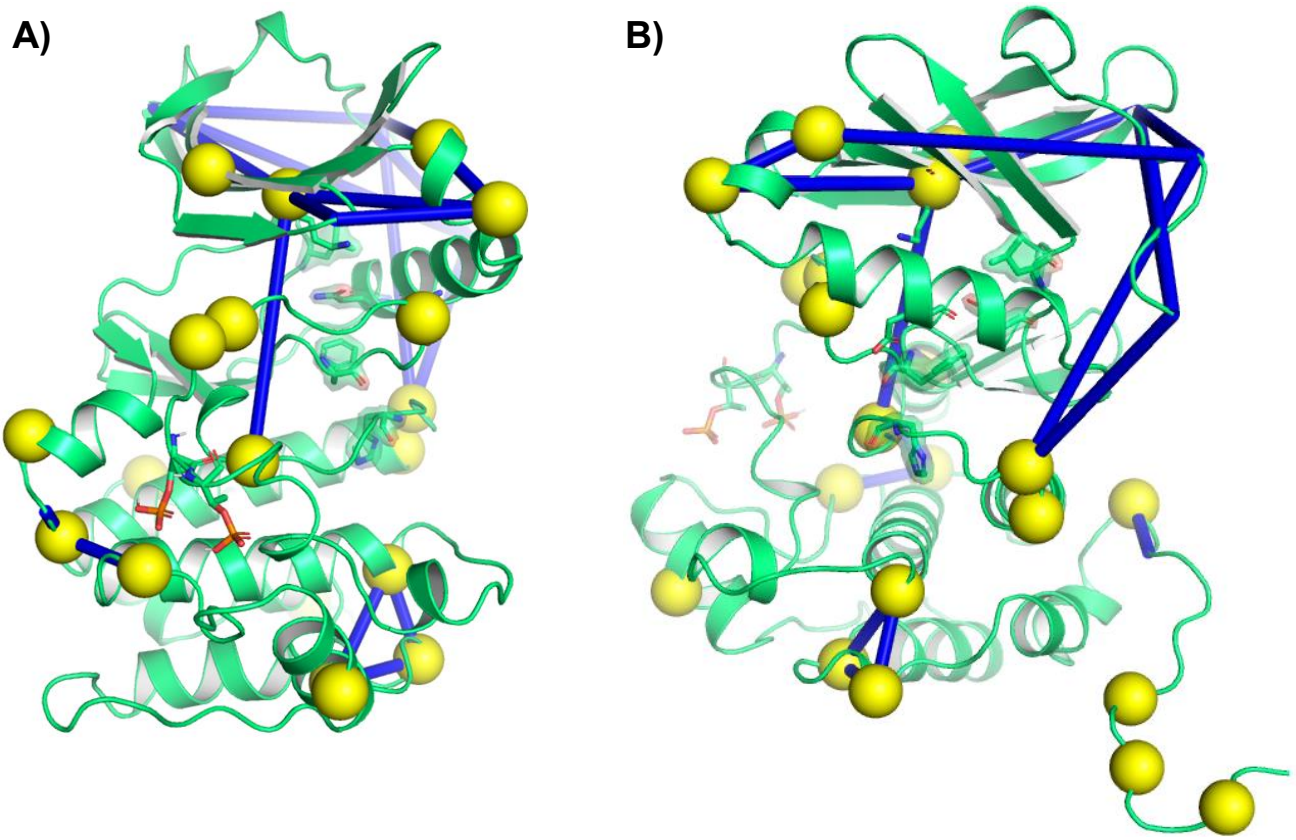


Figure 2-13 AurA Apo crosslinks on 4J8M

MSP1-like model of AurA 119-403, adapted from PBD 4J8M. Blue crosslinks are under 35 Å Cα-Cα, red are over-long. Yellow spheres show the Cα carbon of monolink residues. A) Front face of the model. The NTE bound behind the N-lobe into the F-pocket. B) Side face of the model showing the NTE crosslinks within an acceptable tolerance for DSBU.

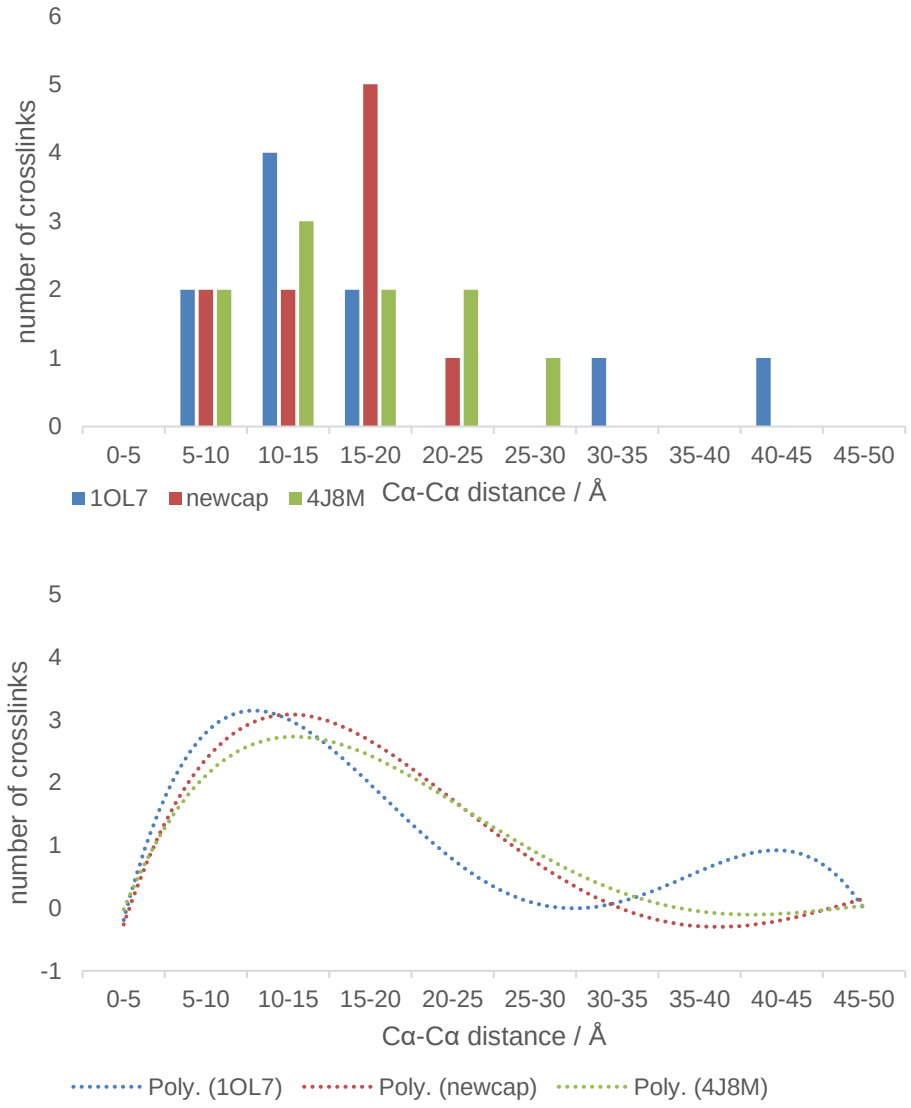


Figure 2-14 Histogram of normalised Cα-Cα distances found for AurA Apo crosslinks annotated on model structures

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM, and >50 % normalisation value.

2.4.2.3 XL-MS of AurA+TPX2 and analysis of the distance restraints generated

2.4.2.3.1 Aurora-A + TPX2 XL-MS

The crosslinking reaction of AurA KD 119-403 with TPX2 1-43 using ≈95-fold molar excess of DSBU generated a band at ≈40 kDa (Figure 2-7). This was excised and once processed and measured the combined results for each condition were searched using MeroX. In addition, a monomeric AurA band

and a faint band corresponding to the weight of 1+1 dimeric complex were observed. However, these were not processed and measured.

2.4.2.3.2 Analysis of crosslinked peptides of AurA +TPX2 by MeroX

After combining the triplicates for analysis, searching with MeroX returned 44 crosslinks found with 467 CSMs. As shown in Table 2-10, aligning then removing crosslinks with a single CSM removes 1.8 % of the total CSMs found and 18.2 % of the crosslinks. Searching the non-crosslinked peptides using PEAKS found coverage of TPX2 6-39 but AurA was “missing” residues 144-162, 220-224 and 340-344.

Table 2-10 Breakdown of crosslinks and monolinks found for in-gel AurA KD crosslinked to TPX2 1-43

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	AurA				AurA-TPX2 protein		TPX2			
	intra-protein		Mono-link				intra-protein		Mono-link	
	Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
n≥1	13	43	20	328	11	96	0	0	0	0
n>1	9	39	17	325	10	95	0	0	0	0

2.4.2.3.2.1 Results of the aligned crosslinks

Figure 2-15 shows the XiNet diagram of crosslinks found with more than one CSM from the triplicate in-gel digested AurA+TPX2 band**.** It shows a noticeable reduction in reactivity in AurA, particularly in the vicinity of the activation loop and in the intra-N-terminal region. However, we do not see a reduction in monolinks or an activation loop to TPX2 inter-protein crosslink. As with the AurA+N-Myc samples, no TPX2 intra-protein crosslinks or monolinks were observed.

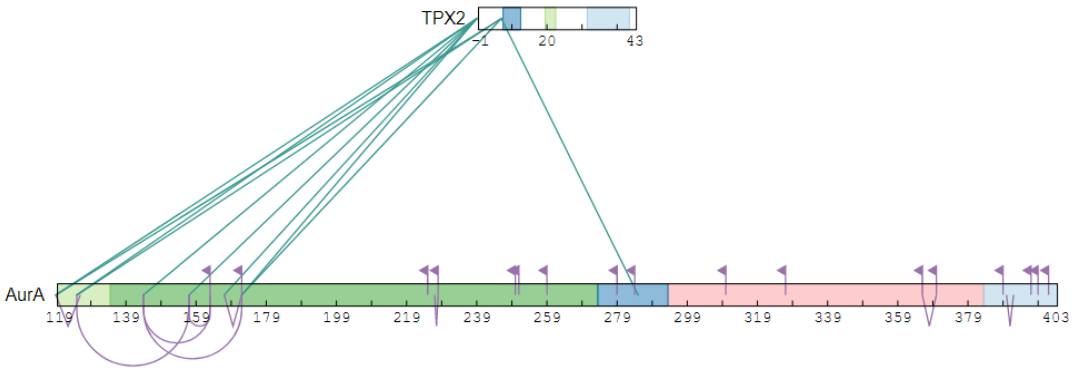


Figure 2-15 Crosslinking diagram of AurA 119-403 with TPX2 1-43 from in-gel digestion using crosslinks with more than a single CSM

Overview of the identified crosslinks between AurA KD 119-403 and TPX2 1-43 with more than one CSM. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For TPX2 (L to R), residues that bind into the Y-pocket, F-pocket and W-pocket. For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.2.3.3 Normalisation of AurA+TPX2 crosslinks

As no TPX2 monolinks were found, this made normalisation problematic, and the nominal CSM value of 0.1 was used, as previously. No trend of residues or domains that were favoured emerged after normalisation. AurA crosslinks were favoured to some extent over AurA-TPX2 crosslinks. A table of normalisation can be found in the Appendix 5.2.2.5.

2.4.2.3.4 Generation and annotation of homology AurA+TPX2 models

PyXlinkViewer was used to annotate two manually produced homology models. The first was constructed from 1OL5 with AlphaFold used minimally to build back in unstructured portions: 119-122/388-403 of AurA, and TPX2 1-6 and 23-29. The second was an alternate model that features an activation loop with a continuous helix, structured subdomain out (PBD code: 4CEP). The annotation shows that the reactivity of AurA NTE changes dramatically with the addition of TPX2. This is to be expected, as the NTE is observed binding to TPX2 in crystallographic structures. The crosslink lengths generated from model annotation can be found in the Appendix 5.2.4.1.3. Figure 2-16 shows that there are three crosslinks with a Cα-Cα

distance greater than 35 Å. But these involve the N-terminus of TPX2, which is unstructured.

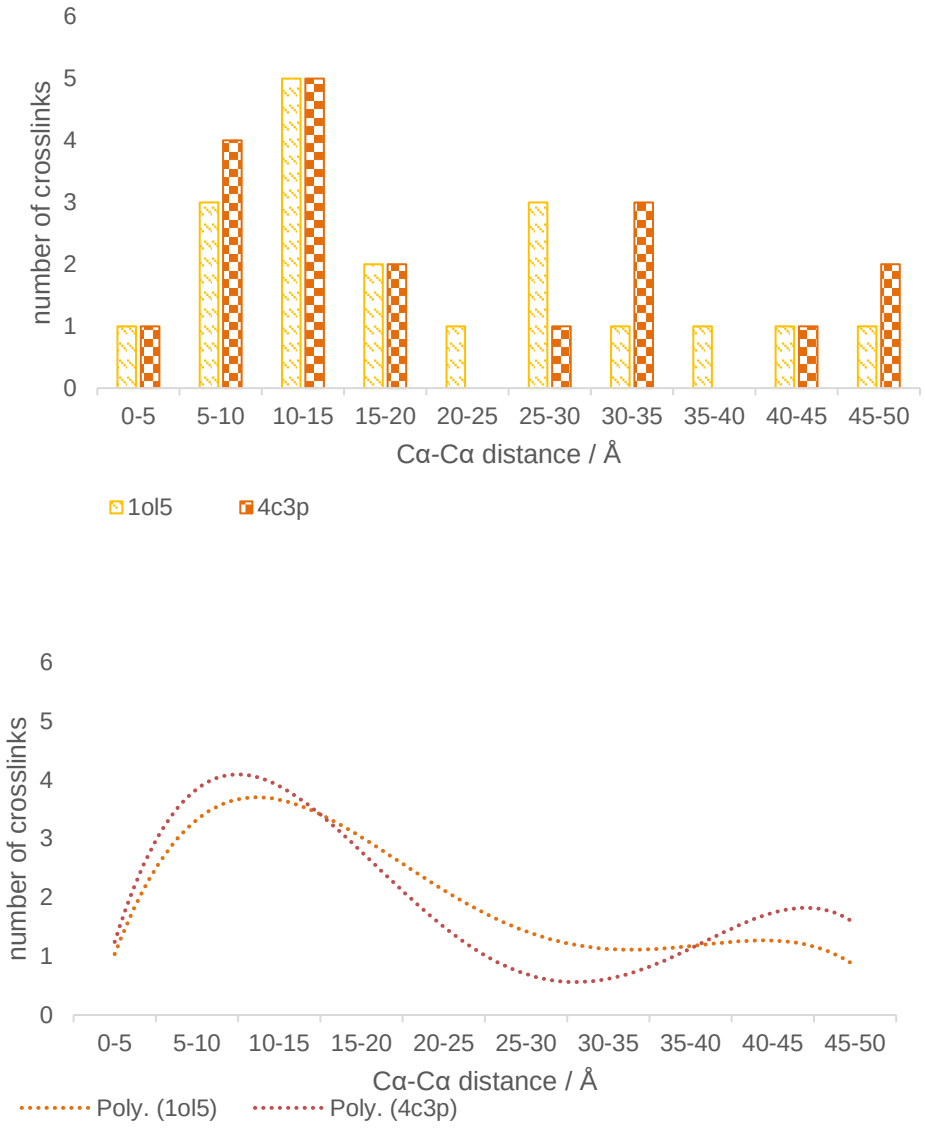


Figure 2-16 Histogram of AurA+TPX2 distances with crosslinks with >1 CSMs

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM.

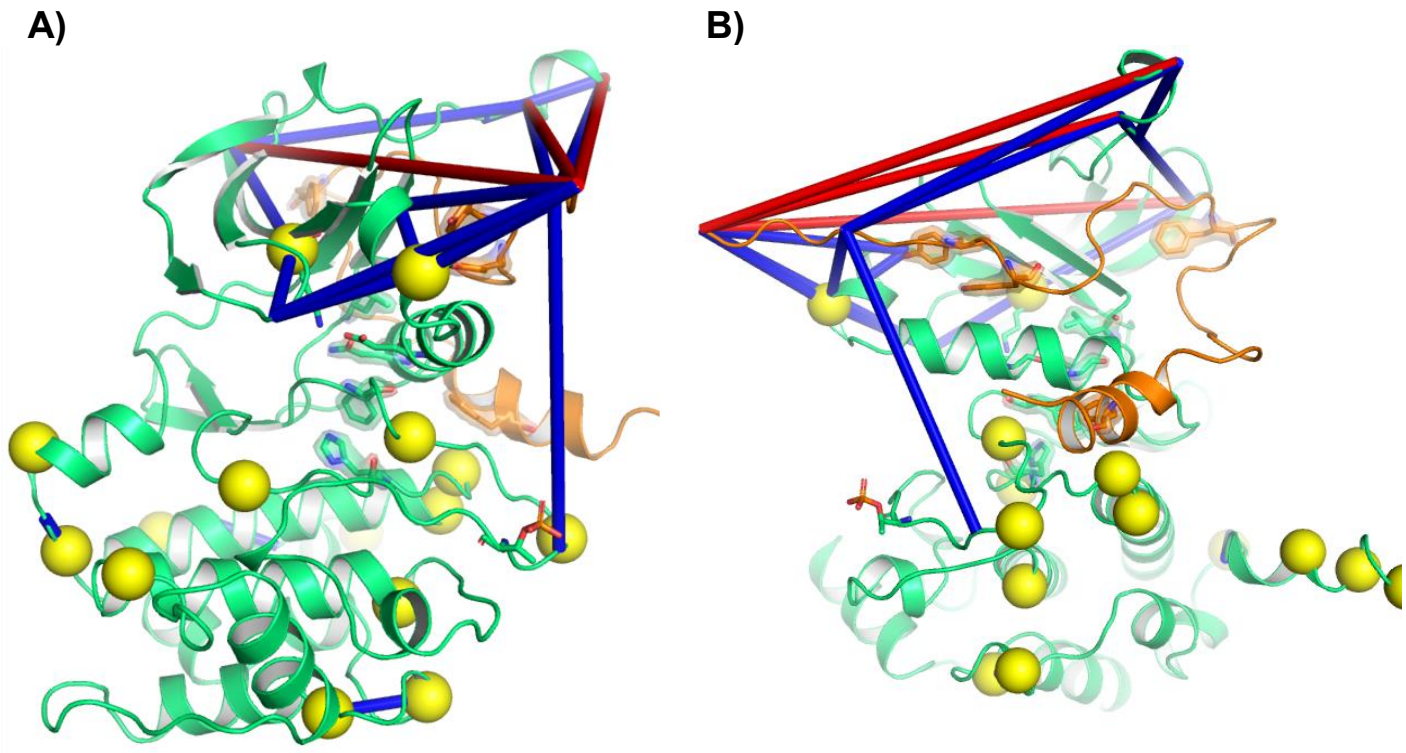


Figure 2-17 AurA + TPX2 crosslinks annotated to 1OL5 structure

Active AurA+TPX2 model, adapted from PBD 1OL5. AurA is in green and TPX2 in orange. Blue crosslinks are under 35 Å Cα- Cα, red are over-long. Yellow spheres show the Cα carbon of monolink residues. A) Front face of the model. The NTE bound behind the N-lobe into the F-pocket. B) Side face of the model showing the NTE crosslinks within an acceptable tolerance for DSBU.

2.4.2.3.5 Structural analysis of the crosslinking of AurA+TPX2 with 1OL5

Figure 2-17 shows the homology model, built from 1OL5, AurA 119-403 and TPX2 1-43 (orange), annotated with crosslinks with more than a single CSM (this was not the normalised data). Blue crosslinks are under 35 Å Cα- Cα, while red are over; yellow spheres show the Cα of monolink residues. The front face of the model shows that the N-terminus of TPX2 is very reactive and generates many over-long crosslinks to the N-lobe of AurA, with no crosslinks spanning the N-/C-lobes. The side face of the model has TPX2 residues shown for the Y-, F- and W-pockets of AurA (definition shown in Figure 1-19). Strikingly, TPX2 Lys38 is not found to have any crosslinks. The crosslink TPX2 Lys5 to AurA Ser284 was observed. It should be noted that no crosslinks involving the activation loop were observed with AurA Apo.

2.4.2.3.6 Structural analysis of the crosslinking of AurA+TPX2 with 4C3P

To explore if a different model of AurA with the activation loop in a different pose was consistent with the crosslinks found, a model generated from a dimeric AurA was generated (Figure 2-18) and annotated with the crosslinking data (Figure 2-19). Zorba *et al* collected a structure of Aurora A 122-403 bound to TPX2 1-45 and an ATP-mimic, AMPPCP, that formed a four-member complex, formed of dimeric Aurora A+TPX2.[141] As part of the crystallographic structure, the dimers swap part of the activation loop phosphorylated and structure subdomains, residues 281-306, (Figure 2-18).

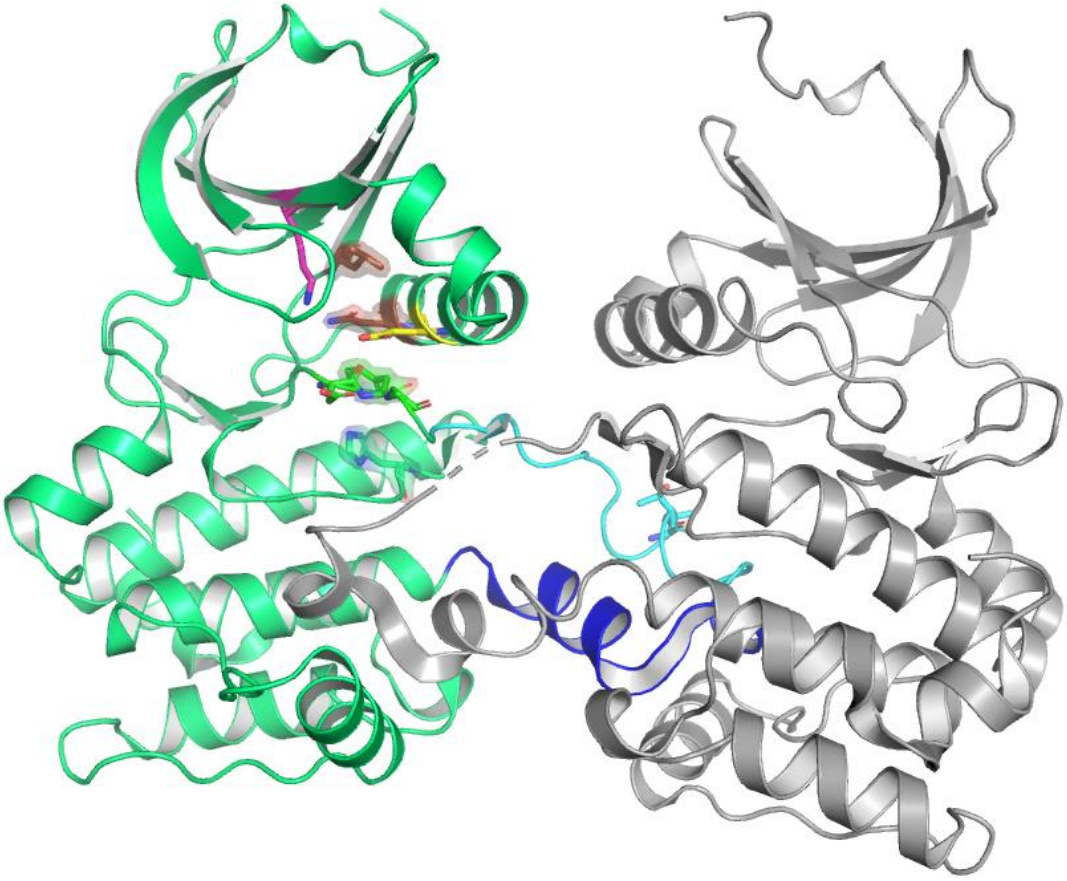


Figure 2-18 Generation of AurA dimer-like structure

The crystallographic structure of dimer swapped 1+1 dephosphorylated AurA+TPX2 complex (PBD code: 4C3P), TPX2 and AMP-PCP removed. Monomer 1 (green) with activation loop structured subdomain shown rotated into the second monomer (silver) and the activation loop phosphorylation subdomain held by the DFG subdomain of monomer 2. 4C3P structured subdomain in navy, phosphorylated subdomain in cyan. TPX2 omitted, exchanged activation loops.

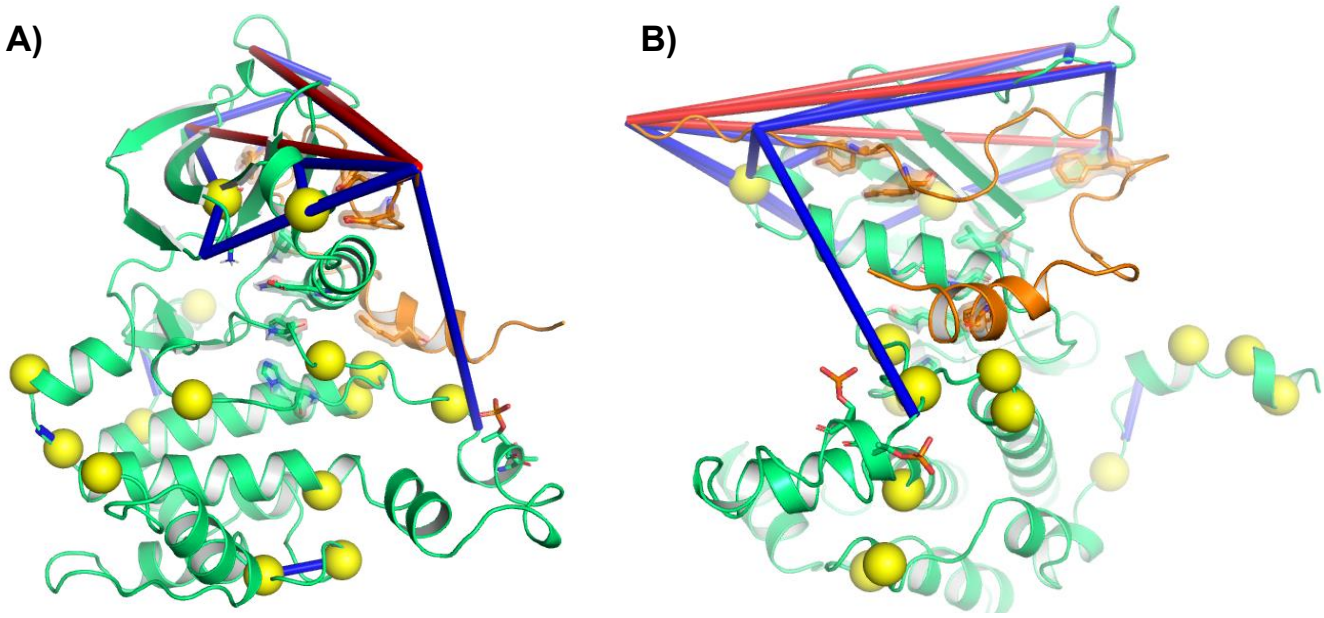


Figure 2-19 Aura+TPX2 crosslinks annotated on dimer-like structure

The annotated homology model, built from 4C3P, AurA 119-403 and TPX2 1-43 (orange) annotated with crosslinks with more than a single CSM. Blue crosslinks are under 35 Å Cα- Cα, red are over-long. Yellow spheres show the Cα of monolink residues. A) Front face of the model with the activation loop in the open pose. The crosslink AurA Ser284-N-terminus of TPX2 is still within tolerance. B) Side face of the model. In the dimeric structure, TXP2 36-43 is absent, leaving the W-pockets of AurA empty.

2.4.2.3.7 Effect of using normalised reactivity cut-off

Neither histogram curves generated from the annotation of crosslinks with the reduced crosslink list present as typical Cα-Cα distance restraint histograms (Figure 2-20). For both models, under both conditions, crosslinks with more than a single CSM and with >50 % normalised reactivity were present, with the portion of over-long crosslinks from the N-lobe, especially the NTE, crosslinked to TPX2. Nonetheless, these overlong crosslinks, especially those found with >50 % reactivity normalisation, are from AurA regions that are unstructured, and that we expect to be dynamic. It is important that these crosslinks do not generate seemingly random crosslinks to unexpected regions of AurA as this would suggest that these disordered regions are held in solvent for most of the conformation's lifetime.

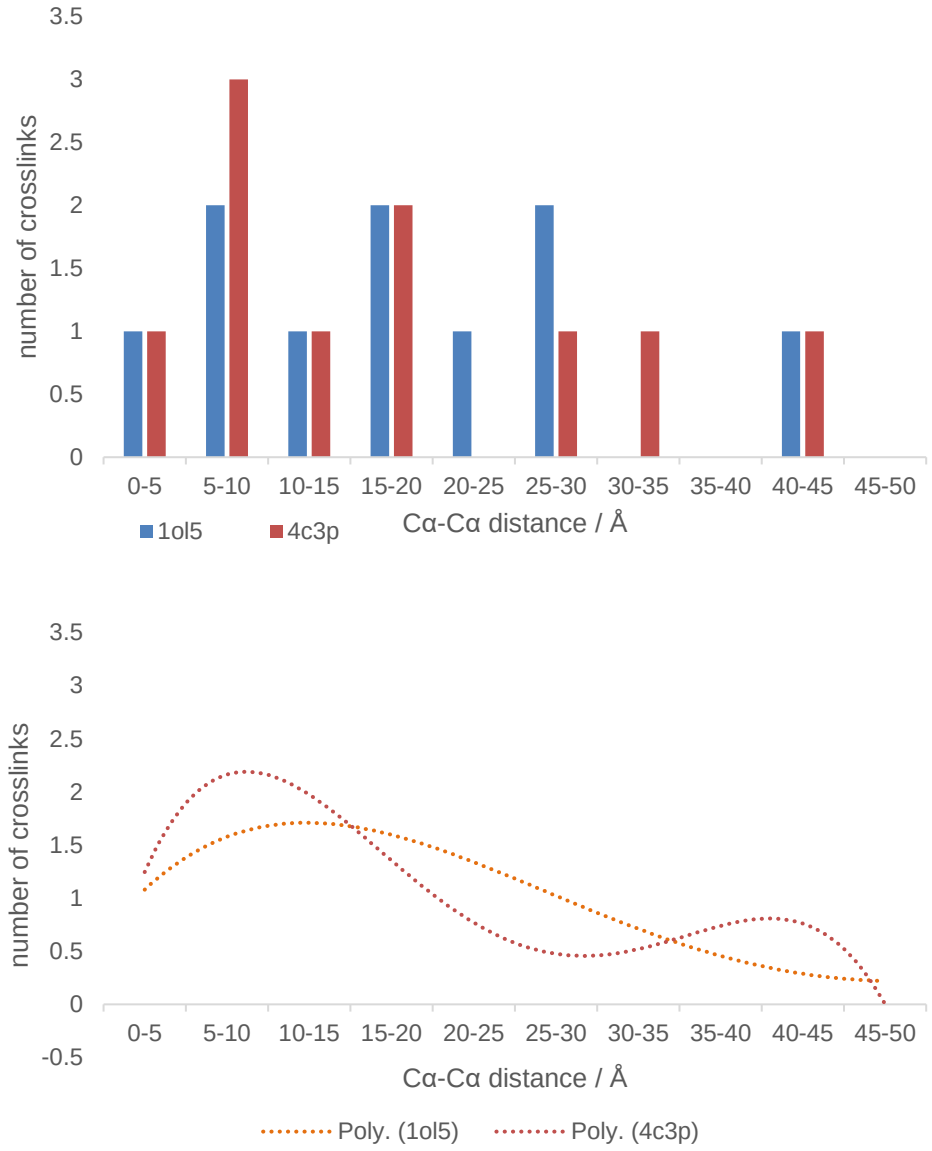


Figure 2-20 Histogram of AurA+TPX2 distances with crosslinks with >1 CSMs and reactivity >50 %

Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM and normalised reactivity above 50 %. A) Cα-Cα atom distances between cross-linked residues are plotted on a histogram. B) Averaged trend lines.

Critically, the activation loop crosslinks, albeit weakly, to the N-terminal residues of TPX2: AurA Ser284-TPX2 Lys5 with 2 CSMs, and AurA Ser283-TPX2 N-terminus with 1 CSM. The addition of AurA Ser283 with 3 MSMs adds to the picture of AurA+TPX2 dynamics that are in line with the X-ray structure and biochemical investigations of the activation loop dynamics of AurA in the presence and absence of TPX2. Furthermore, it provides further evidence for the paradigm that AurA dynamics are tempered by the addition of TPX2, altering the equilibrium of inactive-active conformations of AurA

and stabilising the active position. This could be related to an alternative activation mechanism, seen in other kinase families. No evidence has been seen of this for AurA.

2.4.2.4 Changes to AurA monolinks with the addition of TPX2

We saw with the examination of the crosslinks with AurA+TPX2 that they agree with the crystal structure and biophysical data. However, as we see no crosslinks involving the activation loop in AurA Apo it is harder to conclusively explore the conformational dynamics of this region. An explanation for this poor reactivity, other than reduced coverage due to complications with the digestion and ionisation of a doubly phosphorylated peptide, could be that either more reactive regions out-compete the less reactive serine residues for crosslinkers, or that these residues remain un-reactive due to being bound to the N-lobe, as appears to be the case on the crystal structure. Biophysical data suggests that the activation loop is somewhat dynamic and not wholly bound in place after phosphorylation has occurred. This protein construct contains a double-phosphorylation mark on its activation loop.

To explore this phenomenon, we hypothesised that the normalised monolinks could be used as a post-hoc covalent foot printing experiment. MS foot printing experiments, for example using covalent labelling or hydrogen-deuterium exchange, analyse the rates of labelling under different conditions, using surface accessibility to examine conformational changes or PPIs.[293] We hypothesised that changes to normalised monolinks with the addition of TPX2 1-43 could be used as a post-hoc labelling experiment in addition to the distance restraints found. Figure 2-21 shows the number of monolink MSMs found for each residue of AurA in the presence and absence of TPX2. However, when plotting this as the difference in normalised monolink reactivity generated, inspecting the changes was easier. The changes are displayed in Figure 2-21 with six of the twenty-six residues found to have differences in mono-link labelling greater than one standard deviation (annotated in Figure 2-22 and Figure 2-23). Lys227, adjacent to Helix G, was the only residue of the six that had increased reactivity with TPX2. Residues Lys171 and Lys365 are the most reactive residues of both AurA-TPX2 and AurA+TPX2. However, because fewer monolinks were found overall for AurA+TPX2, looking at the absolute change (upper panel) doesn’t show the changes in residue reactivity.

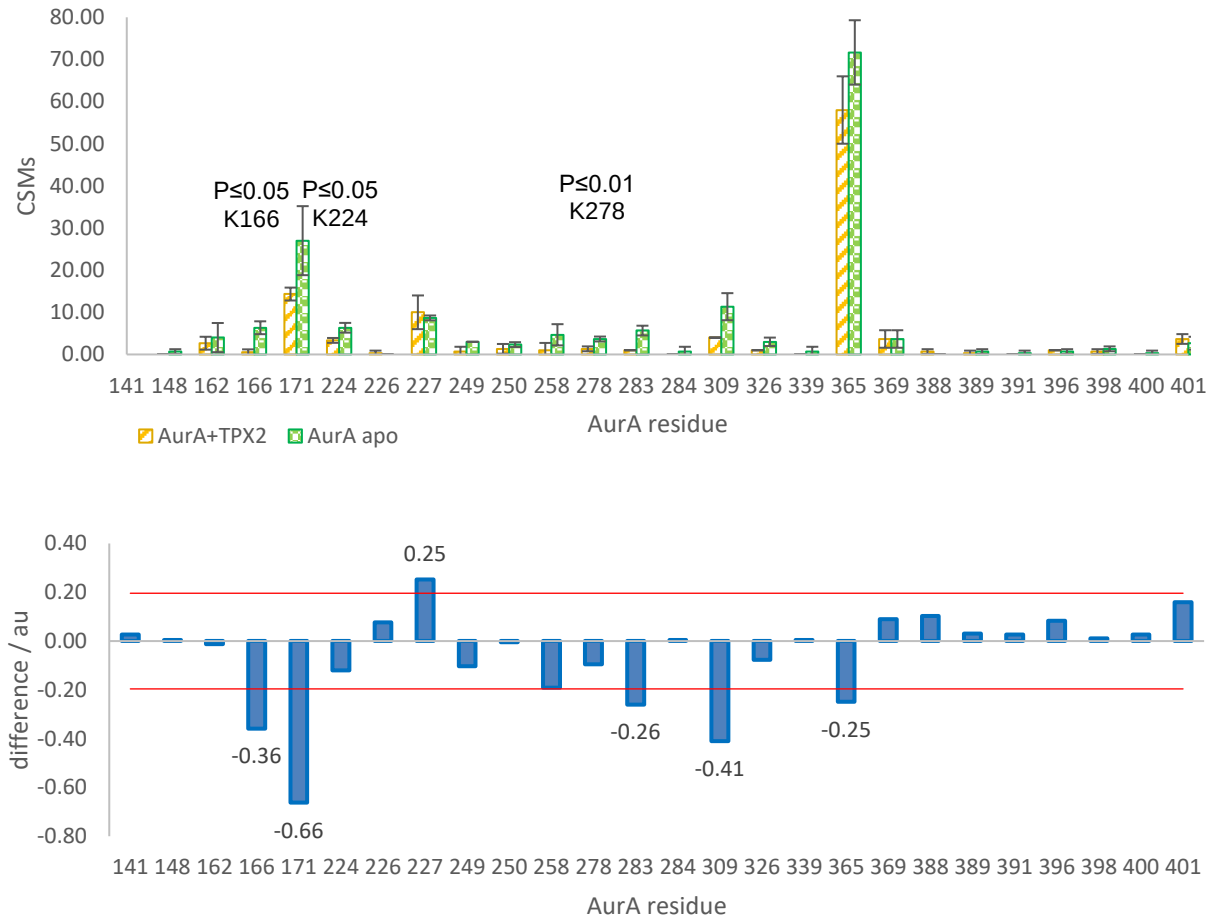


Figure 2-21 Changes to normalised crosslinks AurA±TPX2

Changes in crosslink abundance for each unique monolinked residue site of AurA upon binding of TPX2 were used to explore the changes in monolink reactivity as a post-hoc foot printing experiment. Upper – Average Monolink CSMs for AurA±TPX2; average MSM for each residue shown for both conditions. Each residue is the mean of three measurements, AurA with TPX2 (orange) or without (green). Student's T-test applied (two-tails, heteroscedastic) shows that residues Lys166, Lys224 and Ser278 are significantly different as p value ≤ 0.01. Error bars generated from the standard deviation of the three measurements for each residue. Lower – Changes to normalised monolink abundance with the addition of TPX2 1-43. The difference in normalised reactivity between AurA in the presence and absence of TPX2 is shown for each residue. The increase in reactivity is shown by a positive number; with the addition of TPX2 and subsequent decrease in reactivity shown by a negative number. Residues where the change in normalised reactivity is greater than one standard deviation of the dataset (shown as the red lines) are highlighted. Each condition (±TPX2) is normalised so that changes in reactivity of a residue relative to reactivity of the entire condition can be examined.

For example, Lys227 generated the same number of monolink CSMs for AurA±TPX2, but as AurA+TPX2 has fewer monolinks overall, thus has an increased relative reactivity in the presence of TPX2. As residues Lys166, Lys171, Ser283 and Lys309 fall under or are particularly near the binding site of TPX2, their reduction in reactivity can confidently be explained by uncomplicated foot printing. The change in reactivity of two other monolinks was harder to explain. Firstly, residue Lys227 increases in reactivity with the

addition of TPX2. Secondly, Lys365 decreases in reactivity, which cannot be rationalised by saying that it is directly protected by TPX2 binding. The increase of Lys227 monolinks could be rationalised with the changes to the crosslinks involving Lys227 and Lys326 on Helix D. In the absence of TPX2, we see the Lys227-Lys326 crosslink with four CSMs which is not observed in the presence of TPX2. The pinning or stabilising of the activation loop could potentially reduce the occurrence of this 8.6 Å crosslink, as crosslinks shorter than the average Cα-Cα crosslinked distance are just as disfavoured as over-long crosslinks. Lys326 monolinks are also reduced with the addition of TPX2 suggesting, suggesting that TPX2 binding protects the activation loop. This shows that by looking at both changes in monolink and crosslink abundance, it is possible to unpick and rationalise conformational changes in AurA.

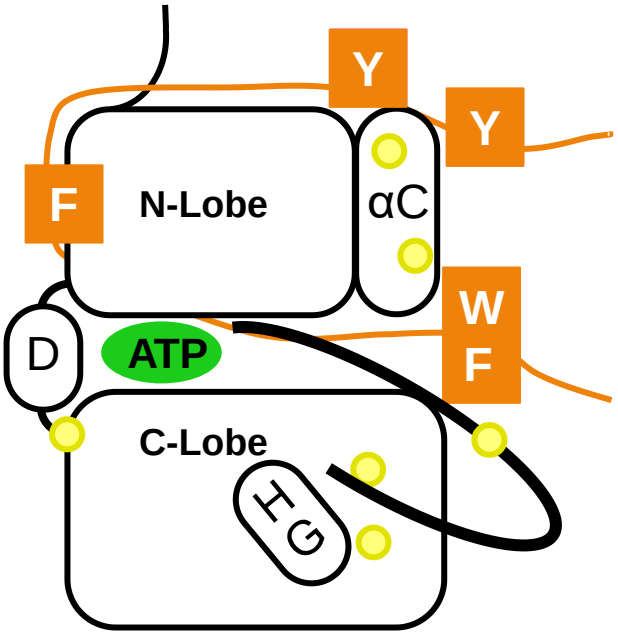


Figure 2-22 AurA+TPX2 monolink cartoon

The monolinks (yellow) with significant difference in normalised reactivity upon addition of TPX2 (orange) annotated to a cartoon of AurA+TPX2. Lys227 of Helix D increases its relative normalised reactivity with the addition of TPX2. The other five monolinked residues are reduced.

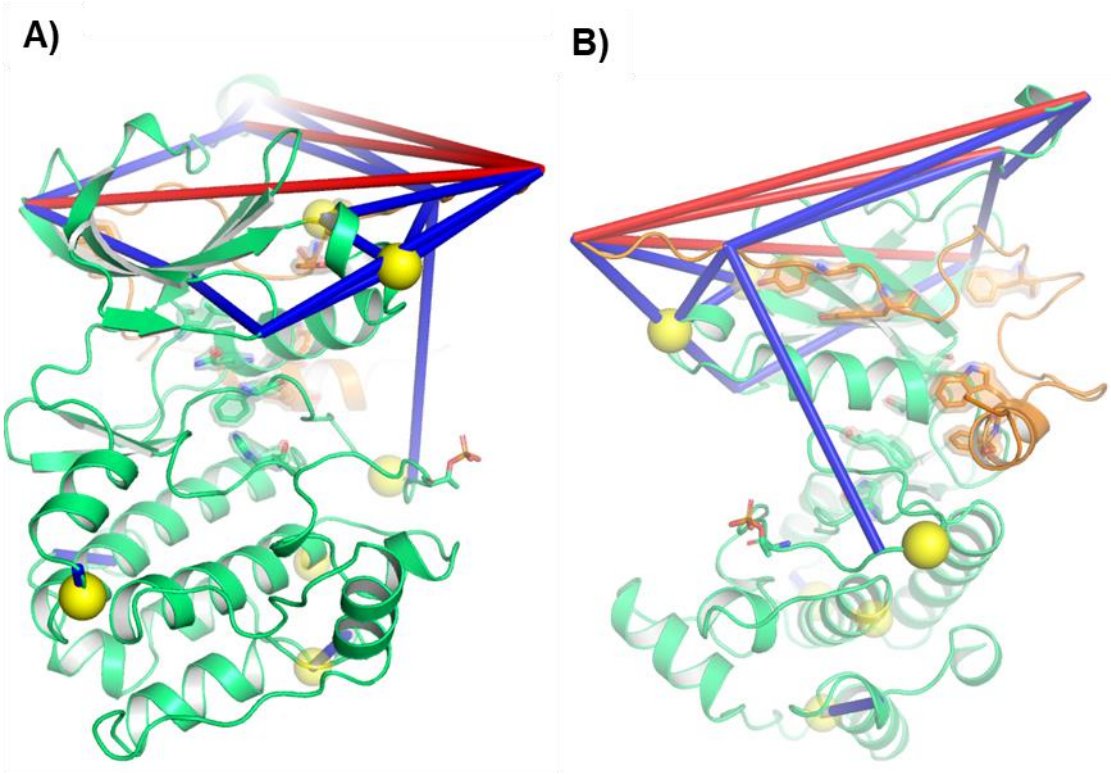


Figure 2-23 AurA+TPX2 annotated with significant monolinks

Active AurA+TPX2 model, adapted from PBD 1OL5. Blue crosslinks are under 35 Å Cα-Cα, red are over-long. Yellow spheres show the Cα carbon of monolink residues that significantly change. A) Front face of the model. Lys227 on Helix D, foreground left and Lys171 on Helix B. B) Side face of the model showing Ser283 (foreground right) on the activation loop.

2.4.3 Applying the kinase-IDR XL-MS protocol to AurA KD±N-Myc TAD

After being able to obtain tagless N-Myc, we re-collected in triplicate the AurA crosslinked with N-Myc TAD, to confirm that the data in Section 2.4.1.3.2 were not being influenced by the tag, with the aim of using this for a similar structural analysis as performed for AurA+TPX2. For this measurement, an Orbitrap Fusion Tribrid was used with a mass inclusion list.[294] In addition, the previously mentioned “sequential digestion”[192] an inclusion list was used to improve the coverage of N-Myc.[295] To prevent the measurement of oligomeric species caused by the crosslinking reaction, in-gel digestion was used.

2.4.3.1 XL-MS of AurA 119-403 and N-Myc 1-137/22-89

As we have shown with AurA+TPX2, XL-MS with DSBU using a sequential in-gel digestion is able to produce distance restraint data that matches the in-solution structure found in literature. The disordered regions of AurA, such

as the NTE, CTE and activation loop, are not excessively reactive; these regions are therefore not over-represented to the extent that they disrupt the generation of crosslinks from ordered regions of AurA. We found with crosslinking the minimum binding domain of TPX2, residues 1-43, that it also generates reasonable crosslinks. The N-terminus of TPX2, arguably the most dynamic region of AurA+TPX2, was very reactive.

Next, we turned our attention back to the N-Myc-Aurora A interaction. In the next section we showed a crosslinked AurA+N-Myc complex could be produced, albeit as part of a heterogeneous mixture of crosslinked species. The XL-MS of AurA+N-Myc 1-137 is not the smallest portion of N-Myc TAD that is still competent to bind to AurA KD. The minimal Aurora-A–interaction region (AIR) was described as N-Myc 28–89.[5] AurA+N-Myc 1-137 is hypothetically more liable to artefactual crosslinking by 'random' sampling of the portions of N-Myc outside the AIR, and so we decided to investigate the crosslinking reaction with N-Myc 28-89 as well.

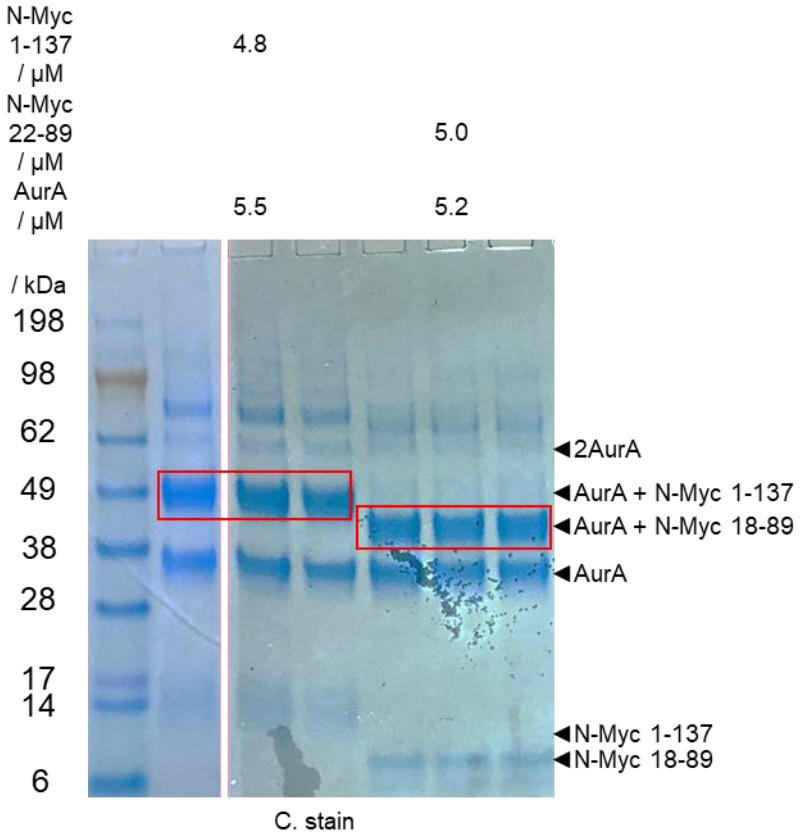


Figure 2-24 Crosslinking reactions of AurA+N-Myc used for measurement.

SDS-PAGE analysis of a crosslinking reaction of AurA+Tagless N-Myc 1-137 and AurA+Tagless N-Myc 22-89. Visualised with Coomassie brilliant blue stain. All conditions used 1.0 mM DSBU for 30 min. A) Replicates of reaction of DSBU, in ≈90-fold molar excess, to AurA KD with Tagless N-Myc 22-89, and B) incubated with tagless N-Myc 1-137 prior to the crosslinking reaction. Indicated bands (red box) excised and processed for MS measurement.

Figure 2-24 shows the SDS-PAGE analysis of crosslinking reactions between AurA+N-Myc 1-137 and AurA+N-Myc 22-89 that were later subject to MS measurement. This gel shows the trend previously seen with the

crosslinking of AurA+N-Myc, namely strong formation of 1+1 complex with the AurA monomer band still present but with minor oligomeric bands. The oligomeric bands in the presence of the tagless N-Myc show a marked increase in sharpness with respect to the previous tagged example reactions.

2.4.3.1.1 Analysis of the chemical crosslinking of AurA with tagless N-Myc 1-137

The crosslinking reaction of AurA+N-Myc 1-137 using ≈95-fold molar excess of DSBU generated an up-shifted band at ≈49 kDa, at approximately the predicted molecular weight of the 1:1 complex (46 kDa). Other bands observed on the gel corresponded to the molecular weight of dimeric AurA and an AurA+N-Myc 1-137 oligomer. In triplicate, the ≈49 kDa band was excised, processed and measured. The combined measurements were searched, which returned 336 CSMs of 43 crosslinks and monolinks that were within the tolerance of MeroX search parameters. As Table 2-11 shows, aligning and then removing crosslinks that have a single CSM removes 4.8 % of the total CSMs and 37.2 % of the crosslinks.

Table 2-11 Crosslinks found of AurA+tagless N-Myc 1-137

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	AurA				AurA-N Myc protein		NMyc			
	Intra-protein		Mono-link				Intra-protein		Mono-link	
	Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
n≥1	11	19	22	283	5	13	0	0	5	21
n>1	4	12	17	278	3	11	0	0	3	19

2.4.3.1.1.1 Results of the aligned crosslinks

It is concerning that the duplicate sample of AurA+FLAG-N-Myc 137, after removing crosslinks with a single CSM, generated 28 crosslinks and monolinks with 113 CSMs. As a result the measurement of the duplicate sample has a comparable amount of crosslinks as the triplicate sample. However, the measurement using tagless N-Myc 1-137 generated ≈5-fold more MSMs than the Flag-tagged N-Myc 1-137 sample with only a very modest increase in the number of monolinks found. We also find a convincing number of N-Myc Lys52 monolinks. It was disappointing to see no increase in inter-protein crosslinks with respect to previously collected duplicate data, and a large percentage of crosslinks with less than two CSMs.

2.4.3.1.1.2 XiNet diagram of AurA KD crosslinked to tagless N-Myc 1-137

The XiNET diagram shows a lack of crosslinks to the activation loop and crosslinks from Lys52 of N-Myc (Figure 2-25). This worse coverage of N-Myc+AurA interprotein crosslinks could be due to several factors. It is possible that the removal of the N-terminal affinity tag altered the digestion and/or measurement of crosslinked peptides. However, it is more likely that the presence of the affinity tag was inducing the formation of non-specific crosslinks, which were removed upon removal of the affinity tag. As we see a good coverage of monolinks and non-crosslinked peptides, it is more likely the latter scenario, validating our strategy to use this dataset of untagged N-Myc for structural modelling/comparison.

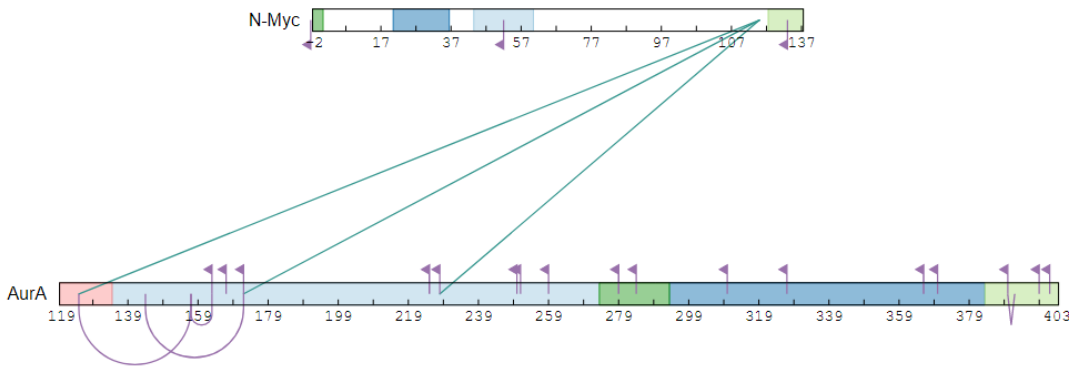


Figure 2-25 XiNet diagram found n>1 crosslinks in-gel AurA KD + N-Myc 1-137

Overview of the identified crosslinks between AurA KD 122-403 and tagless N-Myc 1-137. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. N-Myc 1-137 includes three N-terminal three cloning artefact residues. For N-Myc (L to R), domains are the vestigial residues, Myc Box 0 (N-Myc 20-36), Myc Box 1 (N-Myc 44-63), Myc Box 2 (N-Myc 128-143). For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.3.1.1.3 Normalising the AurA+N-Myc 1-137 crosslinks found

The crosslinks found were normalised, using the same process as for AurA+TPX2, shown in Appendix 5.2.2.5. Five of the seven crosslinks found with AurA+N-Myc 1-137 were found to have a normalised reactivity value of 0, as their Z_n score was 0; this was because the mean and the modal average number of CSMs were both 2. The observation of a low abundance of crosslinks for AurA+N-Myc is concerning. The inability to normalise suggests that re-collection and re-optimisation is desirable to confidently account for the lower reactivity of the system, relative to AurA Apo and

AurA+TPX2, and to confirm that it is due to the conformation of the complex and not poor recovery of crosslinks. The biological implications of this observation are discussed later.

2.4.3.1.2 Analysis of chemical crosslinking of AurA with tagless N-Myc 22-89

Hypothetically, N-Myc 1-137 is more liable to artefactual crosslinking by 'random' sampling of the portions of N-Myc outside the AIR, and so we decided to investigate the crosslinking reaction with a shorter fragment of N-Myc. The crosslinking reaction of AurA+N-Myc 22-89 using ≈95-fold molar excess of DSBU generated an extended band at ≈43 kDa. In triplicate, this was excised and processed and measured. The other gel bands observed were similar to those found for AurA+N-Myc 1-137, albeit with a reduction in oligomeric species. The combined measurements were searched, which returned 284 CSMs found with an average score of 118.72. As Table 2-12 shows, aligning then removing crosslinks with a single CSM removes 3.5 % of the total CSMs and 28.6 % of the crosslinks. Disappointingly, no N-Myc+Aurora A crosslinks were found. PEAKS analysis showed that missing peptides of Aurora A were similar to that of the AurA+N-Myc 1-137. Searching of the non-crosslinked peptides, using PEAKS, failed to find any peptides of the first seventeen residues of N-Myc 22-89, but peptides for N-Myc 36-89 were found. As the sample was measured from an excised gel band corresponding to a 1+1 crosslinked complex, it is highly likely there are missing inter-protein crosslinks, potentially from this region of N-Myc that was invisible (i.e. up to residue 36).

2.4.3.1.2.1 Table of crosslink types found for AurA+N-Myc 22-89

We observed no N-Myc-AurA crosslinks (Table 2-12). A handful of N-Myc monolinks confirm that, along with PEAKS coverage and band weight from SDS-PAGE (Figure 2-24), N-Myc was reacted with DSBU, and that there are missing crosslinks. The normalisation of the reactivity of the crosslinks can be found in the appendix, section 5.2.5.1.3.

Table 2-12 Breakdown of crosslink types found for AurA+N-Myc 22-89

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	AurA				AurA-N-Myc protein		N-Myc			
	Intra-protein		Mono-link				Intra-protein		Mono-link	
	Number	CSM	Number	CSM	Number	CSM	Number	CSM	Number	CSM
n≥1	16	32	17	248	0	0	0	0	2	3
n>1	8	24	16	247	0	0	0	0	1	2

2.4.3.1.2.2 XiNet diagram of AurA KD crosslinked to tagless N-Myc 18-89

Figure 2-26 shows an XiNet diagram of all found crosslinks from the triplicate in-gel digested tagless N-Myc 22-89 to AurA 119-403. We observed subtle changes compared to the AurA Apo crosslinks, including to activation loop monolinks.

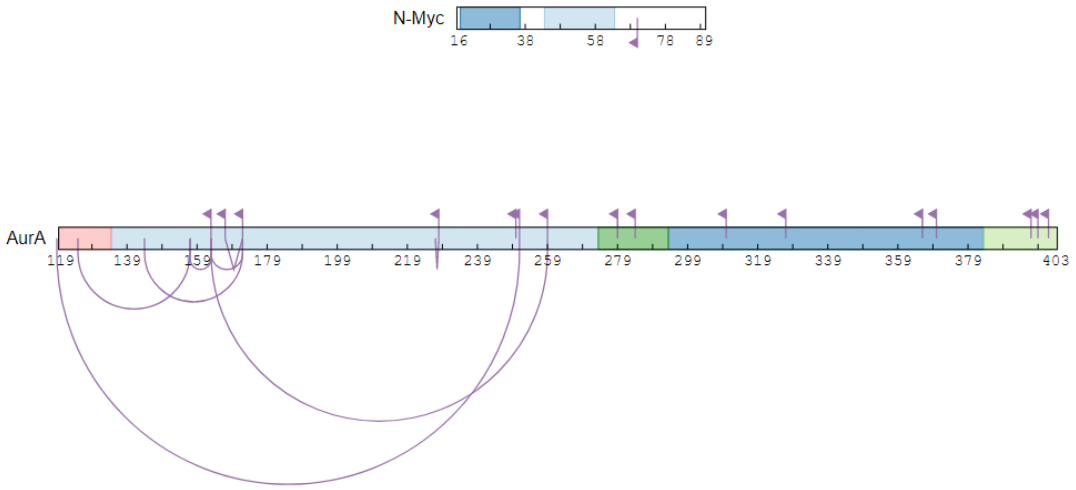


Figure 2-26 XiNet diagram of crosslinks with more than a single CSM in triplicate in-gel AurA KD + N-Myc 22-89

Overview of the crosslinks identified between AurA KD 122-403 and tagless N-Myc 22-89. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. N-Myc contains two N-terminal two cloning artefact residues. For N-Myc (L to R), domains are the vestigial tag residues, Myc Box 0 (N-Myc 20-36), Myc Box 1 (N-Myc 44-63), Myc Box 2 (N-Myc 128-143). For AurA (L to R), domains are NTE (including vestigial tag residues (119-133), N-lobe (134-272), activation loop (273-292), C-lobe (293-382) and CTE (383-403). Flags indicate the presence of monolink label, straight-sided link represents an intra-peptide crosslink, curved link an intra-protein crosslink, and red loop a homotypic crosslink.

2.4.3.2 Changes to normalised monolinks reactivity with the addition of N-Myc

2.4.3.2.1 Lack of AurA+N-Myc crosslinks precludes structural analysis

The poor recovery of N-Myc crosslinks for both constructs prevents confident analysis of the in-solution structure to the same extent as AurA+TPX2. Since N-Myc 90-137 appears to crosslink the surface in a manner that reduces and eliminates AurA intra-protein crosslinks to a worse extent than the test reactions found, and we also found no AurA+N-Myc crosslinks whilst crosslinking with N-Myc 22-89, this analysis was discontinued.

2.4.3.2.2 Changes to AurA monolinks with the addition of N-Myc 22-89

We next turned to the analysis of the monolinks for the AurA+N-Myc experiments, to examine whether the changes to monolink reactivity were able to give some insights into the interaction. We followed the same protocol used for AurA+TPX2.

Despite the lower number of AurA monolink CSMs found in the AurA+N-Myc cross-link data, the changes to normalised reactivity were reminiscent of that found with TPX2 (Figure 2-27, Figure 2-28). This was not unexpected, as N-Myc and TPX2 have been shown, by both crystal structures and biophysical assays, to compete for the same binding sites on AurA. Thus we saw the same reduction in normalised reactivity for Helix D and the activation loop, upon N-Myc 22-89 binding, and the increase in Lys227 reactivity (Figure 2-28). This data is presented in the context of a cartoon illustration of the Aurora-A/N-Myc complex in the annotated cartoon (Figure 2-27).

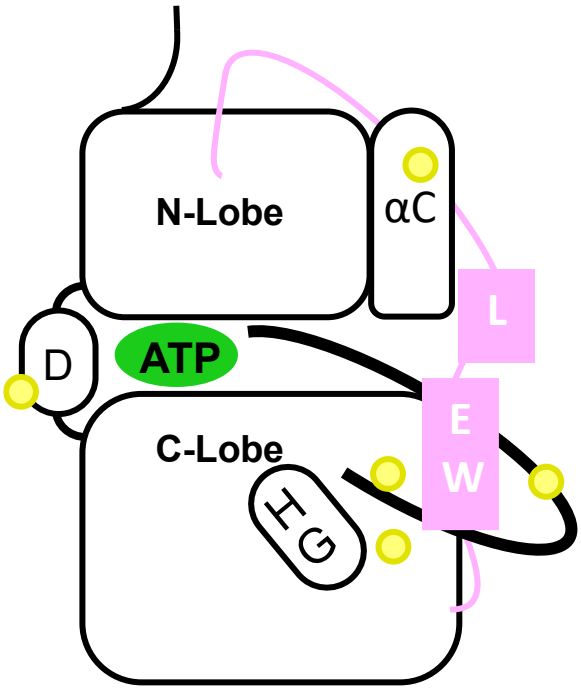


Figure 2-27 Annotation of significantly changed monolinks for AurA+N-Myc 22-89

The monolinks with significant difference in normalised reactivity upon addition of N-Myc annotated to a cartoon of AurA+N-Myc. Lys227 of Helix D increases its relative normalised reactivity with the addition of N-Myc. The other five monolinked residues are reduced.

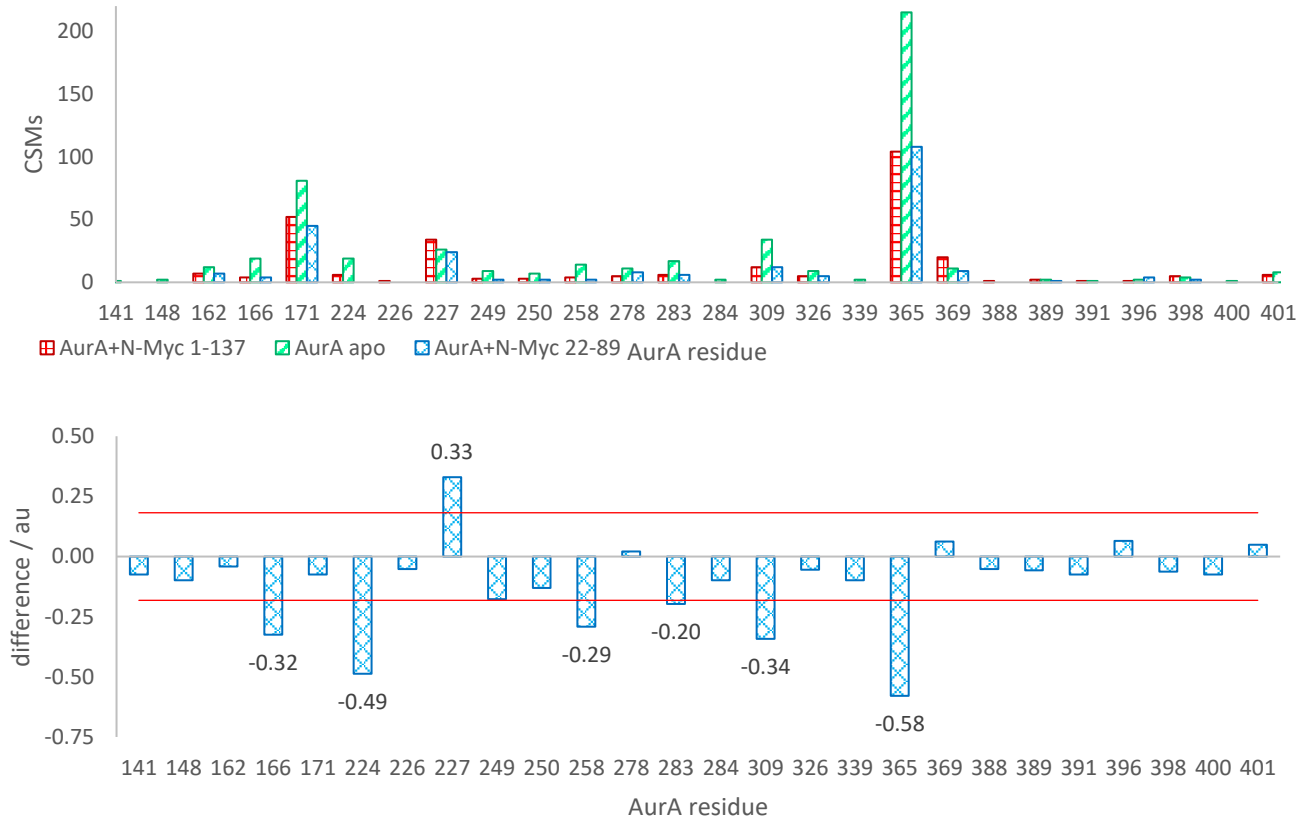


Figure 2-28 Changes to the normalised crosslinks for AurA±N-Myc 22-89

Changes in crosslink abundance for each unique monolinked residue site of AurA upon binding of N-Myc 22-89 were used to explore the changes in monolink reactivity as a post-hoc foot printing experiment. Upper – AurA monolinik residue Z-score change with the addition of an interactor. Total MSM for each residue shown for all three conditions, AurA with N-Myc 22-89 (blue), N-Myc 1-137 (red) or Apo (green). Lower – Changes to normalised monolinks with the addition of N-Myc 22-89. The difference in normalised reactivity between AurA in the presence and absence of N-Myc 22-89 is shown for each residue. The increase in reactivity is shown by a positive number; with the addition of N-Myc 22-89 and subsequent decrease in reactivity shown by a negative number. Residues where the change in normalised reactivity is greater than one standard deviation of the dataset (shown as the red lines) are highlighted. Each condition (±N-Myc 22-89) is normalised so that changes in reactivity of the residue relative to the reactivity of the entire condition can be examined.

2.4.3.2.3 Changes to AurA monolinks with the addition of N-Myc 1-137

Despite N-Myc 90-137 having been shown to be extraneous to the interaction of AurA+N-Myc, inclusion of this region alters the monolink reactivity at AurA sites to give a different pattern than what was observed with N-Myc 22-89 (compare Figure 2-28 and Figure 2-29).

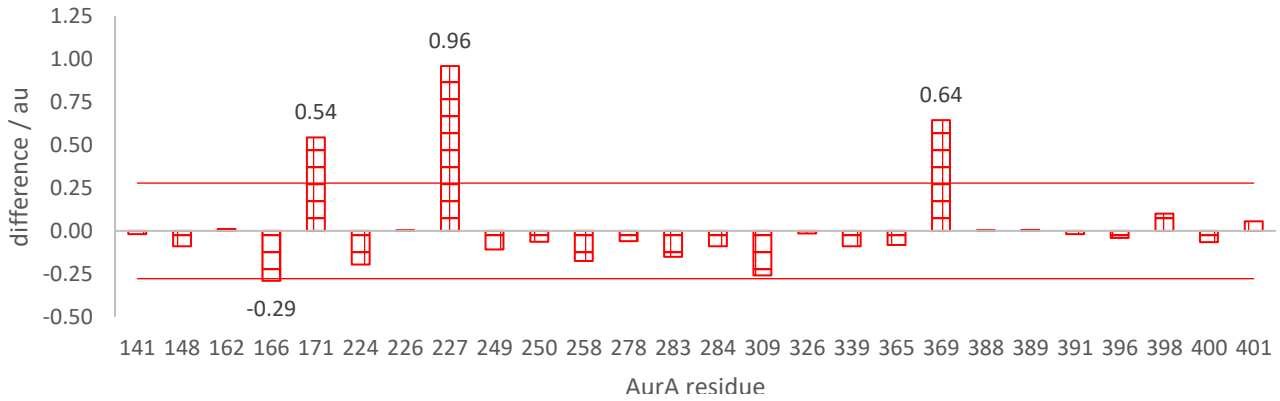


Figure 2-29 Changes to AurA monolinks with the addition of N-Myc 1-137

Changes in crosslink abundance for each unique monolinked residue site of AurA upon N-Myc binding. The difference in normalised reactivity between AurA in the presence and absence of N-Myc 1-137 is shown for each residue. The increase in reactivity is shown by a positive number; with the addition of N-Myc 1-137 and subsequent decrease in reactivity shown by a negative number. Residues where the change in normalised reactivity is greater than 1 standard deviation of the dataset (shown as the red lines) are highlighted. Each condition (±N-Myc 1-137) is normalised so that changes in reactivity of a residue relative to the reactivity of the entire condition can be examined.

2.4.4 XL-MS of Apo N-Myc

As a control, apo N-Myc 1-137 was measured in triplicate to assess its reactivity towards DSBU. A single stage Trypsin in-solution digestion protocol was used to eliminate the possibility that poor extraction of crosslinked peptides from the SDS-PAGE gel was the cause of the potential "missing" AurA-N-Myc crosslink peptides. We wished to examine if the peptides of N-Myc crosslinks were uniquely incompatible with ESI MS/MS, relative to those of AurA, to account for the poor recovery of N-Myc-AurA crosslinks. A Western blot of a N-Myc 1-137 crosslinked sample showed formation of a potential dimeric species (Figure 2-30).

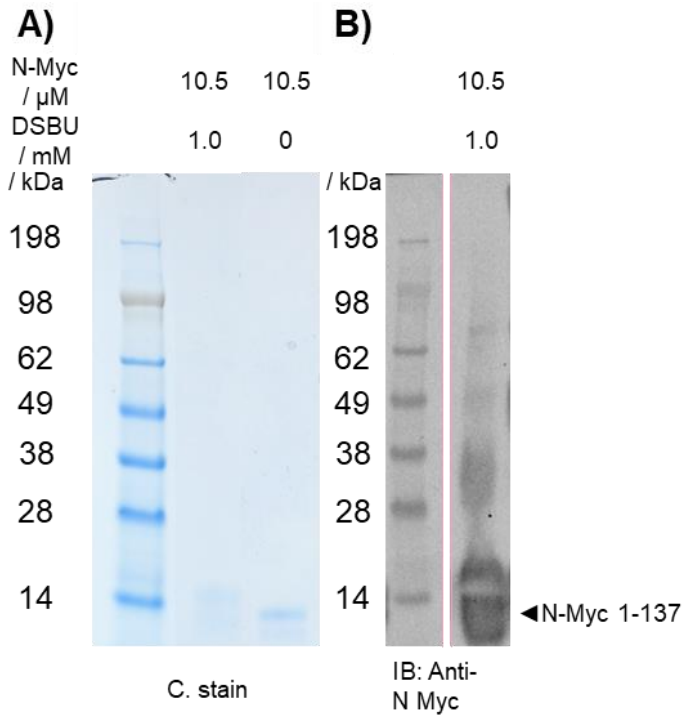


Figure 2-30 SDS-PAGE gel of test reaction of tagless N-Myc 1-137

SDS-PAGE analysis of crosslinking reaction of Tagless N-Myc 1-137 at 10.5 µM, visualised with A) Coomassie brilliant blue stain and B) via Western blotting with an Anti-Myc antibody.

2.4.4.1 Analysis of crosslinked peptides of N-Myc 1-137 by MeroX

In triplicate, the reaction mixture was digested in-solution, and this was then processed and measured. After combining the triplicates for analysis, searching with MeroX returned 29 crosslinks found with 192 CSMs. As shown in Table 2-12, aligning then removing crosslinks with a single CSM removes 3.6 % of the total CSMs found and 27.6 % of the crosslinks. Crosslinks are provided in appendix 5.1.1.8.

The crosslinks with the two largest number of CSMs were the homotypic crosslinks Lys125-Lys125 and Lys133-Lys133. In keeping with the trend, 48.7 % of the monolinks were amide modified. Searching the non-crosslinked peptides using PEAKS found good coverage.

2.4.4.2 XiNet diagram of tagless N-Myc 1-137 Apo

The annotation of N-Myc 1-137 crosslinks on to the XiNet diagram (Figure 2-31), shows that both MBI and MBII are loci of very high crosslinker reactivity, to the extent that MBII could potentially be forming a weak dimer. We would expect non-specific oligomerization to present as non-specific crosslinks down the length of N-Myc. However, this is not the case, as Lys52 and Lys125 have near equal MSMs and we don't observe inter-box crosslinks (Table 2-13).

Table 2-13 N-Myc 1-137 crosslinks after alignment

A breakdown of the types of crosslinks and the respective number of CSMs found after alignment and after the removal of crosslinks with a single CSM.

Cut-offs	N-Myc			
	Intra-protein		Mono-link	
	Number	CSM	Number	CSM
n≥1	13	40	16	152
n>1	8	36	13	149

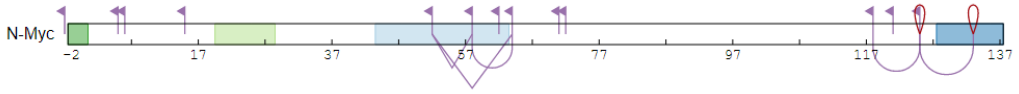


Figure 2-31 XiNet diagram of tagless N-Myc 1-137 Apo in-solution

Overview of the identified crosslinks of tagless N-Myc 1-137. Figure generated using XiNet.[287] Residues are numbered according to their position in the protein. For N-Myc (L to R), domains are the vestigial N-terminal affinity tag residues, Myc Box 0 (N-Myc 20-36), Myc Box 1 (N-Myc 44-63), Myc Box 2 (N-Myc 128-143). Flags indicate presence of monolink label, straight sided link represents an intra-peptide crosslink, curved link is an intra-protein crosslink and red loop a homotypic crosslink.

2.4.4.3 Implications for AurA+N-Myc

N-Myc 1-137 apo presenting as a low affinity dimer perhaps explains previous observations from other Bayliss lab members that the stoichiometry of N-Myc+AurA complex is difficult to quantify by calorimetric methods, and this is dependent on the N-Myc construct used (data not shown). This is also consistent with my observation of the AurA+N-Myc 1-137 generating complex higher stochiometric crosslinked AurA+N-Myc complexes. Comparing the SDS-PAGE analysis of AurA+N-Myc 1-137 and AurA Apo crosslinked with DSBU, we would hypothesise that N-Myc 90-137 is aiding the formation of short-lived or weak dimers of the AurA+N-Myc complex and/or bridging the formation of a trimer comprised of two N-Myc and one AurA. We hypothesise that the crosslinking reaction of AurA+N-Myc 22-89 is generating fewer species (see gel, Figure 2-24) as it is unable to form N-Myc mediated oligomeric species. In conclusion, cross-linking studies on the AurA/N-Myc interaction suggest that the interaction may be more complex

than previously thought (Figure 2-31), but it does not seem feasible to take this approach to resolving its precise nature.

2.5 Discussion

2.5.1 Introduction

The catalytic activity of AurA, like other oncogenic kinases,[296] is regulated through the phosphorylation of its activation loop and control of its conformational dynamics. Insight into these molecular mechanisms is critical in understanding how protein-protein interactions affect changes in kinase activity, and how the causes and effects of mis-regulation drive disease. This knowledge is crucial in devising novel approaches to treatment, such as protein-protein interaction inhibitors that block cancer-specific functions of AurA whilst leaving normal functions intact. The development of clinical inhibitors of AurA has rested on the production of crystal structures of the kinase+inhibitor complex and their validation as part of the drug-discovery pipeline. We wanted to explore the utility of XL-MS to support the structural investigation of kinase conformation. Kinases with IDRs, or that interact with IDPs, are a high value target type which resist crystallographic structural methods. As such, new methods are needed to explore these targets. XL-MS is primarily used for stable complexes but we hypothesise that it is also appropriate for this target type.

The Bayliss lab, having produced many AurA crystal structures, had experience with biochemical orthogonal validation of PPIs of AurA with its interactors. We used AurA±TPX2 to test the feasibility, technical constraints and data generated using the mass-cleavable NHS ester DSBU. Our two-stage approach to developing an XL-MS protocol suitable for IDR containing kinases that bind to IDP targets found that controlling and tempering the reactivity was essential to generating useful structural data. Quantification of the structural changes caused by the binding of the IDP interaction was performed by normalisation, adapted from the protocol outlined by Hogan, of the reactivity of the monolink labels, which allowed for the interpretation of the crosslinks found. Herein, we discuss technical questions before examining the biological implications of our investigations into AurA±TPX2 and AurA+N-Myc, and then finally we discuss improvements with reference to quantitative qXL-MS.

2.5.2 Technical discussion

When planning these XL-MS experiments, we hypothesised that due to the disordered nature of the proteins being crosslinked, we would have to prevent the formation of rare or transient crosslinked species. Our initial conceptualisation of the reactivity of the NHS ester crosslinker DSBU was shown to be incorrect with the exploration of the crosslinking of AurA+N-Myc/TPX2. The adapted protocol required extensive optimisation to reduce the generation of smeared crosslinked species, observed by SDS-PAGE. This was understood to be caused by over-crosslinking of our target protein species, which would potentially have produced XL-MS compromised by non-specific cross-links. The measurement of the optimised reactions generated crosslinks, the majority of which were in tolerance for the Cα-Cα distance of residues linked by DSBU. Additionally, the histogram of crosslink distances generated curves that fitted the expected shape.[257] This suggested that the IDR-containing kinases are closer to standard complexes as targets for XL-MS than expected, after the reduction in reactivity is taken into account, as the distribution of distances is within the expected range for the crosslinker used.

2.5.2.1 Reactivity Normalisation

Our attempt to normalise the reactivity of the found crosslinks, following Hogan, was done in order to identify crosslinks that are critical to the conformation of our complex. This was hampered by poor coverage of both cross- and mono-links caused by several different issues. To calibrate all the residues' reactivity requires monolink coverage, which was lacking in regions of AurA for which we needed valuable structural information.

The reactivity of residue side chains is currently described by a 3+1 paradigm, where the reactivity of residues with respect to NHS esters is governed by the electrostatic energy, the number of hydrogen bonds, and the estimated pKa values of side chains, in addition to their solvent accessibility. We assumed that the more dynamic the protein was, the larger the range in reactivities would be. The residues of a globular protein with a single in-solution conformation would not see any dynamic variation in solvent accessibility, thus the local conditions of an NHS ester reactive side chain would be governed by local effects such as the hydrogen bond network, as well as the charge and hydrophobicity of local side chains. However, the local conditions of a residue sidechain of a protein with IDRs will vary with the equilibrium change in conformation, in addition to the IDRs solvation. Hence, we accounted for the larger than expected number of

residues lacking monolinks (six), especially in the NTE of AurA, due to those residues being so reactive that they always formed crosslinks. However, it is understood that hydrolysis, and thus the generation of monolinks, will always dominate over crosslinking. This has been assumed because, although less reactive, as the solvent of the reaction, the excess water will always out-compete the more reactive residue side chains.

However, our data shows that amide monolinks that are generated at the end of the reaction make up at most 50 % of the monolinks found, which is expected since rates of hydrolysis of NHS esters is >3 hours at 4 ^{o}C.[297] Thus, intrapeptide crosslinks are the main competitor to inter-domain or interprotein crosslinks, and monolinks are generated typically in the absence of available crosslinkable sites. Local effects that appear to catalyse the labelling reaction do still apply, and particularly solvated or residue chains held near catalytic residues will be reactive to monolinks over crosslinks. This is not reflected in our monolink CSMs. It is striking that Lys365 produced the most monolinks; often up to a third of all observed MSMs are from Lys365, followed by residues Lys171, and then Lys227 or Lys309,[298] as the next most reactive. The reactivity of Lys365 is rationalised as, when inspecting crystal structures PBD codes: 4CEG and 1OL7, it is held between a pair of Arginine side chains which is known to catalyse the NHS-ester reaction.[202]

Our other main hypothesis to account for the lack of monolinks was the presence of previously unseen PTM, incorrect sequence or some other reason for the disruption of the searching for using MeroX. MeroX also lacks the ability to search for double labelled peptides. These potentially remained through optimising multiple digestion methods. Guo *et al* suggested that reactive residues screen less reactive residues from crosslinks, and thus reduce the availability of structure-determining crosslinks; this is further demonstrated in our data sets.[289] For the triplicate measured AurA Apo dataset, only five of the sixteen crosslink-able residues of AurA 119-217 were found with monolinks. Guo *et al* pre-treat their proteins with an NHS ester labelling reagent prior to crosslinking to prevent the over-reactive residues from distorting the data.[289] Inadvertently, we found that low concentrations of Tris that is diluted out from proteins prior to crosslinking may increase the rate of monolink generation. Another strategy could have been to use a shorter crosslinker length: Hofman *et al* have shown that for proteins under 50 kDa, crosslinkers with a ≈11 Å spacer-length are better

suited for structural crosslinking.[290] DSAU, not commercially available during this work, has a ≈11 Å spacer-length, whereas DSBU is ≈12.5 Å.[299]

Despite using a range of ≈60-fold molar excess of DSBU from our initial measured AurA+N-Myc sample to the final sample, the reactivity of the regions did not noticeably change, in line with previous data.[193] This is in-line with recent findings that shorter reaction times generate more informative crosslinks, with some finding that 10 minutes in ≈80 fold molar excess of crosslinker yields strong data.[300] This tallies with our data suggesting many un-reacted dead end crosslinks persist until the quench of the reaction (which then react either with added tris or ammonium bicarbonate).

Normalisation of the monolink reactivity allowed us to use the unused monolink data to perform post-hoc foot printing. We assumed that the relative reactivity of a residue was proportional to its MSM, as this has been shown with NHS ester covalent labelling regents.[199] We initially assumed that we would need a method of differentiating crosslinks, but we observed that crosslinks annotated onto AurA-Apo and AurA+TPX2 were for the most part within the tolerance of distances for DSBU.

2.5.2.2 The value of Euclidean Cα-Cα distances in kinase-IDR model discrimination

We chose to annotate our model using Euclidean distance (sometimes called Straight-Line Distances, SLD) between Cα residues, using a PyMOL plugin, as it was a simple and computationally cheap process. However, we found that, using this method, the majority of our crosslinks were within distance, even though some of them clearly passed through solvent in-accessible regions. It became clear that the computationally expensive SASD (solvent accessible surface distance) would better discrimination between crosslinks. This is highlighted by the AurA NTE conformation differences that we observed for AurA Apo. Euclidean Cα-Cα distances can *qualitatively* discriminate between a solvated AurA NTE model and an NTE bound AurA model because of the quantity of highly reactive normalised crosslinks, however, it does not possess the resolution to *quantitatively* discriminate between the two potential NTE models. Bullock *et al* published a novel scoring webserver that uses SASD to evaluate various models.[301] Crosslink evaluation and scoring software that accommodates dynamics in crosslinking has been developed by Sinnott *et al*.[302]

Additionally, quantifying if these models were stable representatives or naïve abstractions, by molecular dynamics simulations or similar, would give

further resolution to discriminate between prospective conformations. This appears to be especially useful in the case of AurA+TPX2. The TPX2 N-terminus appears to be dynamic but it would be useful to identify whether it was in fact occupying two opposed, stable conformations, or simply a range of disordered conformations because the N-terminus lacks a stable binding pose.

2.5.2.3 Kinetic trapping

After we had ascertained that AurA+N-Myc did produce 1+1 crosslinked species, the second area of concern was that the dynamic portions of the complex would kinetically trap or generate artificial or unrepresentative structures. This is a possibility for N-Myc crosslinking data; the consequences of this will be discussed below. The majority of crosslinks found for each crosslinked species were within tolerance of the models generated. This vindicated the optimisation time spent to reduce non-specific crosslinks.

It is hypothesised that complexes crosslinked with NHS ester crosslinkers are prone to being “zippered” or “kinetically trapped”: rare or artificial protein conformations or stoichiometries.[303] This hypothesis states that once one end of the crosslinker has reacted with a protein side chain, given the long half-life of the NHS ester with respect to hydrolysis, it will be able to form a crosslink with another residue, and this will be artificially longer than expected, as the inherent dynamics of a protein means that the two residues will react if they are even briefly in-range: a kinetic trap. Therefore, this minor conformer is magnified as it garners more CSMs of a crosslink than would be expected. In addition, mono-links could potentially induce conformational changes and so expose previously un-reactive residues to become reactive. They hold open the door to partially buried residues, as it were.

Whilst the analysis of monolinks bears out that NHS ester crosslinks do have a long lifetime, the shorter reaction time used in our study appears to have prevented this. Older protocols suggest reactions should be performed at room temperature over 60 minutes; we found that this generated smeared bands, and hence automatically discounted their use in structural studies.

However, the charge that NHS ester crosslinkers are unsuitable for studying PPIs, especially for disordered proteins, is shown to be unfounded, given the evidence of this thesis. The kinetic trap theory recommends that, to prevent kinetic trapping, short- or ultra-short-lived crosslinkers should be used to provide a snapshot of the PPI and the equilibrium of interactors. We have

however found no evidence of the minor conformation of AurA Apo being over represented since, as discussed later, the dimer species appears relevant to the in-solution equilibrium. Rozbeský *et al* found that the crosslinking of human carbonic anhydrase with BS3, a 35 Å Cα-Cα tolerance NHS ester crosslinker, at x5 and x10 molar excess retained ≈80 % of its enzymatic activity.[193] This same protocol was shown by native MS to generate on average two to four cross-links and monolinks per protein. The use of SDS-PAGE gels to evaluate the species' crosslinks, although experimentally simpler, would have been more insightful for our complex crosslink species.

The kinetic trapping theory also posits that NHS esters may disrupt the native conformational ensemble of a protein. Initially, we were disheartened with the generation of a dimeric AurA band in potentially a 1-to-3 or lower dimer to monomer ratio, the biological implications of which are discussed later. Numerous groups using a plethora of techniques have shown that AurA Apo is in an equilibrium with a MSP1-like inactive conformer (Section 1.3.3.2). Examining the crystal structure model suggests that the generation of activation loop to Helix D crosslinks due to the reactivity of Lys227 would be likely. This is not seen in our data. It is vexing that we could not find direct validation of the MSP1-like inactive conformer using non-labelled AurA protein.

As previously shown by Mädler *et al*, equilibrium positions of dynamic complexes are maintained when using optimised reaction conditions.[304] One mistake we seem to have made was to not generate single species by altering the ratio of interactor to AurA. Given that the K_d of TPX2 with AurA is ≈1 µM, we would expect a 1:1 AurA Apo to AurA+TPX2 complex, which is seen when examining the SDS-PAGE gels. We were wary that the enrichment of N-Myc in the presence of AurA, (i.e. not using a equimolar ratio for the crosslinking reaction) would generate multiple species, due to the disordered structure of N-Myc. Although our strategy was vindicated, post-hoc, due to the N-Myc Apo dimerization found, we may have achieved greater coverage of crosslinks if we had optimised for the generation of a homogenous AurA+N-Myc complex that could be digested in-solution for XL-MS. Unfortunately, the AurA+N-Myc interprotein crosslinks were still missing. Despite the changes in digestion conditions, it is on balance likely that this is resulting from poorly ionising peptides or similar mass spectrometric effects. Hence, we cannot confidently describe the conformation of these regions of

N-Myc. If the problem is not detection of crosslinked peptides, then it is likely that the interaction of N-Myc with AurA is preventing the reaction with DSBU.

We were concerned that the presence of N-Myc was inducing artificial stoichiometries with AurA, but the XL-MS of N-Myc in isolation showed a weak dimer forming.[305] Wei *et al* have shown that XL-MS of the IDP p53 respects the correct oligomeric state dependent on the construct and additives used in crosslinking, and so we presume the same for N-Myc.[283]

2.5.3 Discussion of biologically relevant results

2.5.3.1 Examination of the conformation ensemble of AurA

It has been established using FRET, IR and EPR probe-labelled AurA in the presence and absence of activation loop phosphorylation, kinase inhibitors and activating interactors such as TPX2, that AurA is found in-solution in an equilibrium of conformers (Section 1.3.3.3).[171] These are approximated as a canonically active structure (PBD 1OL7, 4CEG, 5DT3) and a DFG-inter/up inactive structure that is trapped by MLN8054 (PBD 2WTV). No MSP1-like structure of AurA has been presented that is not complexed with an inhibitor. Despite extensive searching of the PDB, the only kinase that presents its inactive, Apo form with the activation loop bound to Helix D is the human mitotic checkpoint kinase MSP1 (PBD codes 3HMN, 2ZMC).[306]

Having validated that the crosslinks generated by our protocol were reasonable, we found that AurA Apo with the NTE bound to the W- or F-pockets were best at rationalising the crosslinks found. After these changes to our model, we found all crosslinks were in-tolerance for DSBU crosslink lengths. As we did not obtain sufficient crosslinks involving the activation loop, we could not disprove the presence of conformations of the activation loop other than the canonical active conformer. However, we observed crosslinks that would span the cavity required for the MLN-like pose, Lys162-Lys258, and the highly reactive Helix D, which would be covered in the MLN-like pose, generated no crosslinks to the activation loop. In the triplicate measure of AurA Apo, the only crosslinks observed of the activation loop were monolinks of Ser278, with eleven MSMs. In the duplicate measurement, the AurA band of the AurA+FLAG N-Myc 1-137 was found to have the crosslinks Lys143-Ser284 and Lys171-Ser284, each with two CSMs.

Thus, we question the relevance of the MSP1-like inactive form to the AurA Apo equilibrium, given that an Active-like autoinhibited conformer has also

been described, that involves the rotation of the activation loop, and better fits the crosslinking data.[307] This would bring AurA activation loop dynamics in line with other kinases. However further experimentation is needed to positively discount the presence of a stable MSP1-like conformation.

In addition, as the reactivity of Lys227 increases with the addition of TPX2, in conjunction with the removal of the crosslink Lys227-Lys326, this suggests a change in the conformation of the activation loop. This could be interpreted as rotation and fluctuation of the activation loop occurring and not the formal presence of the MSP1-like structure in the equilibrium of AurA+TPX2.

Two key questions still remain, one technical and one regarding AurA. Firstly, we are unsure how DSBU mass cleavable crosslinks interact with the phosphorylation on the peptide. The activation loop is double phosphorylated, as seen by non-crosslink peptide search, and seen in the work of others,[308] and the MeroX search algorithm cannot search double labelled peptides. In hindsight, pre-measurement treatment of the peptides with calf intestinal alkaline phosphatase would have removed the source of doubt these PTMs cause.[309] Secondly, could treatment of the MLN-like inducing AurA inhibitors prior to crosslinking trap out this conformation to positively show presence of the other possible conformation?

2.5.3.2 The NTE of AurA binds into the F-pocket and competes with interactors

Fu *et al* found that the G198N AurA mutant behaves *in vivo* as an AurB-like kinase and McIntyre *et al* showed that G198N AurA has a reduced TPX2 binding affinity.[86 149] The crystal structures of AurA and TPX2 plainly show the NTE of AurA interacting TPX2, with TPX2 bound into the F-pocket. This displacement can be seen in the change to crosslinks found with the addition of TPX2, which removes crosslinks of the AurA NTE to Lys250, and with the observable crosslinks to the NTE of TPX2. Alignment of the B4 and B5 loops of nine AGC kinases, including AurB, shows that G198 is an outlier in the AGC family (Figure 2-33). Glycine, which lacks a side chain, allows much more flexibility than any other residue.[310]

Disruption of the cap pocket has consequences for its status as a druggable hotspot. McIntyre *et al* performed a fragment screen of AurA binders, and the strongest inhibitors were F-pocket binders (Figure 2-34 panel A).[86]

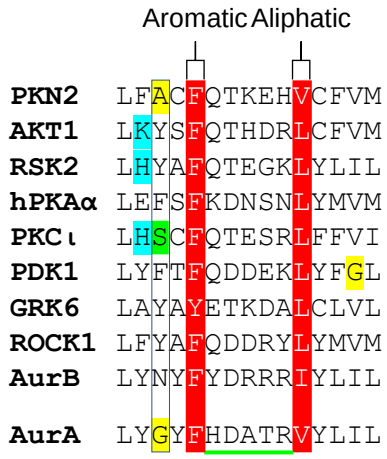


Figure 2-32 Sequence aligment of NTE of AurA with select kinases

Sequence alignment of the Beta-4 and Beta-5 loops of AurA and nine AGC kinases with the position G198 of AurA highlighted. Upper, the highly conserved aromatic and aliphatic residues used to substantiate the alignment. Lower, green bar, the less well conserved charged region. The alignment was manually curated based on X-ray structures of each kinase [PDB IDs: 4CRS (PKN2), 4EKK (AKT1), 4NW6 (RSK2), 4WB5 (hPKA), 4DC2 (PKCι), 4RQK (PDK1), 3NYN (GRK6), 4W7P (ROCK1), 4AF3 (AurB), 1OL5 (AurA)].

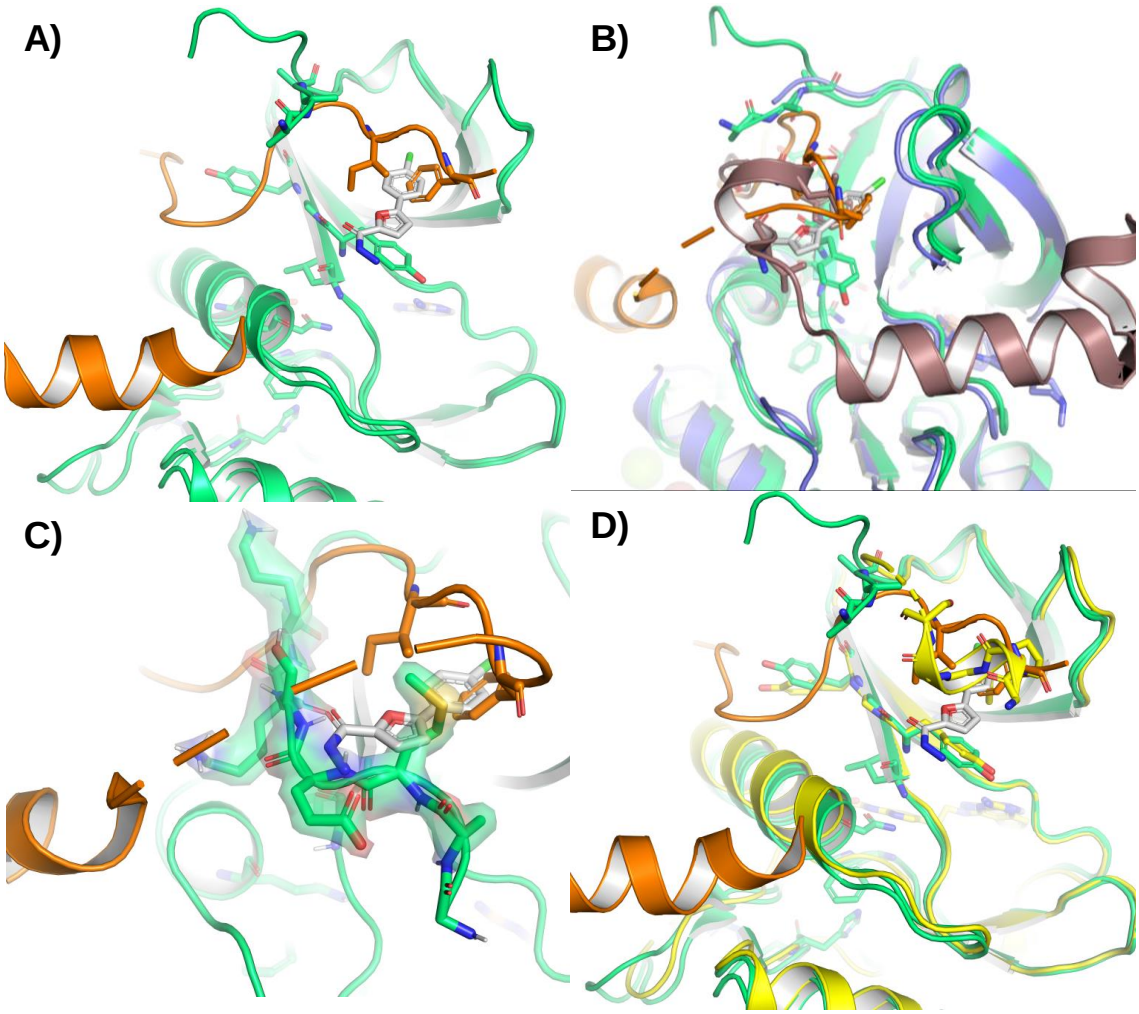


Figure 2-33 Interactions of with the F-Pocket of AurA

Panel A) As part of a high-throughput X-ray crystallography-based screen McIntyre *et al* generated AurA in complex with an F-pocket drug fragment in silver (PBD code: 5OSD) aligned with AurA in complex with TPX2 in orange (PBD code 1OL5) with residues of TPX2 shown bound in the F-pocket. Panel B) shows the same structures as panel A with the addition of AurB in blue in complex with INCENP in brown (PBD code 4AF3). Panel C) shows AurA, TPX2 and the F-pocket drug fragment aligned with and the NTE from the NewCap AurA model. Panel D) shows AurA, TPX2 and the F-pocket drug fragment aligned with and the NTE from the mAurA in yellow (PBD code 3D14).

It is complex to understand the *in vivo* relevance of the F-pocket/NTE interaction when the XL-MS datasets and crystallographic structures were collected using crystallographic constructs that lack the full NTE of AurA. Despite the different residues found in each of the constructs they all place density that overlaps with the small molecule inhibitor of found in PBD structure 5OSD (Figure 2-35 shows the alignment the constructs used in Figure 2-34). The F-Pocket is accessible in other AGC kinases and AurB but the lack of a structured NTE that directly binds into the F-pocket and Gly is unique to AurA.

The only other AurA crystallographic structure, apart from 7O2V, that presents with a structured NTE is mouse AurA (PBD 3D14) that structures the NTE as a short helical turn in the F-Pocket, (Figure 2-34, panel D).

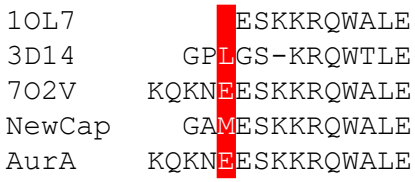


Figure 2-34 Sequences of the NTEs of selected constructs of AurA

The sequence of the NTEs of selected constructs of AurA crystallographic structures aligned with AurA 117-131. The red bar highlights the residues place in the F-pocket. The alignment was manually curated based on X-ray structures of each kinase [PDB IDs: 1OL7 (Apo hAurA), 3D14 (mAurA), 7O2V (hAurA), NewCap (modelled hAurA)].

Given that our crosslinking data places the shortened NTE of AurA into the F-pocket, this may explain why inhibitors that potently inhibit the AurA-TPX2 interaction, poorly bind apo AurA (as this binding site is occluded by NTE binding).[86] It has been shown that a mutation in the NTE (K119G) impairs autophosphorylation at Thr288 that is required for AurA activation, which suggests a role of the NTE in controlling activation of AurA, potentially by binding the F-pocket.[95] The relevance on the NTE of AurA binding into the F-pocket warrants further study.

2.5.3.3 AurA Apo stoichiometry

We have also shown that the in-solution AurA dimer is potentially important to the conformational ensemble of AurA Apo, and is modulated by the addition of TPX2 and N-Myc. Initially, we were surprised to observe that crosslinked AurA dimerized species were present in both, strongly in Apo and weakly in AurA+TPX2. It was previously found by Zorba *et al* that the Cys290 disulfide bridge formation was meditating the generation of a dimer, and they suggested the dimer crystal structures are potentially a step in the AurA activation mechanism.[145] The native mass spectrometric studies of AurA Apo by Tomlinson did not present evidence of the formation of a dimer.[167] Nevertheless, Zorba *et al* demonstrated that this dimer was key to redox dependent activation of AurA. Having performed our crosslinking with an AurA KD construct designed to remove any of the cysteine-mediated oligomerisation in question, the oligomerisation found in our crosslinking data is perplexing. As the AurA+AurA dimer is completely absent from the AurA+TPX2 crosslinking and the crosslinking reactions were performed at 10 μM, below the 25 μM Zorba *et al* performed their kinase assays at, and XL-MS respects the equilibrium positions of conformation ensembles, it is

likely that this phenomenon is not artificial. As the crystallographic constructs lack the complete NTE of AurA further study is needed to understand the place of this transient dimer in the activation and catalytic turnover of AurA.

2.5.3.4 N-Myc binding

Unfortunately, the crosslinking of AurA+N-Myc did not conclude with the presentation of a more extensive model of the interaction that includes MB0. This is potentially due to crosslinking AurA+N-Myc at an equimolar concentration and not saturating N-Myc to force the complete formation of AurA+N-Myc complex. A crosslinking reaction that was subsequently performed with a threefold molar excess N-Myc 1-137 with AurA exhibited the formation of a 1+1 complex with minimal oligomeric bands (data not shown). The low abundance and clearly missing intra- and interprotein N-Myc crosslinks prevent a more impactful analysis of the in-solution structure. Available evidence suggests that N-Myc wraps equatorially around AurA (unpublished observations). This potentially activates AurA, but primarily prevents the modification of N-Myc for proteasomal degradation a mechanism targeted for anti-cancer therapies.[5] However, a complete crystallographic picture of the complex is yet to be presented and our crosslinking has provided new insights into the interaction. Primarily, it has shown that the stoichiometry of the interaction is dependent on the presence of MBII and needs re-evaluation with other in-solution techniques.

Both constructs of N-Myc crosslinked to DSBU show a total CSM count 25-50 % lower per measurement than the previous conditions. The pattern of AurA intra-protein crosslinks and monolinks of AurA+N-Myc appears to resemble that of the AurA+N-Myc 61-89 crystal structure, with changes to the positioning of AurA NTE (Figure 2-9). Apart from the AurA N-term119-Lys250 crosslink, the crosslinks we observed were the same as the crosslinks found in the AurA Apo condition, albeit with a reduced number of crosslinks and CSMs.

Furthermore, the changes to the monolink normalised reactivity of AurA in the presence of N-Myc 22-89 is strikingly similar to that of TPX2. As the reactivity of N-Myc Lys52 is significantly reduced in the presence of AurA, and the changes to monolink reactivity resemble the changes seen with TPX2, in addition to the AurA intra-protein crosslinks being within tolerance of the new cap model, we speculate that N-Myc MBI is binding across the Y-pocket but critically not disturbing the AurA F-pocket and AurA NTE interaction. Interestingly, we do not see discrete AurA dimeric bands in the presence of N-Myc 22-89, but this is seen with N-Myc 1-137, consistent with

a bridging and/or dimerization role of the 90-137aa region of N-Myc in generating higher-order complexes.

Peptide competition assays have shown that N-Myc peptides 19-38 and 28-60 are out-competed by TPX2 for binding with AurA, but the fact that N-Myc 22-89 does not alter the position of the AurA NTE suggests that N-Myc 22-60 is binding over the Y-pocket, as N-Myc 61-89 is shown by crystal structure to bind over the activation loop and W-pocket. The crystallographic structure of AurA when in complex with N-Myc is the same as the AurA Apo structure; hence our observation of the same crosslinks is expected.

Taking the Flag-tagged and tagless N-Myc 1-137, we can tentatively suggest that the presence of the N-Myc 90-137, a region shown convincingly by crosslink analysis to be extremely reactive, is extraneous to the N-Myc+AurA interaction. This reactivity was seen in the N-Myc 1-137 crosslinking data and disrupts the reactivity pattern required to normalise the data for interpretation.

2.6 Conclusion

In this chapter, we have shown it is possible to generate functional structural data of kinase targets with IDRs that are modulated by IDPs. We demonstrated that the reaction of mass-cleavable crosslinker DSBU with AurA+TPX2 generated distance restraints that could be mapped to crystal structure adapted models of AurA+TPX2 and AurA Apo that were validated by literature in-solution data. However, the resolution of these distance restraints could not confidently discriminate between conformers. Happily, we also showed the changes to normalised monolink reactivity could be used as a post-hoc foot-printing experiment. We attempted to apply our validated protocol that to the high-value oncogenic AurA+N-Myc complex however the unexpected stoichiometry of crosslinked species prevents clear analysis to shed light on the rest of the complex.

Both AurA with TPX2 and N-Myc represent high value targets in the development of novel anti-cancer therapeutics, as AurA and AurA+N-Myc are known to be up-regulated and drive tumorigenesis in certain cancers. AurA in the absence and presence of TPX2 was used to investigate if XL-MS was suitable for structural studies on kinases containing and interacting with disordered domains. As AurA+TPX2 is a high value target, numerous crystallographic structures and biochemical studies have already been performed to understand the nuances of this dynamic interaction, hence

making it a suitable testbed. It was found that the crosslinks generated, after optimisation, matched crystal structure conformations. Our data has also revealed that the NTE of AurA is not as dynamic as the crystal conformations imply, and that a major conformation of active AurA kinase domain places the NTE bound to the N-lobe, F-pocket. TPX2 binding has to displace this NTE conformation. This appears to be unique to AurA as the F-pocket of AurB has a Gly to Asn mutant that differentiates the AurA and AurB cofactors. AurA+N-Myc is significant, and is subject to numerous ongoing drug discovery programmes attempting to generate drugs to disrupt the protective effect of AurA binding to N-Myc to treat N-Myc driven cancers. N-Myc 22-89, the AIR, is the minimal protein binding domain, however only N-Myc 61-89 is found in the crystal structure.

Although the data is weak due to poor coverage, the binding of N-Myc 22-89 appears to not disrupt the NTE conformation and modulates the mono-link in the same manner as TPX2 binding. It is probable, consistent with previously published peptide competition assays, that N-Myc 22-44 is binding into the Y-pocket of AurA. However, the crosslinking of AurA+N-Myc 1-137 and N-Myc 1-137 apo, suggests that the stoichiometry of the interaction needs to be re-examined with MBII potentially mediating a dimer of heterodimers type interaction.[171]

It was found that Cα-Cα distances could be used to interpret crosslinks on crystal structure data. Manually reconciling monolink data to a specific conformation of AurA was only possible through the normalisation of the reactivity of the monolinks across crosslinks. Software to streamline this process against diverse models would increase the usefulness and precision of DSBU XL-MS to study kinases in the context of drug development. Molecular modelling may help understand the stability of these models, especially as the interactors are dynamic.

3 Using XL-MS to evaluate MD-simulated models of a kinase with IDRs and its interaction with an IDP

3.1 Abstract

Aurora-A is a serine/threonine kinase that is critical for the successful progression and regulation of mitosis and is frequently overexpressed in cancers. The transcription factor N-Myc is heavily amplified in many childhood neuroblastomas and late-stage neuroendocrine prostate cancers. As Aurora A has a non-catalytic function in preventing the proteasomal degradation of N-Myc, the disruption of the AurA kinase domain+N-Myc transactivation domain complex is a very high value drug target. Efforts to generate small molecule inhibitors of the complex have been obstructed by an incomplete understanding of the conformational dynamics of AurA and the AurA+N-Myc complex. Here, we take our previously collected DSBU XL-MS datasets of AurA, AurA in complex with the allosteric activator TPX2, and AurA+N-Myc and integrate these with molecular dynamic simulations and a previously reported collision cross-section dataset for AurA±TPX2 to produce a 4-conformer model of AurA. We generated a range of hypothetical AurA Apo models which were scored using solvent-accessible surface distance restraints and residue depths, with the highest scoring models subject to evaluation by MD, and comparison of predicted collision cross sections with experimental data. Surprisingly, a model generated from an AurA domain swapped dimer, where the activation loop was rotated out of the C-lobe, rationalised previously seen phenomena and was stabilised by the addition of TPX2. It is unclear whether this conformer is part of the activation mechanism. We also report that the NTE of AurA, unstructured in X-ray structures, stably binds to the F-pocket on the N-lobe, and that TPX2 competes with the NTE for this pocket. Furthermore, we tentatively propose a polyvalent binding model of the AurA KD+N-Myc 20-89 complex.

3.2 Introduction

3.2.1 Summary of chapter 2 and questions arising

In chapter two, we demonstrated that XL-MS can be used to gain insights into the interactions of IDPs with dynamic kinases. However, the capacity of the crosslinks to discriminate between alternative structures was limited by the distance metric used, the Euclidean Cα-Cα distance, to annotate manually generated models. Using the AurA KD±TPX2 1-43 complex as a testbed system to validate the use of XL-MS, with normalisation, we found that the AurA+TPX2 crosslinks were explicable by both the canonically active and a conformationally open structure. We also found evidence that the AurA NTE was bound to or in close proximity to the N-lobe.

Despite the adequate (albeit imperfect) crosslink coverage, we hypothesised that using a more precise scoring method, in conjunction with structural models that were energy stabilised so that the disordered regions were given more flexibility, this would allow for more confident discrimination between potential conformers. We were inspired by the approach taken by Rozbesky *et al* that used molecular modelling to improve the interpretation of crosslinks, which in turn was orthogonally validated with ion-mobility spectrometry (IMS)-MS data.[193] They took this approach because the crystal structure they had produced did not accurately represent the in-solution conformer of their target protein.

3.2.2 High throughput scoring of structures using XL-MS data

3.2.2.1 Using solvent-accessible surface distance and monolink scoring of structures

Bullock *et al* presented a software that scored models using experimentally determined crosslinks and monolinks by evaluating the solvent-accessible surface distance (SASD) of the crosslinked residue side chains, and the depth distance to solvent water for monolink labelled residues.[219 301] It also allows for the higher throughput scoring of hypothetical models than the annotation method used in Chapter 2.

The use of SASD allows for a more precise measurement of the shortest distance a crosslinker can span between two residues. The Cα-Cα measurement takes the shortest distance possible between the alpha carbons of a residue which, for the dynamic and cylindrical AurA, often passes through regions of the protein that a crosslinker could never access. Cα-Cα is a useful approximation that is computationally cheap to calculate.

The SASD calculation requires the algorithmic calculation of the shortest accessible crosslinker distance, and while it does require more computational energy, it can still be performed on a laptop.[219]

3.2.2.2 Mono-links as pseudo-foot printing MS and additional scoring data

In Chapter 2, the only way to utilise the generated monolink information was in a qualitative comparison with the annotated models of the conformation, and in evaluating the quantitative changes between conditions. MoDs, a monolink scoring algorithm within XLM tools, calculates the distance between the residue and the nearest solvent water molecules.[311] This is based on the theory that the further it is from the solvating water, the less likely it is that a residue will come into contact with a crosslinker molecule. Whilst this disregards any local effects, such as hydrogen bonding or pKa that will also directly affect the reactivity of a residue, this simplification was shown to be effective in discriminating between models being tested.[312]

3.2.3 Using MD with orthogonal structural techniques to conquer a dynamic landscape

The approach taken in Chapter 2, which relied on Cα-Cα distances, annotated onto manually adapted crystal structure models, had two potential drawbacks. Firstly, because of the shapes of the complexes being analysed, the Cα-Cα distances represented crosslinks that were impossible. Secondly, the models generated, although in tolerance by Cα-Cα distances, could be misrepresenting the lifetime average in-solution conformation. We were unsure whether both or one of these were occurring, but because the AurA+TPX2 model that was generated from a domain swapped dimer scored as well as the canonically active AurA+TPX2 structure we could not be confident in our methodology. There are two potential solutions for this. The first is to attempt to deal with the conformers by brute force, as it were, and use either a wide range of PDB models, or use in-silico, ab initio methods.[313] The second is to use MD simulations to generate models that approximate metastable conformations.

3.2.3.1 Proteins with IDRs may require special parametrisation

The molecular dynamics simulations of protein complexes that feature IDRs can be challenging and under some conditions warrant the use of unique force fields. This is because MD simulation parameters have been developed to assume the stable secondary structures and the MD simulation of extend disordered states give misleading stable structures.

Here we will use MD to simulate short regions of IDRs and if these generate structures that stabilised into previously unseen conformations we may have to re-adjust our parameterisation of TPX2 and the NTE of AurA.

3.2.3.2 Manual model generation

The modelling of kinases with disordered regions has been performed across numerous kinase families. Here will be using crystal structures of AurA to generate the input structures for MD and if these models fail to help reconcile the crosslinking data, we may have to use adapted structures from analogous proteins.[314 315 316]

3.2.3.3 AMBER MD– our approach to MD simulations

XL-MS has often required the use of hypothetical or MD-simulated structures to interpret patterns of crosslinking.[279] This is especially true when generating distance restraints, for which a structure is required. Homology modelling is often used to backfill crystal structures that are missing regions or lacking structures for IDRs. The process of simulating structures to generate hypothetical metastable structures can be computationally intense, to ensure an accurate representation of the in-solution structure. For poorly studied proteins this is not a trivial process, as MD simulations can be misleading if poorly designed.

There are two approaches to utilising XL-MS distance restraints with MD simulation. The first is to directly include the distance restraints as parameters in the simulation, a method known as harmonic restraints. The second is a crosslinking blind approach, where distance restraints are used to evaluate the performance of the simulation.[317] As AurA+TPX2 is a well-characterised complex with both orthogonal in-solution data and diverse approaches to its MD simulation, we can directly adapt previously used simulation methods to generate a wider range of potential AurA and AurA+TPX2 conformations. However, the results of MD simulations will always be an approximation of the in-solution conformation ensemble, and so will require further validation via methods orthogonal to crosslinking mass spectrometry.[318]

The programme AMBER has been used to simulate AurA to validate distance restraint data generated by FRET probe. Herein we will directly adapt Ruff *et al*'s method.[164] We will use a clustering algorithm to extract average conformations and use RMSD changes and pseudo b-factor calculations to quantify the fluctuations in dynamics of our MD simulations.

Thus, we will be able to qualitatively estimate the stability of the final model generated from the simulation.

3.2.3.4 AlphaFold and ColabFold

Homology modelling uses previous experimentally generated protein structures to computational predict the structure of another protein with similar amino acid sequence. SWISS-MODEL is a web server that uses a four-step process of the submitted sequence to be modelled with template(s).[319] The template(s) are aligned in space then the target and template(s) primary sequences are aligned, from which the model is built. The model structures then are processed to generate an energy minimized predicted structure.

However, AlphaFold uses a deep learning-based algorithm to build the predicted structures from the given templates that is faster and more accurate than conventional approaches.[320] ColabFold is a webserver implementation of AlphaFold2 that allows for public use, as its database require a vast amount of computational storage.[245] ColabFold allows the rapid prediction of kinase structures templated from the vast number of closely related kinase proteins found in the PBD. For example, a potential mutant structure can be predicted using all crystallographic kinase structures of the related kinome family.[321]

3.2.3.5 The two/three-conformer model of AurA of in-solution

As examined in Section 1.3.3.2, the in-solution conformational landscape of AurA±TPX2 has primarily been investigated by FRET sensor decorated AurA to generate time resolved-(TR) FRET measurements that have been used to generate ensemble distances, measuring the distance between Helix D and the activation loop. Several groups have used TR-FRET, and MD simulations, to suggest that AurA is in a dynamic equilibrium of the MSP1-like and canonically active structures. That equilibrium is shifted by the presence of phosphorylation on the activation loop, and the binding of interaction partners like TPX2. For phosphorylated AurA, its estimated populate 75:25 ratio of canonically active:MSP1-like conformer, but this is shifted to an 85:15 ratio with the addition of TPX2.[272 318] It has been shown through the NMR and EPR monitoring of the water network within the active site of AurA that the equilibrium of the active structure can be sub-divided into two conformations.[156] The equilibrium of these 'active' conformations is modulated by the presence of TPX2. This fits into our understanding of the

biological role of TPX2; however, as we cannot see this equilibrium position by the adjustment of the activation loop, we need to be wary that crosslinking data will be blind to its presence.

3.2.4 IMS-MS appears to give actionable insights into the conformational ensemble of AurA and AurA+TPX2

3.2.4.1 The need to understand conformer count and dwell time

In Chapter 2, we were unable to discriminate between the canonically active and inactive structures. Tomlinson *et al* showed that orthogonal MS methods could describe the conformational ensemble present, which could then be rationalised by crosslinking mass spectrometry, using models informed by MD simulations.[308]

3.2.4.2 IMS shows AurA Apo in three conformers

Fortuitously, during the course of this work, Tomlinson *et al*, using ion-mobility mass spectrometry (IMS-MS) and MD simulations, demonstrated that AurA KD presents as a three-conformer ensemble: major, minor and minimal (Table 3-1).[308] IMS-MS uses soft electrospray ionisation to generate ionised protein conformer ensembles that reflect those found in-solution. These can be separated by both m/z and rotationally averaged collision cross-section (CCS). This allows protein conformers of the same m/z to be separated, effectively, by shape.[322] Tomlinson *et al* also showed that the CCS of MSP1-like, autoinhibited active-like or active-like conformers of AurA (PBD: 4J8M, 1OL7) were found to be so similar that it would not be possible to discriminate between them. Using IMS-MS, unphosphorylated AurA KD was found to have a major conformer (≈65 %, Conformer II) corresponded to the size of the inactive or active conformer such a models built from PBD 1OL7. The minor conformer was found to be ≈30 % of the population of AurA (conformer III) and was larger than a MSP1-like, autoinhibited active-like or active-like conformer of AurA and minimal (<5 %) conformer was smaller. The authors were unable to differentiate, by IMS-MS, DFG loop positioning. The AurA+TPX2 1-43 complex also presented as a three-conformer ensemble.

Thus, each model structure generated for AurA and AurA+TPX2 can be scored against distance restraints (crosslinks), pseudo-footprinting (monolinks), stability (MD simulation), size (CCS), and reference to the

crystal structure. This diversity of analytical tools sets us up for success for describing the in-solution conformation ensemble.

Table 3-1 CCS data of conformers found by Tomlinson et al

Table of measured CCS values of AurA 119-403 C290A C393A apo and AurA 119-403 C290A C393A with TPX2 1-43 found by Tomlinson

Conformer AurA APO	CCS / nm²	Conformer AurA+TPX2	CCS / nm²
I	20.4	I*	20.0
II	22.5	II*	22.4
III	24.8	III*	22.6

3.3 Methods

For MeroX searching, the Protein FASTA sequences used are given in 5.2.1.

3.3.1 XLM tools distance and depth model analysis

Depth analysis was conducted using the Depth webserver (http://cospi.iiserpune.ac.in/depth) to generate *-residue.depth files. With these in hand, XLM was run locally using ubuntu and the default settings to score the selected models.

3.3.2 IMPACT CCS prediction

IMPACT was run locally using an ubuntu terminal. IMPACT uses the projection approximation to estimate CCS values for protein models.[255]

3.3.3 Molecular dynamics simulation

An AMBER MD simulation was instantiated using AMBER-ff14SB force field parameters for the protein and TIP3P water model, along with ADP parameters generated by Carlson[323] and accessed from the Amber Parameter Database[324] with phosaa10 parameter set for phosphorylated amino acids.[325]

3.3.3.1 Input structure generation

PDB files were prepared by opening in PyMOL and manually adding phosphate to threonine residue and saving structures without hydrogens but maintaining CONECT records. Residues and atom names were matched to those found in the above-mentioned parameter sets. A modified version of the ‘JFR-AMBER-from-scratch’ script (unpublished; provided by Dr James Ross, University of Leeds) was then used to prepare input files and submit production runs to the ARC3 high performance computer cluster

3.3.3.1.1 PDB structures 3UOK, 6hjk, 2WTV, 5l8k, 5EW9, 1Ol7, 4c3p and 6c2t

Input structures, 3UOK, 6HJK, 2WTV, 5L8K, 5EW9, 1OL7, 4C3P and 6C2T were adapted from structures generated for Dr Mathew Bachelor.[308] To model Aurora A 119-403 (122-403 including the tag) C290A C393A pT287 pT288, the appropriate gifted structure was loaded into PyMOL with Aurora A PBD structure 7O2V. Residues 119-129 from 7O2V were extracted, aligned, and used to replace the sequence in the named structure. The mutagenesis tool was used to correct the sequence of the replaced residues. The structure was then relaxed to a stable pose using the sculpt tool. Hydrogens were then added to the sequence and the structure was

then relaxed to a stable pose using the sculpt tool a second time. Additionally, pThr were added if they were missing. The command *print(cmd.get_fastastr('all'))* was used to ensure the sequence modelled matched the sequence crosslinked.

3.3.3.1.1.1 Ahanger structure

This structure was provided by Dr Syed Ahanger, University of Leeds (unpublished observations).

3.3.3.1.2 AurA TPX2

3.3.3.1.2.1 ColabFold: AlphaFold2 using MMseqs2 generated models

Structures were generated using the ColabFold webserver using the AurA+TPX2 crosslinked sequence. Paramaters included: multimer unpaired+paired and templates settings without amber with 3 recycles. Incorrect bonding waws manually repaired in pymol and pT287, pT288 put in from PBD code 1OL7.

3.3.3.1.2.2 Manual tpx2

The manually generated model adapted using PBD code 6VPJ aligned to the CF A AurA+TXP2 and the residues of 6VPJ 273-308 were used to replace those of the CF A AurA+TXP2 model.

The same process was performed using PBD code 6VPJ.

3.3.3.2 Automated structure cleaning

‘pdb4amber’ module was used to prepare structures for parameterization with LEaP.

3.3.3.3 Parameterization of the WT AMBER simulations

The modified structure was passed to LEaP of AMBER 20[260] to add hydrogens and generate AMBER-ff14SB parameter (.parm7) and coordinate (.rst7) files for simulation under the AMBER MD engine, including the parameters for ADP and phosphothreonine mentioned above. An explicit TIP3P cubic solvation box was added, extending 12 Å beyond the protein, and neutralizing counterions (Na+ and Cl-) were added. Further ions were added to reach an ionic concentration of 150 mM. The ParmEd tool within AmberTools 18 (35) was used to repartition the hydrogen mass of the AMBER parameter file generated by tLEaP, as per Hopkins *et al*.[266]

3.3.3.4 Minimisation and equilibration for the WT AMBER simulations

Minimisation was performed with 2500 steps of steepest descent followed by 5000 steps of the conjugate gradient method, with 1.0 kcal/mol restraints placed on backbone atoms. The system was then equilibrated with Langevin dynamics for 125,000 1ps time steps, heating to 303.15 K in an NVT ensemble. The SHAKE algorithm was used to constrain bonds involving hydrogen and omit force evaluations for these bonds.

3.3.3.5 Production simulation for WT AMBER simulations

Molecular dynamics was performed as an NPT ensemble, using Langevin dynamics at 303.15 K and the MC barostat at 1.0 bar. 4 ps timesteps were afforded by the hydrogen mass repartitioning. All atomic restraints were removed. The simulations were restarted until a total simulation time of 0.4 µs was achieved.

3.3.3.6 Data analysis

Trajectories from the production molecular dynamics were concatenated and analysed using the CPPTRAJ of AMBER 18.

A clustering analysis was performed to identify different conformational states sampled during the MD simulation by grouping molecular structures using hierarchical agglomerative clustering, into subsets based on their conformational similarity. The RMSF (root-mean-square of atomic fluctuation) and the RMSD (root-mean-square of deviation) were calculated across the trajectory.

3.4 Results

3.4.1 Can the use of algorithmic scoring of XL-MS of MD simulated structural models increase the utility of this approach for IDR-containing kinase targets?

In Chapter 2, crystal structure-generated homology models were used, along with the manual assessment of the mono-link data. Incomplete crosslink coverage, straight-line distance scoring and the dynamic behaviour of AurA±TPX2 meant that two differing conformers were broadly supported. Although it was not conclusively shown, a monomeric conformer generated from a dimeric crystal structure also complemented the experimental distance restraints found. We hypothesised that evaluating energy-minimised AurA-interactor models using SASD calculated distances between crosslinked residues, and also incorporating monolink scoring, would enable better evaluation of the proposed structures.

Previously, structures were generated by adapting incomplete crystallographic structures of AurA±TPX2, using molecular graphics software, PyMOL, to incorporate these regions from the artificial intelligence structure prediction software AlphaFold. The tolerance of Cα-Cα distance restraints applied to these structures had to accommodate both crystallographic artifacts and the dynamic movement of disordered regions, which thus reduced the resolution and confidence of the analysis.

Molecular dynamics simulations of these adapted crystal structure models were proposed, in order to generate energy minimised conformations. It was hoped that this approach would aid in understanding the dynamic IDR parts of the structures. Previously, we annotated these structures directly; here in Chapter 3 we performed MD simulations of them and then annotated them using SASD methods.

An additional goal was to assess the utility of this pipeline: was it worth the time spent on generating energy minimised conformations for evaluation, or were the limits of this process determined by the confidence in the XL-MS data?

It was seen in Chapter 2 that the AurA and AurA+TPX2 crosslink datasets fitted the conformational model generated from X-ray structures, albeit with the position of the NTE altered. Euclidean distances (sometimes called straight line distances or SLD) between the Cα of residues that had formed crosslinks were used to discriminate between hypothetical structures generated from literature. For several reasons, covered in Section 2.5.3, the

lack of crosslinks involving the activation loop of AurA meant that it was difficult to determine the placement of the activation loop with a high degree of confidence (Figure 3-1). The conformation and dynamics of the activation loop are key to understanding the activation of AurA and designing anti-cancer agents for AurA driven cancers, but is still an unanswered question. However, the majority of crosslinks annotated on the homology models generated from the active Apo AurA structures gave Euclidean Cα-Cα distances that were within the tolerance for the crosslinker used (DSBU, Section 2.4.3.1.2). This was taken to indicate that the NTE altered AurA homology model represented the major in-solution conformation that this AurA construct adopts.

As previously discussed, (Section 3.2.3.5), the active AurA KD Apo state presents as an equilibrium of two conformers MSP1-like and active-like that is independent of phosphorylation status. Gilburt *et al* have calculated that the population ratio of inactive/active is ≈30:70.[272] Although there exists no crystallographic structures of the inactive conformer in an Apo form, significant effort has gone into developing small molecules to induce stabilisation of the inactive conformer, thereby inactivating AurA.

Ruff *et al*, using FRET and EPR probes, and Tomlinson *et al*, using native and IMS-MS, report that AurA is a three conformer system, although they come to differing conclusions as to the role of the new third conformer.[164 308]

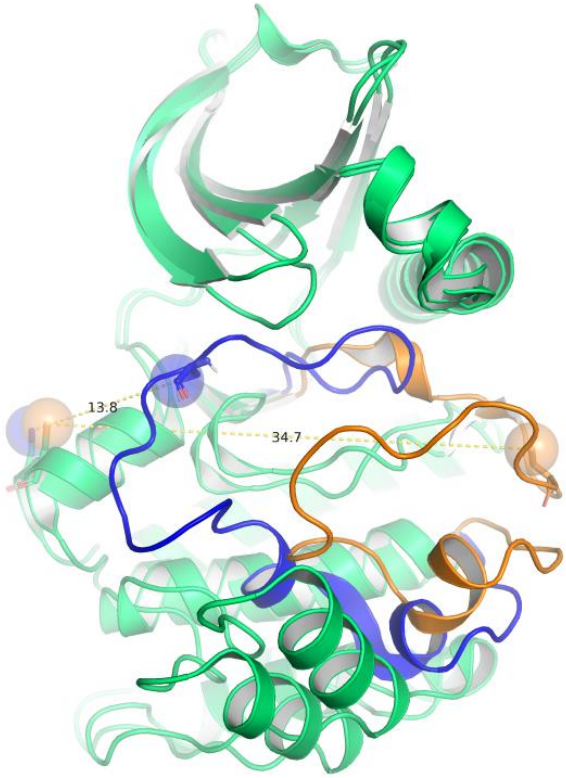


Figure 3-1 Canonical activation loop positions for active and inactive AurA

Crystallographic structures of AurA active conformer (PBD 1OL7, orange activation loop) and inactive (MSP1-like) conformer (PBD 4J8M, blue activation loop). The Phospho-Thr on the activation loop (dark blue) to Helix D distance, Lys224-Thr287 Cα-Cα distance, for the active is 34.7 Å and the inactive is 13.8 Å.

Together, the IMS-MS and MD simulations present a three conformer system, one highly populated compact conformer with a minor open structure, similar to the structure found in dimeric AurA; an ultra-compact low-abundance structure not rationalised to any crystal structures, and the minimal conformer that is smaller than the canonically active structure. The current model of active AurA is that it is composed of a dynamic equilibrium of three conformers. The canonical kinase active structure with a stabilised, phosphorylated activation loop in position, a second conformer that is comparable but inactive due a rotated DFG motif unable to perform phosphorylation, and a final inactive structure where the phosphorylated subdomain of the activation loop is rotated into an inactive position like that found with the MLN- or CD-series small molecule allosteric inhibitors of AurA.[325] The model proposes that allosteric activator proteins such as TPX2, Bora or CEP192, as well as phosphorylation status, alter the equilibrium between these states, by disfavouring the inactive conformations. However, the three conformer model of AurA activation is complicated by the report of a domain-swapped Aurora A dimer being responsible for auto-phosphorylation.[145] Currently there has been little investigation that includes dimerization as an equilibrium state.

However, the uncertainty surrounding this model increased when it was proposed[156] that phosphorylation did not alter the subpopulations of active and inactive AurA, and that only activation interactors, such as TPX2 or INCEP, were required to fully activate AurA. The use of DEER spectroscopy with MD simulations concluded that phosphorylation eliminates an inactive DFG-In state from the conformation ensemble of AurA.[164]

Here, we hypothesise that the additional resolution that energy stabilised models and SASD distances provide will allow us to refine the three-conformer hypothesis, by evaluating a wider range of AurA conformers than we assessed by the AurA crosslinking dataset in Chapter 2. For these conformers, their compaction and stability also needs to be predicted. Gilburt, Ruff and Tomlinson appear to agree that the binding of TPX2 does not fundamentally alter the conformation of AurA, merely inducing a change in the relative abundance of AurA Apo conformers. Consequently, conformers that match the compaction of postulated conformers and that are stable in a MD simulation will be directly adapted for use in the evaluation of AurA+TPX2 conformers. The qualitative evaluation of AurA+TPX2 already suggested a second conformer was responsible for a subset of crosslinks.

The caveats presented in Chapter 2 still remain.

As reported by Lake, how TPX2 and N-Myc modulate the activation loop position of AurA is divergent.[326] As we currently lack experimental CCS data for AurA+N-Myc, and the coverage of crosslinks that was reported in Chapter 2 for AurA+N-Myc is poor, conclusive explorations of this system are not possible. The assessment of the stability of various hypothetical structures by simulation is possible, and preliminary evaluation of these models by AurA intra-protein crosslinks can be performed in the context of the wealth of biochemical data available.

3.4.1.1 Generating a range of testable models of AurA

The vast majority of AurA kinase domain crystal structures surveyed in the PDB were found to vary only by the position of the activation loop, with the hinge flexing to accommodate the activation loop position. Of these, a majority only varied by the position of the DFG and phosphorylated loop subdomains. To accommodate the positioning of the activation loop, there are minor differences in the N-/C-lobe hinging openness. This continuum is expected for kinases that do not feature any control/regulation domains as part of their KD.[318] As the PDB contains such a wide diversity of structures of AurA Apo structures, we considered it unnecessary to generate ab initio AurA models (Figure 3-2).

A panel of AurA structures was generated from the PDB and a colleague's X-ray crystallographic structure (unpublished, Dr Mohd Syed Ahanger). The structures were selected to give the widest range of activation loop positions; from the activation loop bound in the ATP binding pocket to the other extreme of the activation loop extended into solvent. The Cα-Cα distance between 244-287 is used to quantify the activation loop position (the smaller the distance, the closer the activation loop is to the nucleotide pocket and Helix D). This was done to ensure the range of structures showed different positions of the activation subdomains; DFG subdomain, phosphorylated loop subdomain, and structured subdomain.

To ensure that the only variable during evaluation by distance restraint data and predicted CCS was the conformation, the structures were modified so that the residues matched that of the AurA construct used for XL-MS, and phosphorylation pT287/288 and hydrogens were added if they were missing. Any ligands or other modifications were removed. The NTE was taken from the previously adapted 7O2V structure from Chapter 2, as was the CTE from AlphaFold. Where the crystal structures were missing residues from their activation loops, residues from models generated by Matthew Batchelor were used.

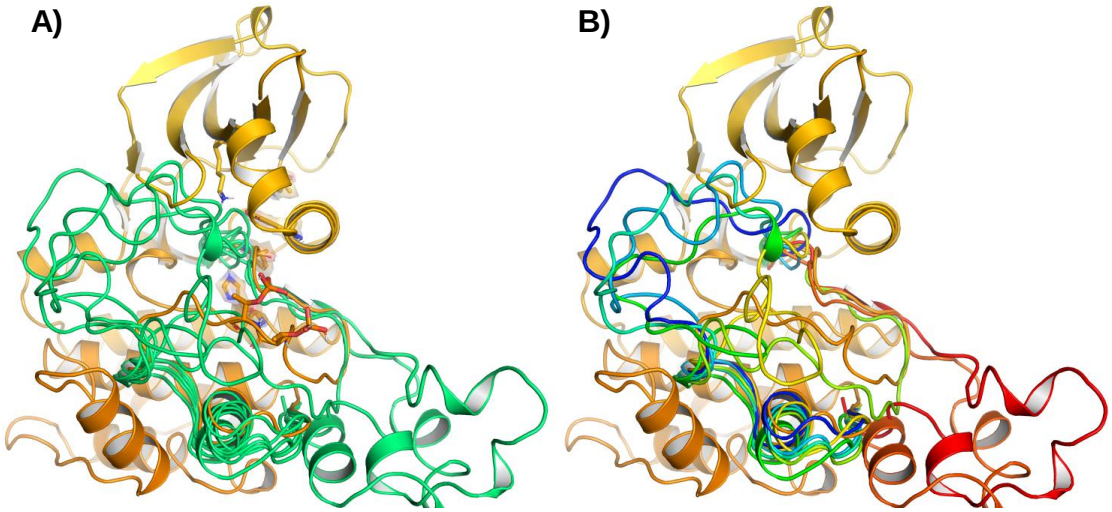


Figure 3-2 Composite figure of activation loop positions used in this study

Panel A) shows residues of AurA 272-307 (lime green) for all structures, overlaid on 1OL7. Panel B) differentiates those chains by colour: dark blue 3OUK, turquoise SYED, turquoise bright green 6HJK, green 2WTV, chartreuse 5L8K, yellow 5EW9, orange 1OL7, terracotta 4C3P, red 6C2T.

3.4.1.1.1 Structures included

Eight crystallographic structures were taken from the PBD, with one unpublished structure also selected (Ahanger structure): 3UOK, 6HJK, 2WTV, 5L8K, 5EW9, 1OL7, 4C3P and 6C2T. These were used to construct all-atom models that matched the AurA construct with appropriate phosphorylation that was used in the crosslinking reactions.

The activation loop opening was assessed by measuring the angle between the α-carbons of Val324, Glu308, and Ala172, as the overall compaction of the structure is related to the hinging open of the C- and N-lobes (Figure 3-3). The angle measured is proportional to the openness of the lobes; the smaller the angle, the more closed the cleft.[160] Additionally, we quantified the activation loop position using the Lys224-Thr287 Cα- Cα distance.

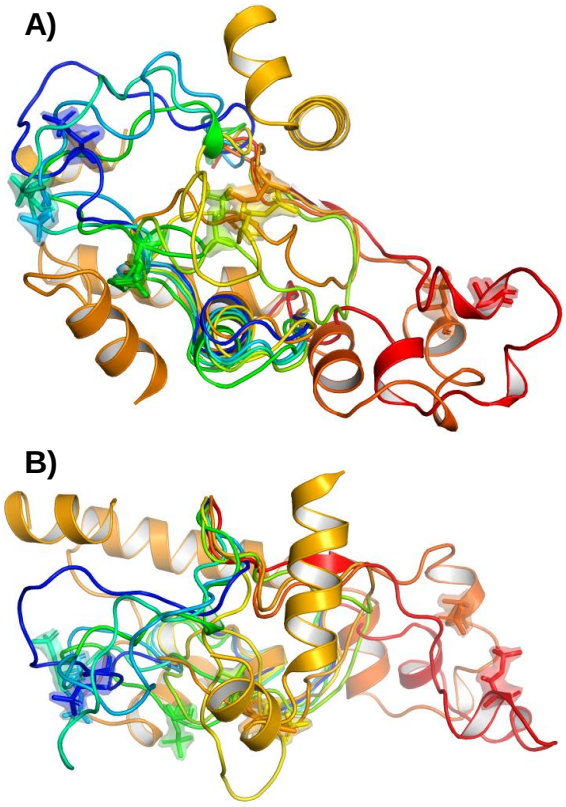


Figure 3-3 Composite highlighting the activation loop positions used in this study

Panel A) is an overlaid diagram of the selected chains, highlighting the αC helix, helix G, D and F of 1OL7. Panel B) is the same figure rotated 90° out of the page on the east-west axis.

3UOK – MSP1-like, DFG-inactive

Martin *et al* presented a face-to-face dimer where the DFG and phosphorylated loop subdomains of the activation loop are pinned, in an inactive conformation, over the D helix, by the small molecule inhibitor YL5-81-1 (Figure 3-4). This is sometimes referred to as 'MLN-like', as this conformation was first seen with MLN8054 in 2WTV.[326] Additionally 3UOK, chain A, is the only model where the 224-287 Cα-Cα distance is less than 35 Å, and the N- and C-lobe are forced wider to accommodate the dynamic position of the activation loop. It is also reminiscent of the inhibitor induced MLN-like structure of an exciting new AurA inhibitor that prevents the formation of the AurA+N-Myc complex (PHA-680626; PBD 2J4Z).

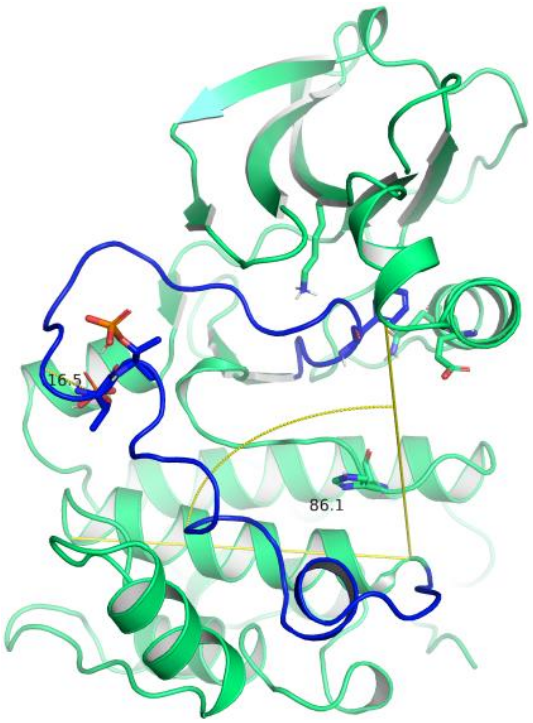


Figure 3-4 Structure 3UOK

Input model from 3UOK. Phospho-Thr on the activation loop (dark blue) in contact with Helix D with a Lys224-Thr287 Cα-Cα distance of 16.5 Å and an angle formed between the Cα's of residues 172-308-324 of 86.1°.

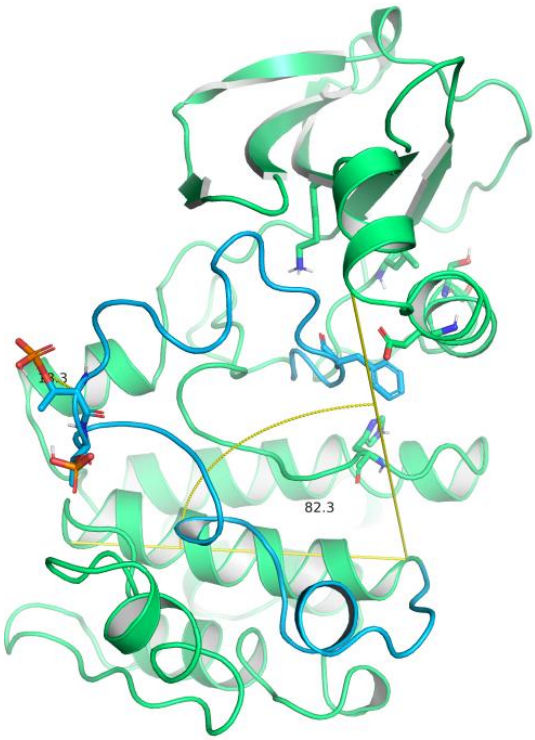


Figure 3-5 Ahanger structure

Input model similar to 1MUO. Phospho-Thr on the activation loop (light blue) in solvent with a Lys224-Thr287 Cα-Cα distance of 18.3 Å and an angle formed between the Cα's of residues 172-308-324 of 82.3°.

Ahanger structure – MSP1-like, DFG-active

Dr Mohd Syed Ahanger (University of Leeds). This unpublished structure was used in place of similar structures 1MUO and Chain A of 2C6E but with a fully structured activation loop requiring re-modelling (Figure 3-5). This is the only structure where the glycine rich loop is directly in contact with activation loop, so that roof of the active site is displaced out of the position.

6HJK – MSP1-like, DFG-NT

Ocasio *et al* presented AurA in complex with a type II inhibitor that presents as an MLN-like inactive structure (Figure 3-6).[327] It is distinct from the previous structures because the inhibitor binds into the nucleotide pocket and displaces the catalytic lysine, which further disrupts the binding of the DFG subdomain to the N-lobe. The phosphorylated loop subdomain is unstructured.

A homology model of the missing residues was generated as part of previous work by a colleague.

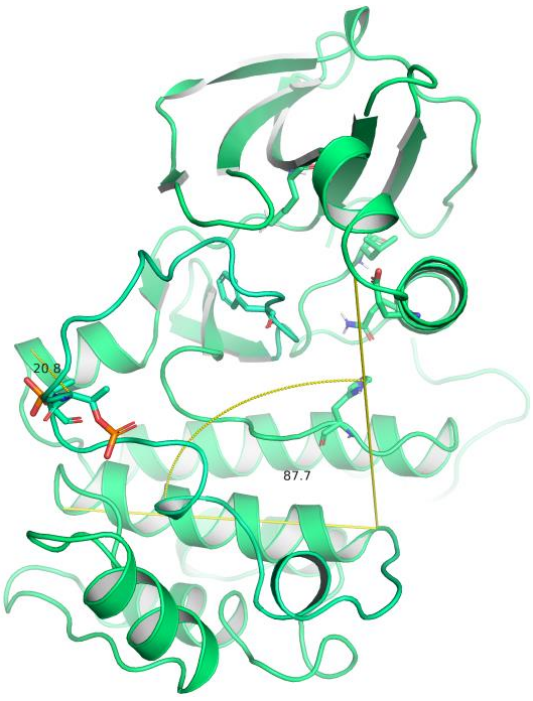


Figure 3-6 6HJK structure

Input model from 6HJK. Phospho-Thr on the activation loop (teal) in solvent with a Lys224-Thr287 Cα-Cα distance of 20.8 Å and an angle formed between the Cα's of residues 172-308-324 of 87.7°.

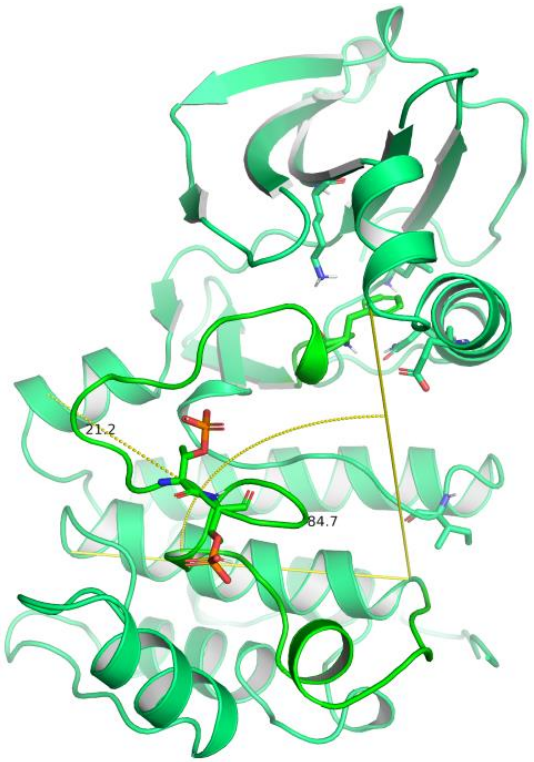


Figure 3-7 2WTV structure

Input model from 2WTV. Phospho-Thr on the activation loop (pea) in contact with the αC-helix with a Lys224-Thr287 Cα-Cα distance of 21.2 Å and an angle formed between the Cα's of residues 172-308-324 of 84.7°.

2WTV – Alpha-helix contact, DFG-inactive

Dodson *et al* generated an AurA KD mutant, 122-403 L215R T217E R220K, that imitates the residues with which AurB binds to MLN8054 (Figure 3-7).[328] The structure contains the AurA KD crystallised with MLN8054, hence the generator of the MSP1-like class of conformers. However 6HJK, 2WTV and the Ahanger structure all differ in their placement of the phosphorylation marks found in the crystallographic structures.

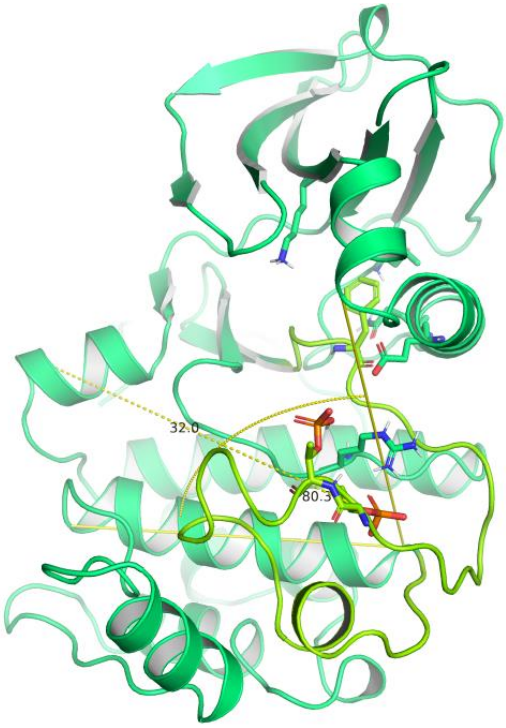


Figure 3-8 5L8K structure

Input model from 5L8K. Phospho-Thr on the activation loop (limon) in active conformation with a Lys224-Thr287 Cα-Cα distance of 32.0 Å and an angle formed between the Cα's of residues 172-308-324 of 80.3°.

5L8K – Autoinhibited, DFG-out

Bayliss *et al* presented AurA in complex with a synthetic vNAR single domain binding into the PIF-pocket, 5L8K (Figure 3-8).[329] The vNAR is an allosteric inhibitor, engaging the Y-/W-pockets, and overlapping with the TPX2 binding sites. The activation loop is placed in an active-like conformation with a DFG motif that is still broken. The nucleotide pocket is still accessible. It is important that this structure is included, as the conformation of the DFG motif has been experimentally shown to be analogous to the auto-inhibited DFG-out inactive structure present as the activation loop in inactive conformer.

5EW9 – Alpha-helix contact, DFG-out

de Groot *et al* presented AurA in complex with a type I inhibitor where MK-5108 (VX-689) cleft between the lobes is compacted, which displaces the phosphorylated loop subdomain into solvent (Figure 3-9).[330] The structure is notable as the C-lobe is twisted ≈5-8 Å clockwise, around a north-south axis, so that the Helix G is in line with the αC helix.

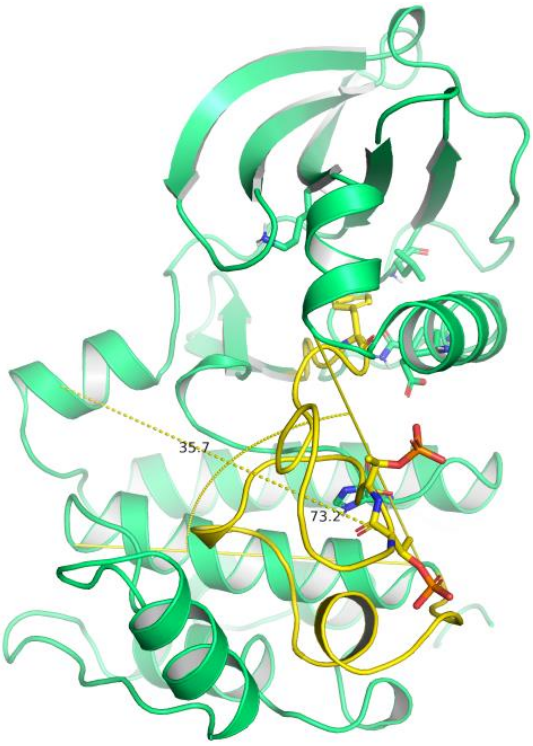


Figure 3-9 5EW9 structure

Input model from 5EW9. Phospho-Thr on the activation loop (yellow) in contact with the αC-Helix with a Lys224-Thr287 Cα-Cα distance of 35.7 Å and an angle formed between the Cα's of residues 172-308-324 of 73.2°.

1OL7 – Active

This structure is believed to represent the canonically active AurA conformer, as the R-spine and salt bridge are formed, and the activation loop is positioned to accept substrate (Figure 3-10).[144]

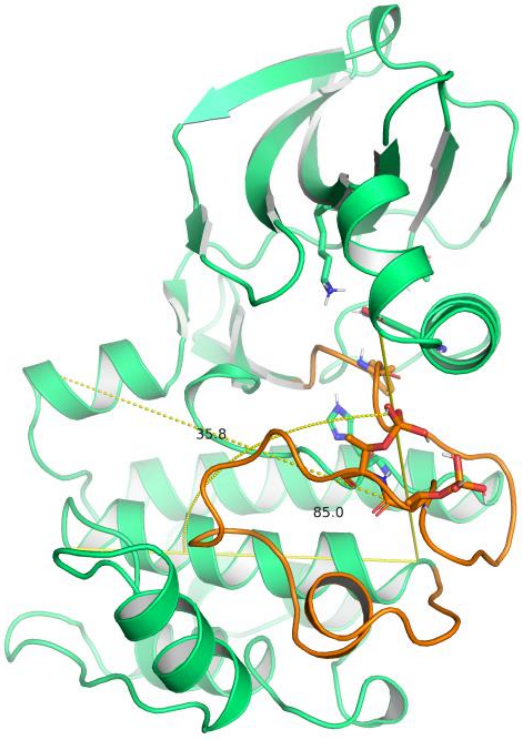


Figure 3-10 1OL7 structure

Input model from 1OL7. Phospho-Thr on the activation loop (orange) in contact with the αC-Helix with a Lys224-Thr287 Cα-Cα distance of 35.8 Å and an angle formed between the Cα's of residues 172-308-324 of 85.0°.

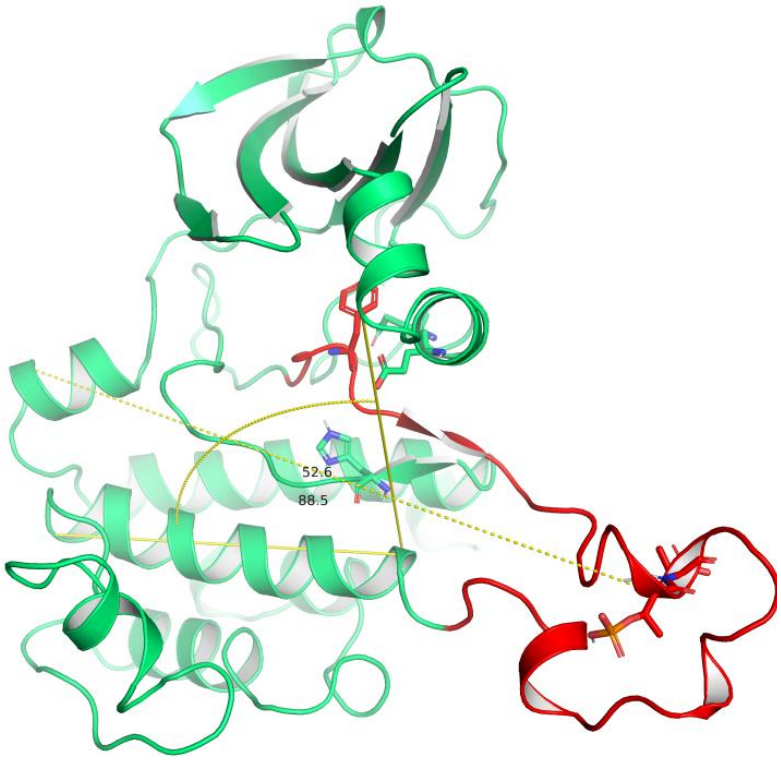


Figure 3-11 4C3P structure

Input model from 4C3P. Phospho-Thr on the activat (red) in solvent with a Lys224-Thr287 Cα-Cα distan Å and an angle formed between the Cα's of residue 308-324 of 88.5°.

4C3P – Structured subdomain out, continuous helix, DFG-active

Zorba *et al* presented a structure where the activation loop is fully out of the cleft and helix 298-305 is moved out to support the loop (figure 3-11).[145] The conformation of Trp313 prevents collapse of the C-lobe. In addition, Arg251 seems to support the helix of pT287/T288 in its extended position. The activation loop structured subdomain is rotated as an extension of helix F.

6C2T – Structured subdomain out, discontinuous helix, DFG-inactive

Gong *et al* generated a face-to-face domain swapped dimer that presents as a monomeric structure (Figure 3-13).[331] The activation loop is fully out of the cleft, almost forming a discreet domain, supported by the stretching of the C-lobe and structured subdomain. R-spine could be formed with minor adjustments to Phe275.

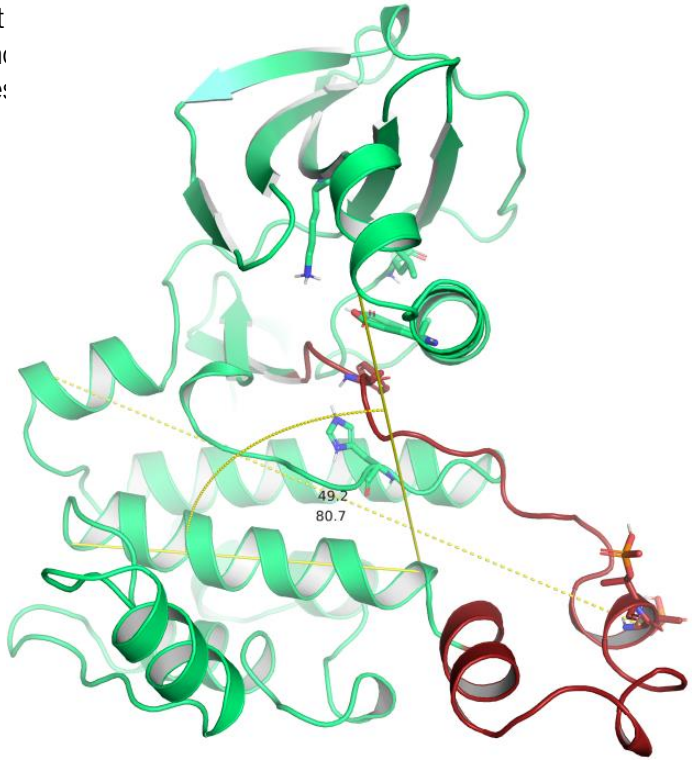


Figure 3-12 6C2T input structure

Input model from 6C2T. Phospho-Thr on the activation loop (ruby) in solvent with a Lys224-Thr287 Cα-Cα distance of 49.2 Å and an angle formed between the Cα's of residues 172-308-324 of 80.9°.

3.4.1.1.2 Assignment criteria

The PBD contains >150 crystallographic structures of human AurA with a plurality of them containing AurA complexed with small molecule inhibitors. To ensure that we achieved a representative range of potential or hypothetical AurA structures,[139] the manually selected eight structures were assessed using quantitative and qualitative metrics, summarised in Table 3-2. As previously discussed, the distance from the activation loop to Helix D, N- to C-lobe openness, and the theoretical likelihood of kinase activity, represent the primary structural variations of AurA. A kinase structure is likely able to perform catalysis if the DFG motif is correctly positioned for a nucleotide binding, the Lys-Glu salt bridge is formed, and the R-spine is formed. If these are present in the structure with an activation loop able to bind substrates, then it is likely to be catalytically active.

To appraise the capacity for catalytic activity of the kinase models using categorisation of the DFG position, we used the Dunbrack category webserver, in addition to inspection of the R-Spine and salt bridge formation.[49] The structure 4CEG was used as a template for R-Spine and salt bridge formation, as it accords with previously validated active positions.[331] The DFG category, formation of the R-Spine and salt bridge can be used to infer if the model is active, since this will be the case if the DFG is in and R-Spine and salt bridge are formed, (Table 3-2). This approach ensured that each structure included was unique, and that the largest range of AurA structures were included.

Selecting AurA models where the activation loop structured subdomain is displaced into alpha-helix contact or continuous-/discontinuous-helix contact was challenging, as these structures were found to contain extended unstructured portions of the activation loop and/or be modelled as dimeric structures. Hence structures from PBD codes 5EW9 and 6HJK were selected as Bayliss group member Dr Matt Batchelor had previously generated all-atom models of these crystallographic structures. The Ahanger structure was included as it has a fully structured activation loop.

The continuous-/discontinuous-helix activation loop conformers were more challenging as the majority of these structures were generated with redox modified residues and ligands. The discontinuous-helix structure 6C2T was selected since it lacked an activation loop cysteine, and many of the comparable structures were generated with a disulfide bridge present in the dimeric AurA structure. The structure 3O51 was also considered but it

appears to have a very similar activation loop position to 6C2T, but with fewer residues modelled.

For the continuous-helix model, 4C3P was chosen, as its activation loop is fully modelled. 2BMC was considered as an intermediate structure between 1OL7 and 4C3P with a 224-287 Cα-Cα of roughly ≈40 Å. However, 2BMC has fewer activation loop residues structured, and is not dissimilar enough to 4C3P to warrant inclusion, since the autoinhibited DFG motif matched that found in 5L8K and 6C2T.

Table 3-2 Summary of AurA APO models

Summary of input AurA Apo structures constructed for SASD evaluation

PDB	3UOK	Ahanger	6HJK	2WTV	5L8K	5EW9	1OL7	4C3P	6C2T
Parent Construct	123–390	119-403 C290A C393A	119-403 L210C, C290A, and C393A	119-403 L215R T217E R220K	119-403 C290A C393A	123-390	119-403 C290A C393A	122-403	122-403
Crystallographic ligand(s)	Type I SM	ADP nVAR cysteine reactive covalent SM	Type II SM	Type II SM	ADP nVAR	Type I½ SM	ADP	AMPPCP TPX2 1-43	Type I SM
224-287 Cα-Cα / Å	16.5	18.3	20.8	21.2	32.0	35.7	35.8	49.2	52.6
172-308-324 Angle / °	86.1	82.3	87.7	84.7	80.3	73.2	85.0	80.7	88.5
DFG category	DFGinter	DFGin	DFGout	DFGinter	DFGinter	DFGinter	DFGin	DFGin	DFGinter
Dunbrack category	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	BLAminus	Unassigned	Unassigned
Rspine (275 + 254 + 185 + 196)	Broken, Phe275 and His254 flipped	Broken, Phe275 and His254 rotated	Broken, Phe275 flipped	Broken, Phe275 and His254 flipped	Broken, Phe275 flipped	Broken, Phe275 and His254 flipped	Formed	Formed	Broken, Phe275 flipped
Salt bridge (162 + 181)	Broken, Phe275 bisects residues	Broken, Trp277 bisects residues	Broken, Lys 162 in nt pocket	Broken, Phe275 bisects residues	Broken, Phe275 bisects residues	Broken, Phe275 bisects residues	Formed	Formed	Broken, Glu181 held between backbone CO of 275 and 177
Predicted CCS / nm^2	23.16	23.79	23.66	22.95	22.54	22.90	22.56	25.27	25.47

3.4.1.1.3 Evaluation criteria

The eight AurA Apo models were evaluated using the experimental crosslink data set, and their predicted CCS values were compared to experimental CCS data from the literature.

IMPACT was used to predict the CCS values of the AurA models. An average projection approximation value from three separate calculations was used.[332] All settings were set to default, except 40,000 independent runs were performed insetead of the default 32. No correction factor was applied.

XLM tools was used to score the AurA models against the collected XL-MS dataset.[219] When scoring a set of potential models using experimentally determined crosslinks and monolinks, XLM tools first scores each model using crosslinks and monolinks separately, and then calculates a combined crosslink and monolink Z-score.

XLM tools is not able to weight crosslinks or monolinks relative to the number of cross-link-spectrum matches (CSM) found for each crosslink or monolink. Scoring was thus performed using crosslinks with more than a single CSM, or using the normalised >50 % cut-off established in Chapter 2. This should enable analysis of the models using the most relevant selection of crosslinks and monolinks.

MoDS was used to score monolinks, which calculates the negative sum of the predicted residue depth. The residue depth prediction is performed by the *Depth* web server,[311] which predicts the distance, in angstroms, to the nearest bulk water molecule to residues of the models. The maximum score is therefore 0 Å, where all surfaces' residues are fully solvated, and the lowest score is infinite. For the 1OL7 model, the smallest residue depth was found to be 3.62 Å on Lys401, and the largest 9.47 Å on Lys162. For example, If all the monolinks were on residues as deep as Lys401, this would score -70.2 Å and if all were on Lys162 the score would be -189.4 Å.

The crosslink-only score is MNXL, which is comprised of three parts: number of violations (NoV), non-accessible crosslinks (NoNa), and expected SASD (ExSASD). The NoV is the number of crosslinks that exceed the distance tolerance of the crosslinker; NoNa is the number of crosslinks that involve a non-accessible residue, and ExSASD is the expected SASD between inter-subunit crosslinked residues.

If a crosslink is from an inaccessible residue this scores a flat -0.1 penalty, if it scores over the accepted nz-nz SASD crosslinker distance this also scores a flat -0.1 penalty, and if the crosslink is within tolerance it is scored positively on a Gaussian distribution centred on the ideal distance; the maximum is +0.06. Thus for our eighteen crosslinks, the theorical highest score possible is +1.08 and the lowest is -1.8.

$$Z = \frac{(x - \tau)}{\sigma}$$

Table 3-3 Variables for normalisation equation

Normalisation of the SASD or MoDs used by XLM tools

Symbol	Variable
Z	Z-score - normalised SASD or MoDs score
x	SASD or MoDs score
τ	τ is the average SASD or MoDs score
σ	standard deviation of SASD or MoDs score

Consequently, for a given crosslink+monolink list with a set of models, the calculated scores are converted into Z-scores, and a combination score (XLMO) is calculated by summing the two scores.

The scoring of closely inter-related models of monolinks and crosslinks by XLM tools can on occasion be counterbalancing. With more open and extended structures, monolink scores will increase (the negative penalty reduced) but crosslink scores will also be reduced because the distances between residues increase. The converse is also true: crosslinks scores will increase with a more compact structure, centring on the ideal distance, and the negative penalty monolinks increase as the depth of the residues increases. We noticed that there are occasions where inter-related models are scored dramatically differently by XLMO, the combination score, but most of the models score well by the crosslink and monolink scores. This led to a need to carefully interpret the difference in XLMO score between models and the crosslink dataset, along with the specific distance restraints that are violated and the number of such violations, as well as the correlation of the distance restraint and mono-link score. Where the differences between model structures are slight, or sub-populations of crosslinks are within tolerance for groups of models, the MNXL, and thus the XLMO, score offers little insight. This also demonstrates that the selection of crosslinks and monolinks used in scoring closely inter-related models can dramatically change the scores of each of the models and needs careful consideration.

3.4.1.1.3.1 Using AurA Apo crosslinks and monolinks with more than a single CSM

We saw in Chapter 2 with SLD and models that, apart from the placement of the NTE of AurA, there was little to differentiate the active and inactive structures, since all the crosslinks with more than a single CSM were within tolerance of the Cα-Cα distance for DSBU. All structures had crosslinks within tolerance of at least thirteen of the eighteen crosslinks (Table 3-8). Models 1OL7 and 6HJK had sixteen crosslinks within the acceptable SASD for the crosslinker. For the 1OL7 input structure, crosslinks K124-K156 and K124-K171 were over-long, see Table 5.2.3.1.3. Intriguingly, the smallest structure by CCS estimate, 1OL7, was scored highest using jwalk (crosslinks only) and the largest, 6C2T, was scored highest by MoDS (mono-links only).

Table 3-5 Input table data from crosslinks with more than a single CSM

All input structures subject to XLM tools analysis using AurA Apo n>1. MoDs is the evaluation by monolink depth (average -100.5±3.5), MNXL is the evaluation by crosslink SASD (average 0.48±0.08), again the larger the better. XLMO is the combined score, the larger the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off.

PDB	3UOK	Ahanger	6HJK	2WTV	5L8K	5EW9	1OL7	4C3P	6C2T
MoDS	-102.74	-95.85	-105.80	-103.01	-101.47	-97.89	-104.29	-97.52	-96.46
MNXL	0.50	0.36	0.43	0.45	0.53	0.34	0.57	0.59	0.53
XLMO	-0.31	-0.02	-1.60	-0.80	0.28	-0.72	0.00	1.76	1.41
Crosslinks over	2	3	1	2	1	3	1	1	1
Place by XLMO	6th	5th	9th	8th	3rd	7th	4th	1st	2nd
Taken forward for MD	No	No	No	No	No	No	Yes	Yes	yes

The average MNXL score of the nine structures was 0.26 (stdev=0.13), and the average MoDS score was -95.55 (stdev=3.5). Examining the XLMO scores of the nine structures, we can divide them into three classes: high scoring (4C3P and 6C2T), good scoring (5L8K, 5EW9 and1OL7) and low scoring (3UOK, Syed, 6HJK and 2WTV). These three classes roughly correspond to the three classes of activation loop: out, active and inactive. The crosslink, Lys156-Lys162, was not found in tolerance for any model evaluated (Table 3-7). Lys156-Lys162 was normalised with a value of 5.5, and so was to be weighted as important for AurA Apo. This is potentially explained by the flexing of the glycine rich loop but

Table 3-4 Distances of the AurA Lys156-Lys162

The crosslink Lys156-Lys162: 6 CSMs was not found in-tolerance (< 35 Å) with any model AurA structures in Chapter 2

4C3P		1OL7		6C2T	
SASD / Å	Euclidean / Å	SASD / Å	Euclidean / Å	SASD / Å	Euclidean / Å
40.1	17.9	42.6	18.2	45.7	18.2

this is just a hypothesis at this stage. Structures that place the activation loop covering the nucleotide pocket scored poorly. The active structures 1OL7 and 4C3P scored well. Using crosslinks and mono links with a count larger than 3 to evaluate the structures placed 1OL7 into fourth place, with 5EW9 in third, and 4C3P and 6C2T still first and second respectively. As 1OL7 is thought to be the canonically active structure, and 4C3P and 6C2T scored highly, these three structures were taken forward to MD. Figure 3-13 shows that the general trend of each of the structure's SASD distances is similar.

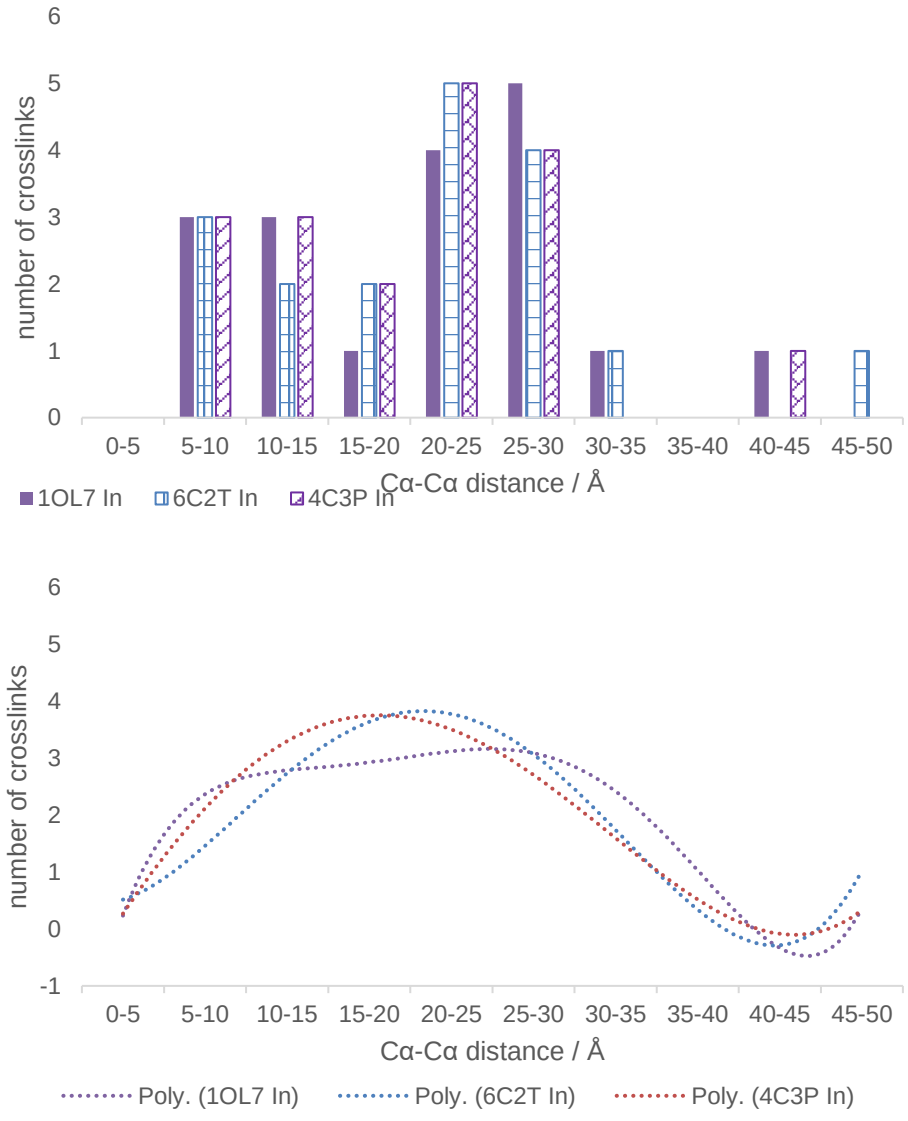


Figure 3-13 Histogram of distances from AurA Apo models

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM.

3.4.1.1.3.2 Using AurA Apo crosslinks and monolinks with more than a single CSM and normalised reactivity of > 50 %

We did not observe any improvement in the discrimination between models using the normalised cut off of crosslinks which suggests some improvements need to be made to either crosslink coverage or models being assessed (Table 3-9).

Table 3-6 Input table data – Normalisation cut-off

All input structures subject to XLM tools analysis using AurA Apo n>1 and normalisation of >50 %. MoDs is the evaluation by monolink depth (average -60.5±3.4), the larger the better. MNXL is the evaluation by crosslink SASD (average 0.27±0.06), again the larger better. XLMO is the combined score, the larger the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off.

PDB	3UOK	Syed	6HJK	2WTV	5L8K	5EW9	1OL7	4C3P	6C2T
MoDS	-63.17	-56.94	-65.34	-62.18	-59.67	-59.05	-66.08	-58.22	-56.08
MNXL	0.32	0.21	0.18	0.32	0.26	0.19	0.29	0.33	0.31
XLMO	0.12	0.07	-1.97	0.35	0.15	-0.64	-0.84	1.32	1.45
Crosslinks over	1	2	1	1	1	2	1	1	1
Place by XLMO	5th	6th	9th	3rd	4th	7th	8th	2nd	1st
Taken forward for MD	No	No	No	No	No	No	Yes	Yes	yes

3.4.1.2 Molecular dynamics simulations of high scoring Aurora A 122-403

3.4.1.2.1 Preparation of molecular dynamics simulations of high scoring Aurora A 122-403

Molecular dynamics simulations of the selected AurA KD conformers were performed to investigate the mobility of the IDRs and to generate “stabilised” or energy minimised structures that approximate the in-solution equilibrium conformation of the respective structure. The structures generated from the domain-swapped dimers, 4C3P and 6C2T, potentially constitute unstable monomers, and were expected to undergo a longer stabilisation, or perhaps to be inherently unstable. Performing MD simulations of AurA KD with pT287, pT287+pT288 or without phosphorylation, in the presence and absence of interactors, to study the conformational changes seen in experimental data has been established by several groups.[318]

We wanted to simulate the conformers as the AurA protein would theoretically appear in the crosslinking reaction. To this end, all models were simulated with an identical sequence of the construct of AurA used in the reaction as pT287/pT288, 150 mM NaCl and no nucleotides such as ADP or AMP-PCP. These conditions were chosen because Musavizadeh *et al* demonstrated that the choice of simulating AurA APO with pT287/pT288 or

pT287 affects the dynamics of the activation loop, as does the presence of ADP+2Mg^{2+}.[318] This is despite the variation in sequence and phosphorylation seen in the adapted PDB structures . Both Tomlinson *et al* and Burgess *et al* found by native MS that the AurA 119-403 double cysteine mutant construct produced a doubly phosphorylated activation loop, hence here we will run MD simulations with pT287/pT288.[308 274] Each selected input structure was run once, in 150 mM NaCl buffer, at 303.15 K for 0.4 µs minimum. The entire simulation was then subject to algorithmic clustering into ten structures. Pseudo b-factor analysis of each residue was calculated for the entirety of the simulation.

3.4.1.2.2 Initial analysis of the output of MD simulations

Examining the RMSD traces, which are a measure of the positional difference between every frame of the simulation and the initial input structure, we see that all the structures undergo a period of dramatic stabilisation, followed by an extended search for a stable conformation with more minor changes and a final period of stability until the end of the simulation (Table 3-10). The simulation of the model constructed for 1OL7 appears to have found a stable conformation; this is not unexpected as 1OL7 is the canonically active AurA conformer, (Figure 3-14). The models for 4C3P and 6C2T fluctuated more, but not to the extent to suggest a broken or unstable conformer. This is further evidenced by the stability of the temperature and predicted CCS, an indirect measure of the volume of the conformer.

The average predicted CCS of all the clustered structures, explained below, is 23.85 nm^2, which is roughly midway between those of conformer I and II (23.65 nm^2); as defined by Tomlinson *et al.* The smallest predicted size from a cluster structure for the 1OL7 simulation at 22.64 nm^2, and the largest is from a cluster structure for 4C3P at 25.57 nm^2. It is interesting to note that structures 4C3P and 6C2T, both underwent a minor amount of compaction.

Table 3-7 Data from AurA Apo MD

Key output metrics for the three AurA Apo simulations.

Simulation	Average temp / K	Average RSMD / Å	Time / µs	Input predicted CCS / nm^2	Average predicted CCS / nm^2	Stdev of CCS
1OL7	303.42	3.90	0.616	22.56	23.12	0.43
4C3P	303.41	6.84	0.426	25.27	24.22	0.60
6C2T	303.38	6.06	0.510	25.47	24.92	0.83

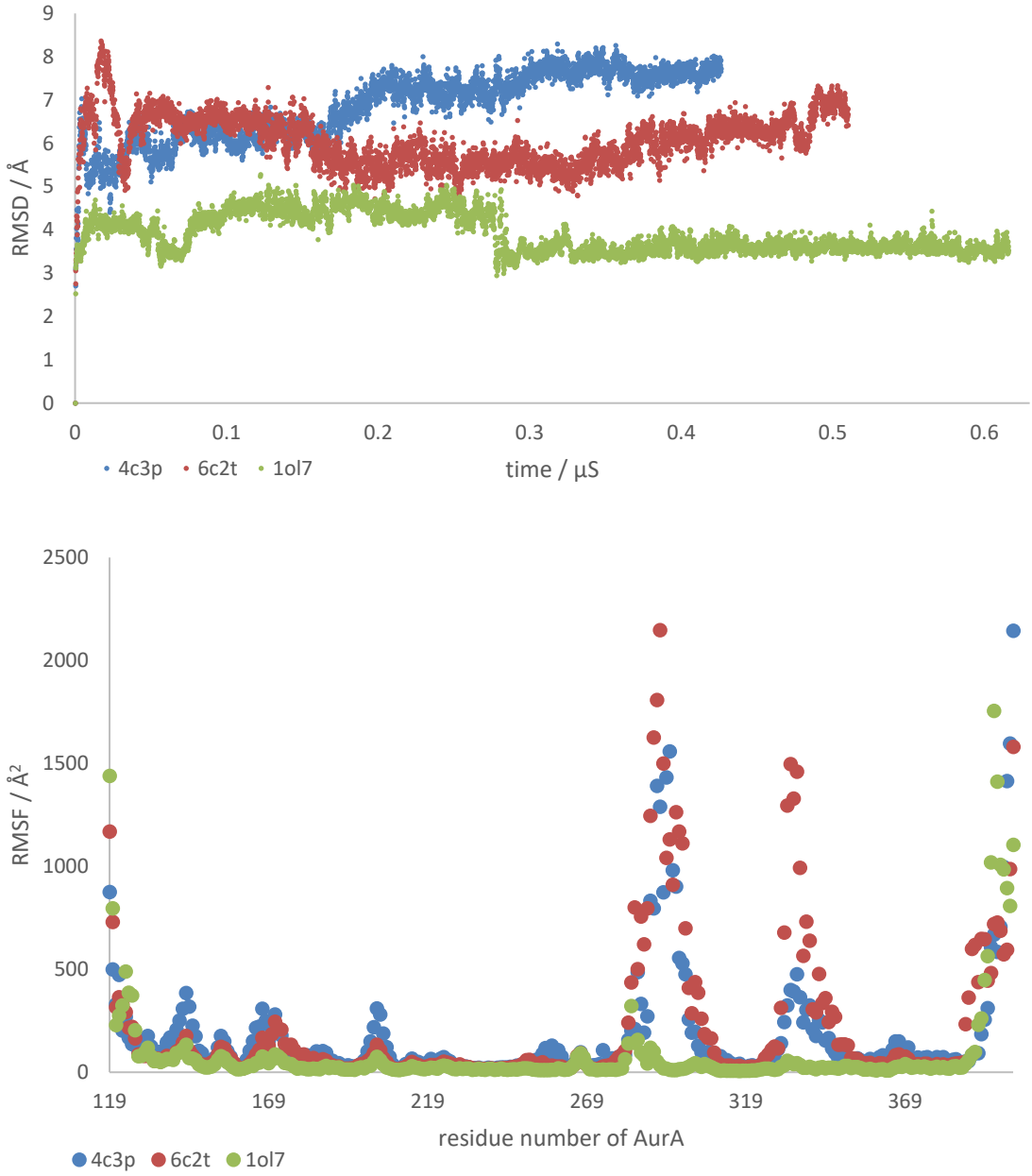


Figure 3-14 RSMD traces and pseudo b-factor traces from AurA Apo MD length of each simulation

(Upper) RMSD trace for each simulation. (Lower) Pseudo b-factor analysis calculated for the entire course of the simulation for each residue.

The calculation of pseudo b-factors for each residue from the entire length of the three simulations revealed that 4C3P and 6C2T had two regions of mobility that do not present in the 1OL7 simulation, the activation loop and Helix-G. A pseudo b-factor is analogous to the b-factor used in crystallography to measure the uncertainty and thus the mobility of an atom or residue. Examining the pseudo b-factors, the activation loop ≈283-305 was equally mobile in both 4C3P and 6C2T. The second region, the Helix-G ≈330-345, was significantly more mobile in 6C2T. As the activation loop has been rotated out of position, it is expected that the Helix-G will become mobile, as it now lacks the support of the C-lobe; it is striking that the mobility is significantly higher in one conformer, when both structures appear to have been equally supported initially.

3.4.1.3 Evaluation of clustered output structures

3.4.1.3.1 Clustering process

The hierarchical agglomerative algorithm, implemented in the cpptraj module of AMBER, was used to produce ten structures from the simulation. It groups similar conformations to give an ‘average’ structure with the fraction of the simulation used to generate the cluster reported (see tables in appendix 5.3.2.2-4). These are named C0 to C9, with C0 being the cluster generated from the largest fraction of the simulation (most frames) and C9 being generated from the smallest fraction. For each simulation a cluster was generated at the end of the simulation that represents the stable conformer found; for the 1OL7 and 6C2T simulations, this cluster was generated from a large proportion of frames at the end of the simulation. This was interpreted to mean that these simulations had produced stable conformations. However, for 4C3P, which underwent a collapse of the nucleotide pocket during the simulation, the cluster with the largest number of frames, hence the average structure, came prior to this collapse, suggesting that this conformation is unstable.

From the three simulations, the clustered structures with the largest three fractions of frames were taken forward for analysis by XLM tools. This was to ensure that transient conformers were not subject to evaluation.

3.4.1.3.2 Using crosslinks and monolinks with more than a single CSM to evaluate the clustered structures.

As with the input structures, most crosslinks fell within the SASD tolerated by the crosslinker. For the nine structures, the average number of matched crosslinks was 15.8 out of 18 total observed crosslinks (Table 3-11).

Examining the standard deviations for MoDS (monolink), we see that they are comparable for input structures, but for MNXL (crosslink) scores, the deviations have slightly increased. As all the structures present with a similar number of in-tolerance crosslinks, ≈80-90 %, XLM tools cannot meaningfully differentiate between them, beyond confirming that they are accurate representations of the in-solution structure. This is a remarkable finding as one model, 1OL7, is representative of the canonically active structure of AurA, while the other two models were generated from domain swapped dimers (Figure 3-16).

Table 3-8 Scoring of clustered output models from AurA Apo MD using all crosslinks and monolinks

All output clustered structures subject to XLM tools analysis using AurA Apo crosslinks and monolinks with more than a single CSM. MoDs is the evaluation by monolink depth (average -99.4±3.9), the larger the better. MNXL is the evaluation by crosslink SASD (average 0.41±0.11), again the larger the better. XLMO is the combined score, the larger the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off; 224-287 Cα-Cα is the measure of the activation loop relative to Helix D; 'Predicted CCS' as found using IMPACT, and 'Conformer' denotes which experimental conformer the CCS value fits in. Conformer descriptors defined in table 1-4.

PDB	**1OL7**			**4C3P**			**6C2T**		
Cluster	C0	C1	C2	C0	C1	C2	C0	C1	C2
Fraction of sim / x/1	0.53	0.04	0.396	0.517	0.141	0.141	0.362	0.316	0.201
MoDS	-100.5	-103.44	-99.9	-107.55	-102.41	-101.18	-96.39	-98.32	-96.46
MNXL	0.48	0.32	0.33	0.45	0.52	0.44	0.47	0.43	0.44
XLMO	0.21	-1.43	-0.67	-1.35	0.11	-0.20	0.97	0.30	0.75
Crosslinks over	2	3	3	2	2	2	2	2	2
Place by XLMO	4th	9th	7th	8th	5th	6th	1st	3rd	2nd
224-287 Cα-Cα / Å	37.2	38.5	38.5	44.4	47.3	47.3	51.8	44.9	46.0
Predicted CCS / nm^2	22.92	22.7	22.92	23.53	23.99	24.22	24.67	22.7	24.16
Conformer	II	II	II	III	III	III	III	II	III
Conformer descriptor	Active-like	Active - like	Active -like	Structured subdomain out, continuous helix, DFG-inactive	Structured subdomain out, continuous helix, DFG-inactive	Structured subdomain out, continuous helix, DFG-inactive	Structured subdomain out, dis-continuous helix, DFG-inactive	Structured subdomain out, dis-continuous helix, DFG-inactive	Structured subdomain out, dis-continuous helix, DFG-inactive

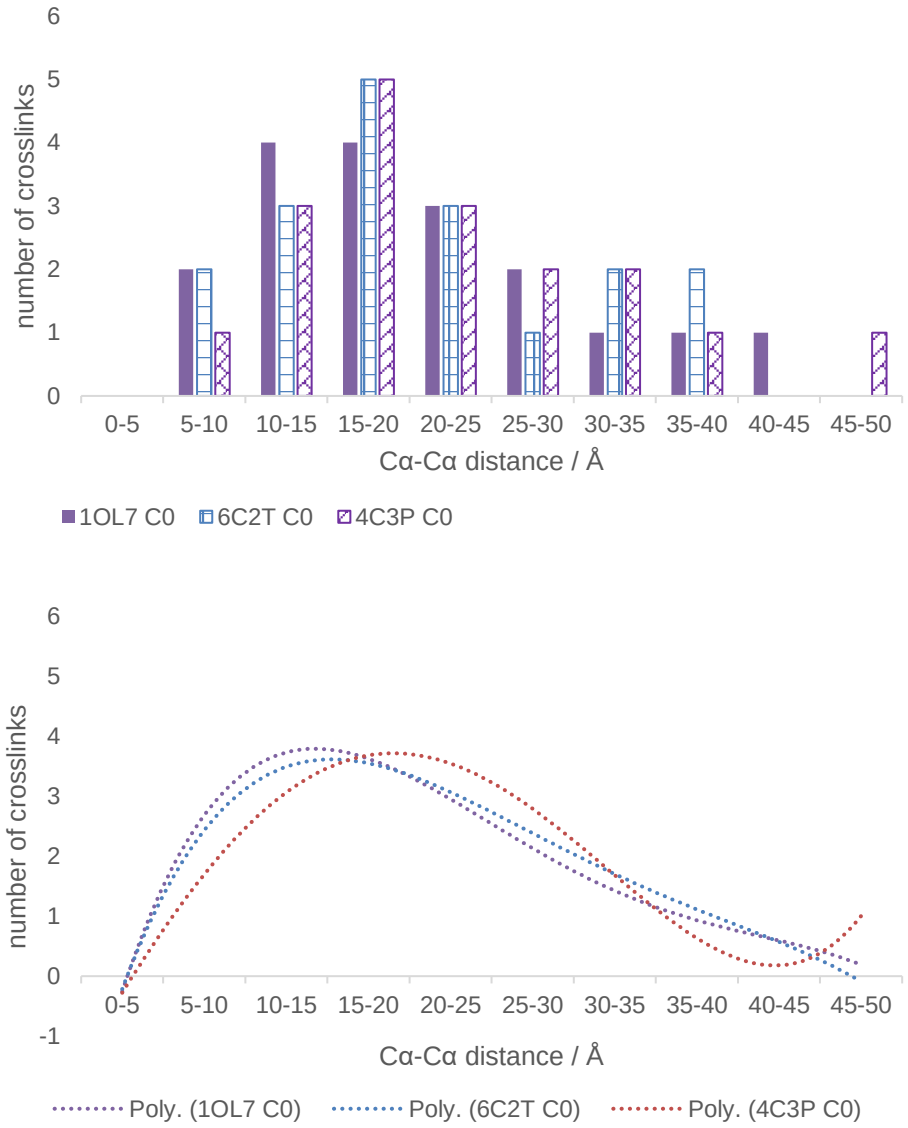


Figure 3-15 Histogram of normalised Cα-Cα distances found for AurA Apo crosslinks annotated on C0 cluster MD structures

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM.

3.4.1.3.3 Annotated output models

The clustered models with the three largest amount of frames for each of the input structures was annotated using SASD crosslinks and scored using XLM tools (Table 3-12), surprisingly showing 6C2T models scoring best.

Table 3-9 Scoring of clustered output models from AurA Apo MD using normalised selected crosslinks and monolinks

All input and output clustered structures subject to XLM tools analysis using AurA Apo n>1 and normalisation of >50 %. MoDs is the evaluation by monolink depth (average -59.1±3.7), the larger the better. MNXL is the evaluation by crosslink SASD (average 0.24±0.08), again the larger the better. XLMO is the combined score, the larger the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off.

PDB	**1OL7**			**4C3P**			**6C2T**		
Cluster	C0	C1	C2	C0	C1	C2	C0	C1	C2
MoDS	-60.97	-63.52	-61.37	-61.87	-59.4	-59.49	-54.27	-54.27	-54.93
MNXL	0.31	0.16	0.19	0.26	0.36	0.28	0.32	0.29	0.29
XLMO	0.25	-1.56	-0.87	-0.39	0.91	0.27	1.62	1.33	1.26
Crosslinks over	1	2	2	1	1	1	1	1	1
Place by XLMO	6th	9th	8th	7th	4th	5th	1st	2nd	3rd
Predicted CCS / nm^2	22.92	22.7	22.92	23.53	23.99	24.22	24.67	22.7	24.16
Conformer	II	II	II	III	III	III	III	II	III

3.4.1.3.3.1 SASD annotated 1OL7 model

The 1OL7 C0 clustered model annotated using SASD crosslinks generated using XLM tools scored well and strongly resembled the original crystallographic structure with a stabilised CTE bound to helix E and the NTE bound into the F-pocket (Figure 3-17).

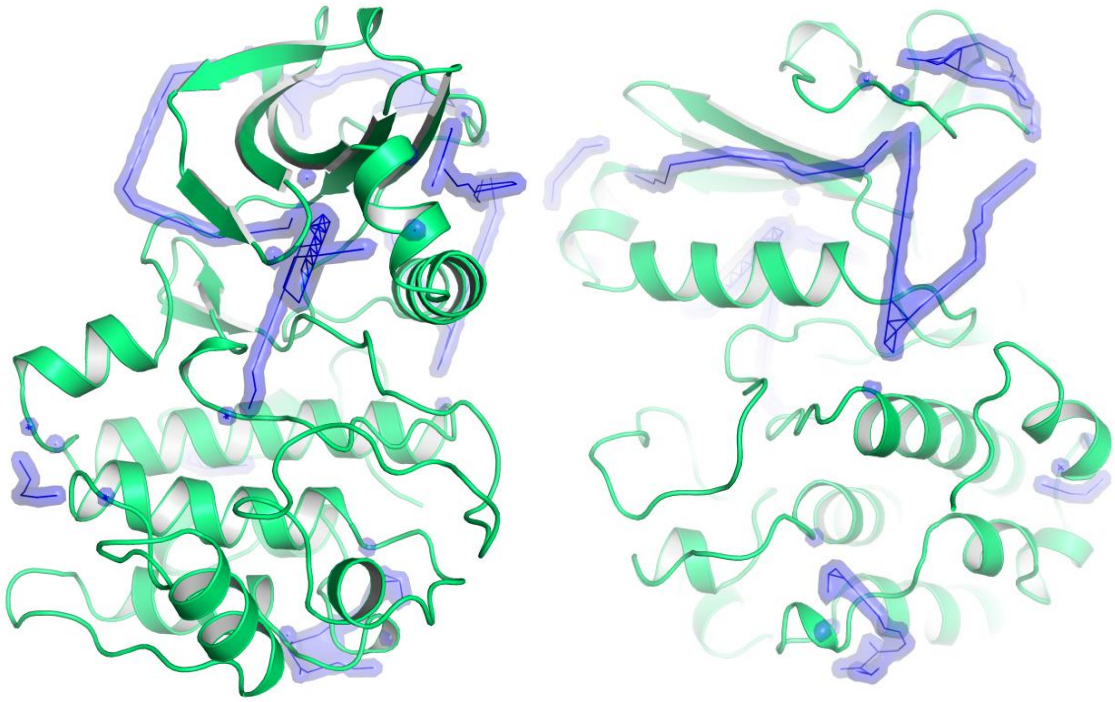


Figure 3-16 1OL7 C0 model annotated with SASD

3.4.1.3.3.2 SASD annotated 6C2T model

The 6C2T C0 clustered model annotated using SASD crosslinks generated using XLM tools scored well and, despite the fully exposed activation loop, was stable when compared to the 4C3P MD simulation. Interesting there was some contact between the NTE and CTE (Figure 3-18).

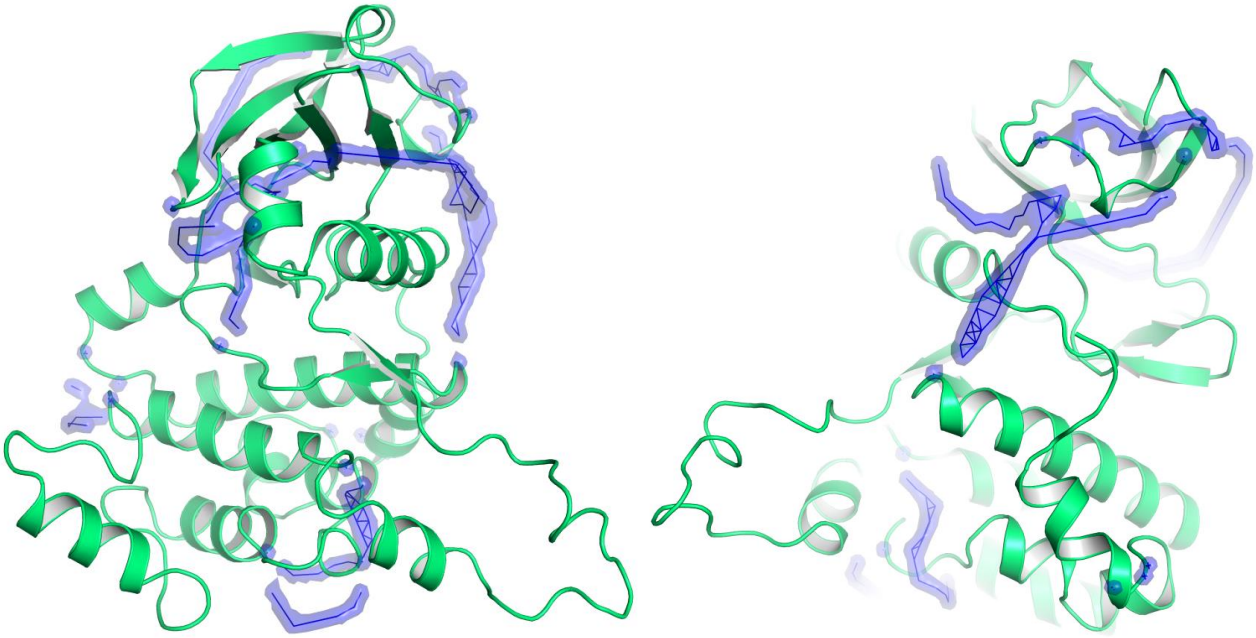


Figure 3-17 6C2T C0 model annotated with SASD

3.4.1.3.3.3 SASD annotated 4CEP model

The 4C3P C0 clustered model annotated using SASD crosslinks generated using XLM tools scored adequately despite the complete collapse of its N-lobe on the C-lobe and breaking of the αC-helix integrity (Figure 3-19).

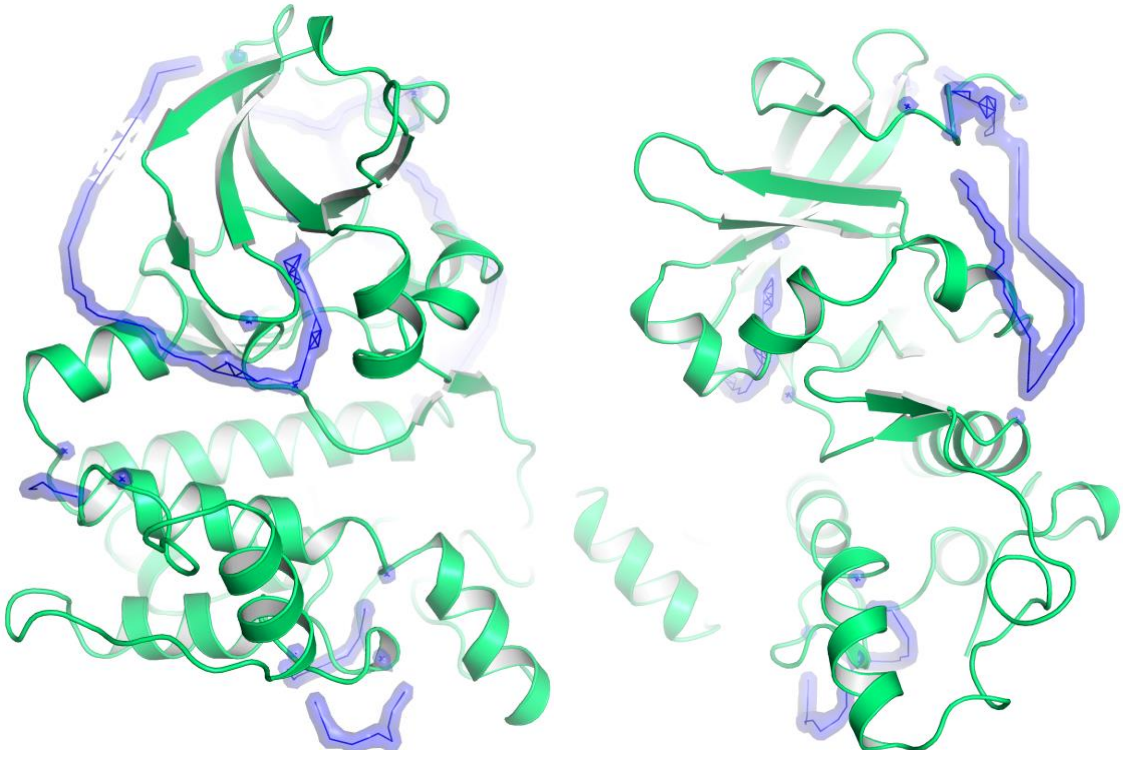


Figure 3-18 4CEP C0 model annotated with SASD

3.4.1.3.4 Using crosslinks and monolinks with more than a single CSM and reactivity normalised value of >50 % to evaluate the clustered structures

The use of the reactivity normalised values to attempt to remove crosslinks and monolinks that have little or no structural value to aid discrimination between the clustered output structures gave a similar result to the previous dataset. We see that XLM tools scores only a single crosslink of the nine used as over long, suggesting that we at resolution limit for this dataset and model set.

3.4.1.4 Structures taken forward for evaluation of the AurA+TXP2 complex

As has previously been shown by other in-solution methods, the interaction of TPX2 with AurA alters equilibrium of activation loop conformers by stabilising the catalytic active conformers. In Chapter 2, we also saw that TPX2 displaces the position of the AurA NTE, which switches from being bound to the N-lobe to being bound to TPX2. We wanted to assess the stability, using MD simulation, and evaluate the generated conformer using XLM tools, SASD-based scoring of the crosslinks, and depth-based scoring of the monolinks. Although SLD evaluation of the input model generated histograms of the correct shape, SASD evaluation showed that they possessed too many long crosslinks to give the expected distribution. However, using the cluster generated from the most frames for each simulation the crosslink distributions presented as expected. Remarkably, XLM tools scoring of the C0 clusters showed barely any discrimination between them, whether using crosslinks and monolinks with a single CSM/MSM or the normalisation cut-off. This is striking, as the 6C2T (Structured subdomain out, dis-continuous helix, DFG-inactive) model and the 4C3P model (Structured subdomain out, continuous helix, DFG-inactive) scored as well as the canonically active model (PBD code: 1OL7). This suggests that other in-solution methods are required to confirm or refute the presence of these novel monomeric structures in the in-solution ensemble of AurA Apo. If it were not for the Tomlinson *et al* ESI-MS native experiment values suggesting that these models fit a third, minor conformer of AurA Apo (Section 3.2.4.2, Table 3-1) we would discount their authenticity solely on the basis of their novel structure.[308] Pleasingly, the C0 cluster for the canonically active model (1OL7) matches the experimentally found CCS value for the major conformer. As all three C0 clusters are generated from frames at the end of simulations, and both structured subdomain-out models (6C2T and 4C3P) are equally scored, they were both taken forward.

3.4.2 Evaluation of MD simulated conformers of AurA+TPX2 using XL-MS dataset with IMS-MS

3.4.2.1 Preparation of MD simulations of high scoring Aurora-A 122-403

Literature data consistently reports that the binding of TPX2 alters the conformational ensemble of AurA, by stabilising active conformers and disfavouring inactive structures. It was recently shown that the activation of AurA by TPX2 is driven by a population shift from a DFG-out to DFG-in state, and not by triggering a conformational change of AurA.[156] We therefore decided to adapt the clustered structures representing the largest fraction of the active AurA Apo conformers with one notable additional cluster that was produced as part of the Apo assessment, and other AurA+TPX2 structures found in the PDB, if any could be found that differ significantly from previously modelled structures. Due to 4C3P undergoing nucleotide collapse, we included a cluster generated from frames prior to the collapse. We chose to use the same MD simulation protocol as that used in the AurA Apo. As before, the aim was to simulate each conformer for the construct used in the crosslinking reaction. Hence each conformer used the AurA Apo construct, 119-403 C290A C393A pT287/pT288 and TPX2 1-43 construct, both containing the vestigial residues of the purification affinity tags.

The opensource structure prediction platform “ColabFold”[245] that makes use of the neural network-based AlphaFold2[244] was also used to generated AurA+TPX2 conformers.

We repeated the process of using XLM tools to evaluate a range of hypothetical, generated AurA+TPX2 conformers and used IMPACT to approximate their CCS value, in order to compare against previously reported experimental data.[255] Two relevant structures were then subject to MD simulation to generate energy stabilised structures, and the XLM tools and IMPACT evaluation process repeated.

3.4.2.2 Generation of input structures

3.4.2.2.1 Summary

Despite having generated AurA Apo validated structures, we were concerned that not including a range of AurA+TPX2 structures would not adequately substantiate these generated structures. We saw with AurA Apo that scoring a selection of similar structures could not confidently differentiate between them if all of them had the same number of in-tolerance crosslinks. For this, eight AurA+TPX2 structures were constructed to represent a range of potential conformations, aiming for variation mainly of the AurA activation loop (their scores and associated data are given in Table 3-13). Clustered structures from AurA Apo were adapted directly with the manual addition of TPX2. As with AurA, selected crystal structures were manually adapted to match the construct used in the crosslinking reaction. In addition, AurA+TPX2 structures were generated using ColabFold. It should be noted that PDB structure 4C3P was not generated manually again, as it was included in the AurA Apo MD clustered structures.

Table 3-10 Summary of input AurA+TPX2 structures constructed for SASD evaluation

Table shows the range of input AurA+TPX2 structures constructed for SASD evaluation.

PDB	**3UOK**	**(CF A)**	**(CF B)**	**6VPJ**	**1OL7**	**4C3P**	**4C3P**	**6C2T**
Parent model	3UOK apo input	ColabFold	ColabFold	6VPJ PDB	1OL7 C0	4C3P C0	4C3P C7	6C2T C0
224-287 Cα-Cα / Å	16.4	36.5	36.8	37	37.5	44.6	51.2	51.9
172-308-324 Angle / °	85.9	85.3	84.7	81.4	78.7	59.3	82.5	88.2
Rspine (275 + 254+ 185 + 196)	Broken, Phe275 and His254 flipped	Formed	Formed	Broken, Phe275 flipped	Formed	Collapsed α-Helix	Formed	Partially formed, Phe275 flipped
Salt bridge (162 + 181)	Broken, Phe275 bisects residues	Formed	Formed	Broken, Lys 162 in nt pocket	Formed	Formed with Gln177	Formed	Broken, Glu181 out of place

3.4.2.2.2 ColabFold generation of AurA+TPX2

As we were constructing the AurA Apo models, we were concerned that evaluating models where the structural variables were not controlled would lead to misleading scoring. For AurA Apo, this meant using NTE and CTE structures that were identical across the model set, since we were testing the difference that activation loop position would make to the scoring. For this purpose, we envisioned a neutral TPX2 and NTE structure for the model

set used in MD simulation. However, generating a neutral TPX2 structure is a complex task, as TPX2 interacts directly with the activation loop in the W-pocket.

To this end, the ColabFold webserver was used to generate predicted structures using the AurA and TPX2 constructs that matched those used in the crosslinking experiments. The ‘use_amber’ option was not selected, to save computational resources. This generated structures that were not energy minimised, and the bonding angles and valency of certain residues had to be manually adjusted. The server lacks the capacity to include PTMs, and so the phosphorylations had to be added manually, to ensure all AurA+TPX2 structures had the same sequence and modification, and matched the protein used in the crosslinking reaction.

Each ColabFold run generates five models of the inputted sequence, each with a confidence score with the predicted structure. These scores were ignored, as the two models that were selected represented two different structures, both with the NTE of AurA and N-terminus of TPX2 in-solvent. The ColabFold models all closely matched the AurA structure found in 1OL5, apart from the unmodelled parts of the crystal structure. Model CF A showed the CTE of AurA interacting with TPX2, and model CF B placed it into the solvent (Figure 3-20).

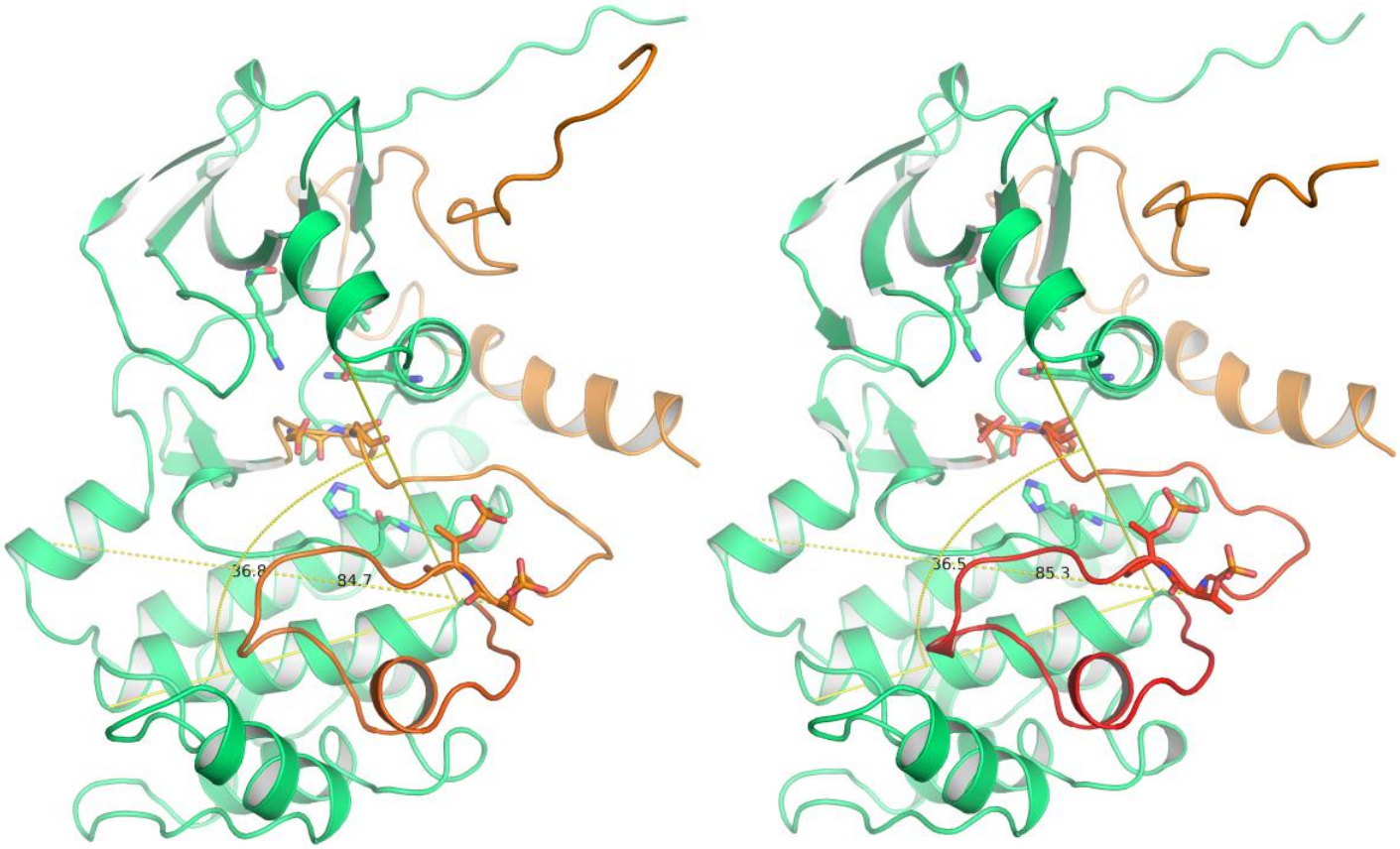


Figure 3-19 Input model AurA+TPX2

Input model AurA+TPX2, generated using the ColabFold webserver. The activation loop (red/orange) is held in the active conformer Left) Model CF A Right) Model CF B. Both models present with a Lys224-Thr287 Cα-Cα distance of ≈36 Å and an angle formed between the Cα's of residues 172-308-324 of 85°.

3.4.2.2.3 AurA Apo input models

Each of the four clusters of models were separately aligned to AurA residues 119-191 of 1OL5 in PyMOL, and then combined with the TPX2 1-43 from the Model CF B (Figure 3-20). As previously stated, the use of the ColabFold TPX2-bound AurA structure was to ensure that the minor changes that would arise from separately generating a homology model were avoided. As the AurA Apo places the NTE of AurA into the F-pocket, which TPX2 also binds into, the NTE was remodelled, as AurA Met121 clashes with TPX2 Phe19. This ensured that the placement of TPX2 matches that found in the crystal structure; the conformation of TPX2 relative to the Y- and F-pockets was kept constant.

3.4.2.2.3.1 4C3P models

Initially, we had selected the three clusters from each AurA Apo simulation; however, we included cluster C7 from 4C3P, as C0 from 4C3P was found to have a collapsed nucleotide pocket (Figure 3-21). This meant that the conformation of the activation loop required adjustment of the TPX2 alignment. It is striking that C7 4C3P had both a formed R-spine and salt bridge and was scored second of all the input and output AurA Apo models, the first being C0 6C2T. Despite C7 4C3P bearing a striking resemblance to crystallographic structures, it is of questionable value for the continued analysis of AurA+TPX2.

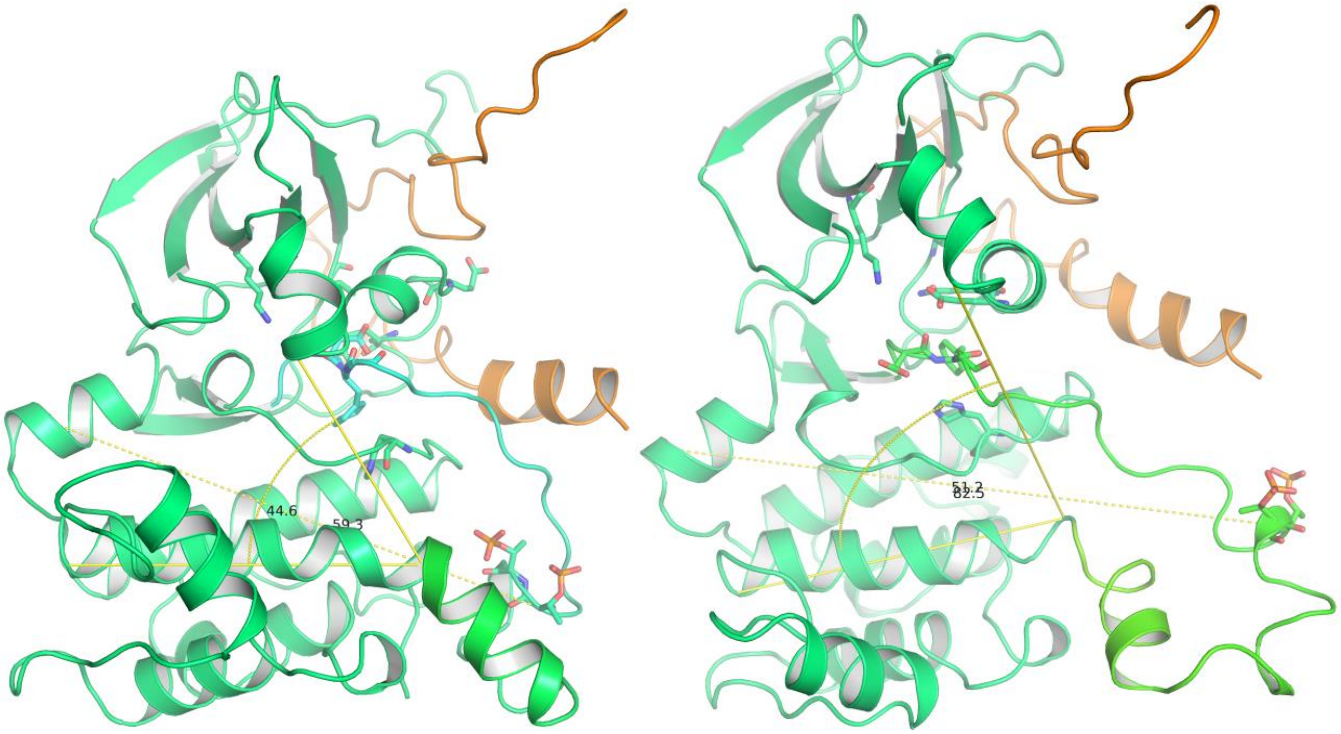


Figure 3-20 Clustered Input model 4C3P AurA+TPX2

Left) C0 4C3P has collapsed alpha helix; despite it being clustered from the highest number of frames from the MD simulation, it bears little resemblance to monomeric conformations found in AurA crystal structures. Right) C7 4C3P generated from the largest number of frames where the nucleotide pocket has not been collapsed. This strongly resembles the initial dimeric crystal structure conformation. Strikingly, the Lys224-Thr287 Cα-Cα distance differs between C0 (44.6 Å) and C7 (51.2 Å) as does the an angle formed between the Cα's of residues 172-308-324 for C0 (59.3°) and C7 (62.5°).

3.4.2.2.3.2 6C2T model

The construction of the C0 AurA+TPX2 was straightforward, with the activation loop poised to support the W-pocket site binding residues of TPX2 (Figure 3-22, left).

3.4.2.2.3.3 1OL7 model

The construction of the C0 AurA+TPX2 was straightforward, as it matched the canonical structure (Figure 3-22, right).

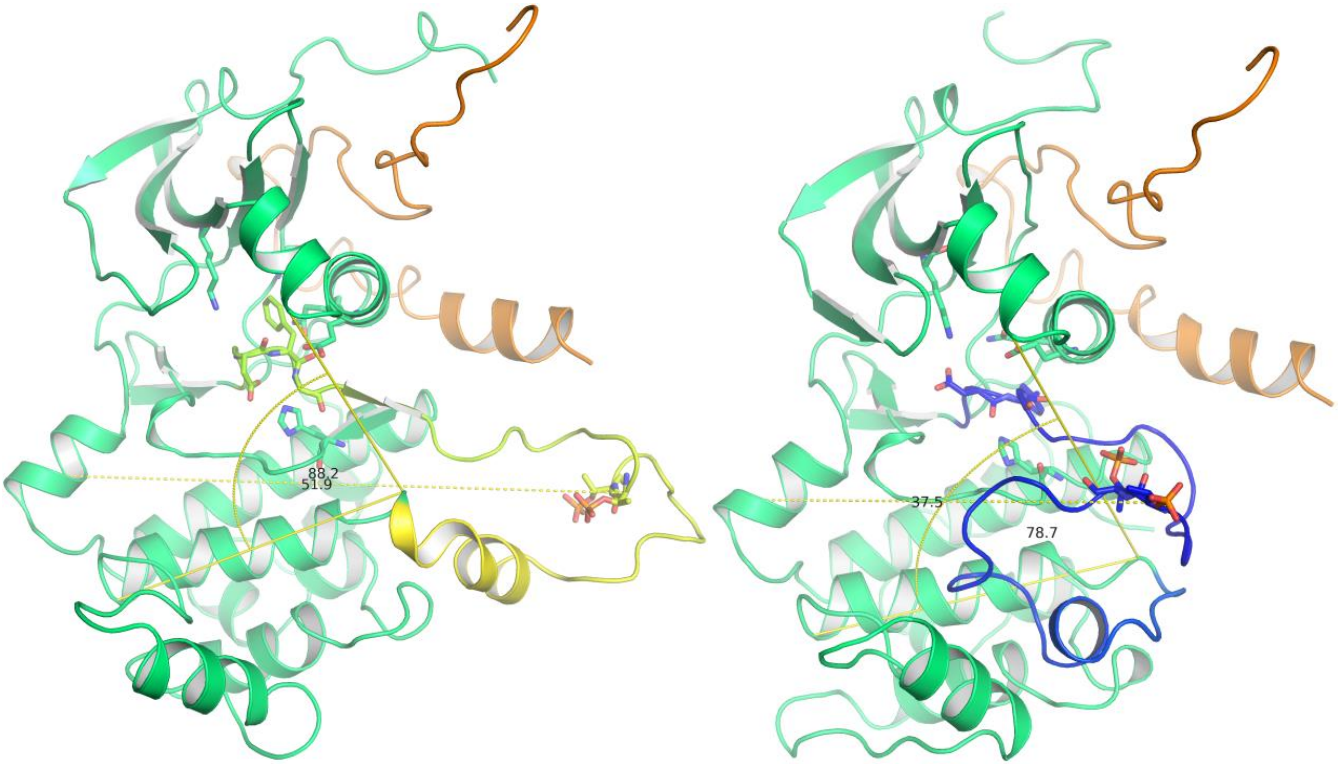


Figure 3-21 Clustered Input models 6C2T and 1OL7 AurA+TPX2

Left) C0 62CT clustered input model, The activation loop (yellow) is held open conformation with a Lys224-Thr287 Cα-Cα distance of **51.9** Å and an angle formed between the Cα's of residues 172-308-324 of 88.2°
Right) C0 1OL7 clustered input model, the activation loop (blue) is held active conformation with a Lys224-Thr287 Cα-Cα distance of **37.5** Å and an angle formed between the Cα's of residues 172-308-324 of 78.7°.

3.4.2.2.4 AurA+TPX2 input models

3.4.2.2.4.1 3UOK Model

As previously mentioned, we were concerned with the exclusion of models that we expected to scored poorly. Hence an inactive, MLN-like structure was manually constructed using the initial AurA apo model (Figure 3-23). A model constructed from 3UOK, that represented TPX2 interacting with a DFG-out, alpha-helix contact structure, was included to ensure the widest range of structures. Despite the removal the activation loop, the W-pocket is still available for TPX2 to be placed into.

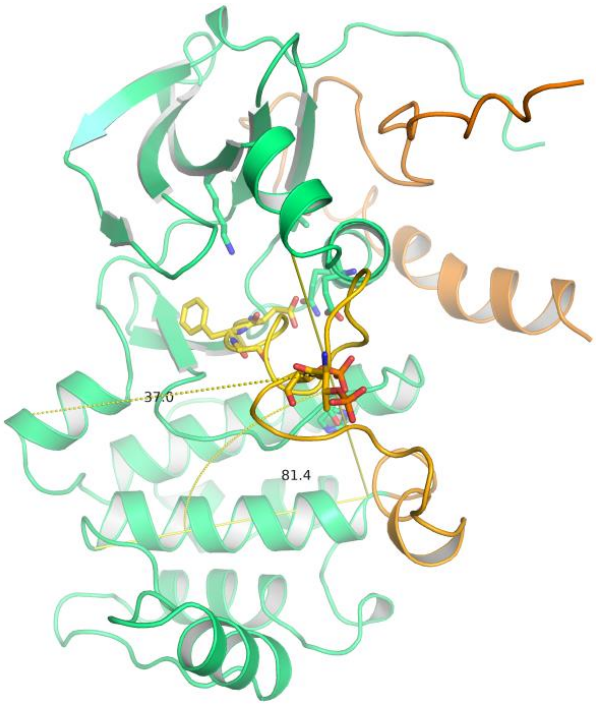


Figure 3-22 Input models 3UOK AurA+TPX2

3.4.2.2.4.2 6VPJ model

Lim *et al* reported a structure involving the chimeric fusion protein TPX2 7-20 C-terminally fused to de-phosphorylated AurA 116-389 C247V + C319V in complex with AMP-PNP.[166] This structure, 6VPJ, has a face-to-face, domain swapped dimer, the opposite orientation to 4C3P (Figure 3-24). The TPX2 fused to the AurA in the biological assembly is bound to a third AurA in the asymmetric unit. The TPX2 pose, as it is bound to AurA in the asymmetric unit, matches that found in the other AurA+TPX2 crystal structure, PBD code 1OL5.

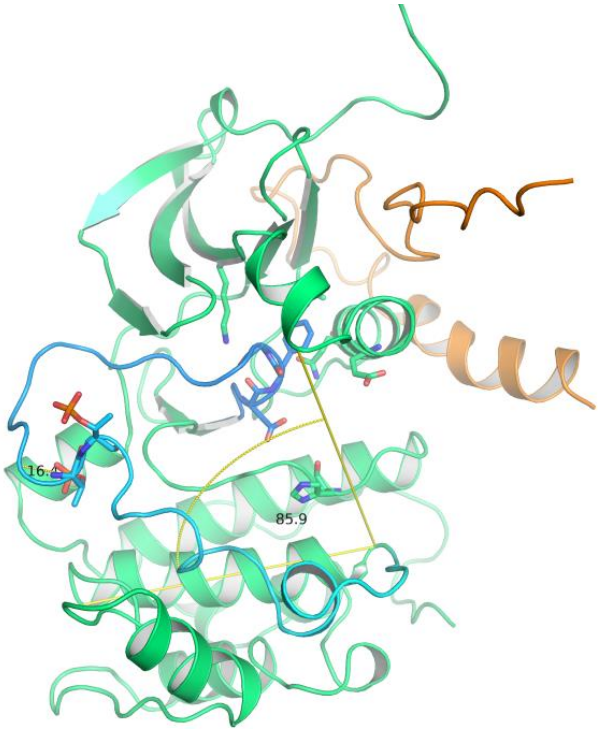


Figure 3-23 Manual 6VPJ input model AurA+TPX2

The activation loop (light blue) is held inactive conformation with a Lys224-Thr287 Cα-Cα distance of 16.4 Å and an angle formed between the Cα's of residues 172-308-324 of 85.9°.

3.4.2.3 Evaluation of input structures using AurA+TPX2 crosslinks and monolinks

3.4.2.3.1 Summary

Once compiled, the structures were evaluated using XLM tools, using both all crosslinks and monolinks with more than a single crosslink, and additionally using the reactivity cut-off list. The scores generated showed that it was ill-advised to evaluate non-stabilised models. The C0 6C2T model scored highest using both CSM>1, and CSM>1 with >50 % normalisation list. This model was shown to have fifteen of the nineteen crosslinks within the SASD tolerance, and six of nine for the normalised dataset. The models that closely resemble the canonically active AurA+TPX2 structure (ColabFold A, ColabFold B and C0 1OL7) scored poorly relative to the inactive or control structures (3UOK Apo input and 6VPJ PDB), (Figure 3-25). As the standard deviation of the MoDs scores was very low, the majority of the structural discrimination was caused by the MNXL scores. Unfortunately, five of the nineteen crosslinks involved the NTE, which was previously standardised across the model, and was for this model. XLM tools was thus attempting to evaluate structures where disordered regions had either been standardised, to allow for the non-NTE crosslinks being the variable scored, or had been stabilised so that the average lifetime position of the NTE could be scored.

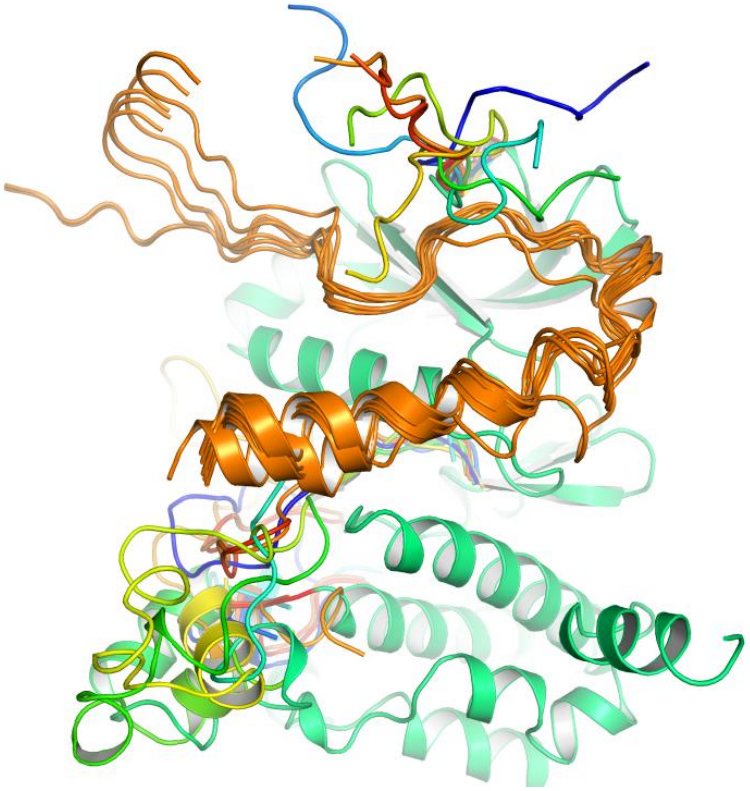


Figure 3-24 Overlaid input models of AurA+TPX2

Overlaid TPX2 models showing the differing NTE positioning TPX2 and activation loop on the ColabFold A model. 6VPJ (yellow), 3UOK (light blue) C0 4C3P (cyan), 1OL7 (royal blue), 6C2T (split pea green), C7 4C3P (grass).

The monolink standard deviation was not large enough to confidently distinguish the structures. It is likely that even standardised structures would generate a similar range of results, and that generating stabilised structures to then re-test with the XLM-tools dataset is the sensible method.

Additionally, the range of average predicted CCS is larger than the largest experimentally found conformer, adding credence to the idea that the generated input models do not accurately represent the range of possible conformers for AurA+TPX2.

Furthermore, XLM-tools could be generating an average structure of two conformers. Especially as we successfully found two differing conformers of AurA Apo.

The combined scores of the input models for AurA+TPX2 is like that of the C series, where there is little differentiation by scoring of the models. 6C2T and 6VPJ do score well, but do not have significantly more distance restraints in tolerance than other models. This is potentially due to the lack of energy minimisation of the disordered or dynamic regions of the structure. This is supported by residue TPX2 G1 being very promiscuous, involved in eight of the nineteen crosslinks. Given that we are evaluating non-stabilised structures with dynamic IDRs, that dominate the number of crosslinks found, and also thus dominate the scoring, the low deviation between metrics means that we cannot confidently differentiate between this model set.

3.4.2.3.2 Structures taken forward for evaluation of the AurA+TPX2 complex

The 6C2T matched a structure generated as part of the AurA Apo evaluation. Furthermore, a canonically active AurA+TPX2 structure, seen in crystal structures, was also be taken forward. Inactive or MLN-like structures were omitted. These structures will be revisited if the structures generated from clustering the MD simulations fail to explain certain key crosslinks. Table 3-14 gives a summary of the structures scores and which structures were taken forward.

Of the AurA+TPX2 input structures generated, the ColabFold A structure was taken forward as it matches the crystal structure 1OL5. This is not the highest scoring 1OL5-like structure, but it features NTE and CTE in solvent, and will be a neutral input structure for the MD simulation.

Table 3-11 Summary of input AurA+TPX2 structures scores

All input structures subject to XLM-tools analysis using AurA+TPX2 n>1 and n>1 with > 50 % normalization. For MoDs, the evaluation by monolink depth, the larger the score the better. MNXL is the evaluation by crosslink SASD, again the larger the better. XLMO is the combined score, the larger the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off; total crosslinks were 19 and 11 respectively. The average MoDs for n>1 was -82.9±4.5, and the average MNXL was 0.12±0.21. For n>1 with > 50 % it was -46.9±3.3 and -0.10±0.15 respectively.

PDB		3UOK apo input	ColabFold A	ColabFold B	6VPJ PDB	1OL7 C0	4C3P C0	4C3P C7	6C2T C0
CSM>1	MoDS	-83.82	-84.15	-82.71	-78.8	-83.81	-92.86	-76.48	-80.84
	MNXL	0.11	0.1	-0.07	-0.08	-0.24	-0.49	-0.35	0.08
	XLMO	0.54	0.48	0.16	0.67	-0.47	-2.45	0.2	0.87
	Crosslinks out of tolerance	6	4	5	5	6	8	7	4
	Place by XLMO	3rd	4th	5th	2nd	7th	8th	6th	1st
CSM>1 and >50 %	MoDS	-48.87	-49.44	-48.40	-45.52	-48.18	-51.10	-42.49	-41.10
	MNXL	0.08	0.03	-0.16	-0.17	-0.24	-0.24	-0.27	0.14
	XLMO	0.40	0.06	-0.54	-0.01	-0.83	-1.43	0.15	2.19
	Crosslinks out of tolerance	4	4	5	5	6	5	6	3
	Place by XLMO	2nd	4th	6th	5th	7th	8th	3rd	1st
Predicted CCS / nm^2		26.53	27.22	26.85	27.76	26.38	26.71	28.07	27.94
Conformer		III	III	III	IV+	III	III	IV+	IV+
Taken forward for MD		No	Yes	No	No	No	No	No	Yes

3.4.2.4 MD simulations of active and activation loop out, discontinuous helix conformations of the Aurora A 122-403 + TPX2 1-43 model

3.4.2.4.1 Evaluation of simulation and generation of AurA+TPX2 clustered output structures

Before we evaluated the output structures of XLM tools using the AurA+TPX2 crosslinked datasets, we examined the stable structures generated to evaluate if the TPX2 dynamics were observed. We observed a reduction in predicted CCS for both models over the course of the simulations, suggesting that they both underwent compaction as part of their stabilisation. As with AurA Apo simulations, a clustering algorithm was used to generate stable structures from each of the simulations. We also compared the changes to the structures with the addition of TPX2, to assess the impact that the addition of TPX2 had on the conformer (Table 3-15).

The clusters comprised of the three largest numbers of frames were taken forward for analysis by XLM tools, and comparison of their predicted CCS to the experimental values generated by Tomlinson *et al*.[308]

Table 3-12 Summary of Key output metrics for the two AurA+TPX2 simulations

Key output metrics for the two AurA+TPX2 simulations.

Simulation	Average temp / K	Average RSMD / Å	Time / µs	Input predicted CCS / nm²	Average predicted CCS / nm²	Stdev of CCS
6C2T C0	303.40	4.96	0.482	27.94	26.13	1.71
CF A	303.40	5.37	0.491	27.22	24.84	1.99

3.4.2.4.2 Evaluation of the stability of an active AurA KD conformer in the presence and absence of TPX2 1-43

3.4.2.4.2.1 Summary

The ColabFold model A, CF A, closely matched the conformation of the crystal structure 1OL5; unsurprising, as it is the canonically active AurA structure bound to a stabilising IDR, we see little difference in the D Helix to activation loop distance and the hinge angle (37.3 Å and 85.4^{O}). The cluster comprised of the largest number of frames only shows subtle conformational changes for AurA, but more dramatic changes for TPX2. These changes could reasonably be due to energy minimisations from the crystallographic pose (Figure 3-26). The AurA

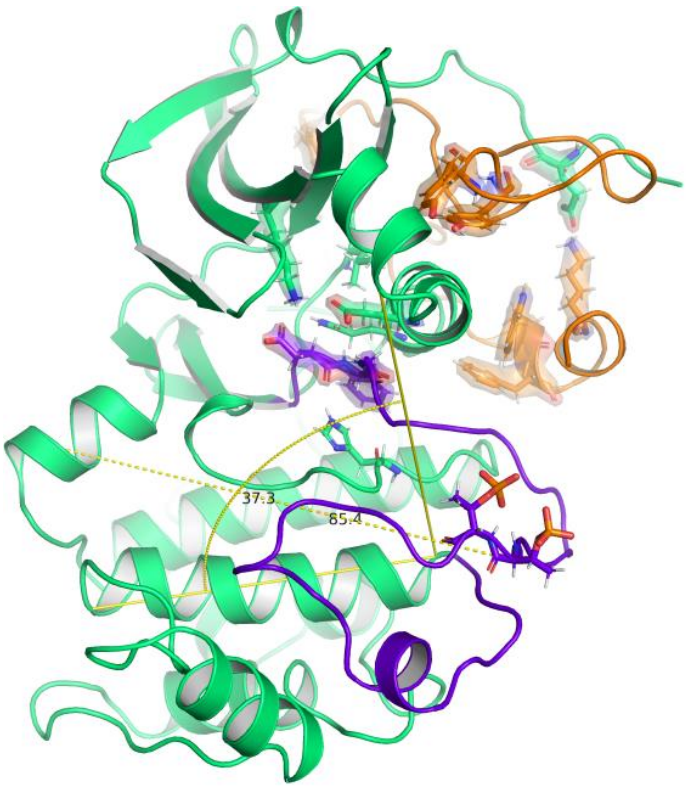


Figure 3-25 Cluster C0 output models of CF A AurA+TPX2

This MD simulated model of AurA bound to TPX2 (orange) presents the canonical model of active AurA with a novel interaction between the NTE of AurA and TPX2. Phosphorylation on the activation loop (purple) in an active conformation with a Lys224-Thr287 Cα-Cα distance of 37.3 Å and an angle formed between the Cα's of residues 172-308-324 of 85.4°.

NTE appears to bind into the cleft between the Y-pocket binding helix and activation loop binding helix of TPX2. The lack of TPX2 Lys38 crosslinks or monolinks seemed to be explained by this simulation as here Lys38 appears to hydrogen bond to a residue in the AurA NTE, namely Glu122. Comparing the pseudo b-factors of AurA in the active conformation (simulation 1OL7) for the AurA+TPX2 and AurA Apo simulation shows no significant differences. The MD simulations' trend towards stabilisation, and later conformational changes typically occur to increase the stability of the complex by increasing energetically favourable contacts.

3.4.2.4.2.2 MD data for active conformer

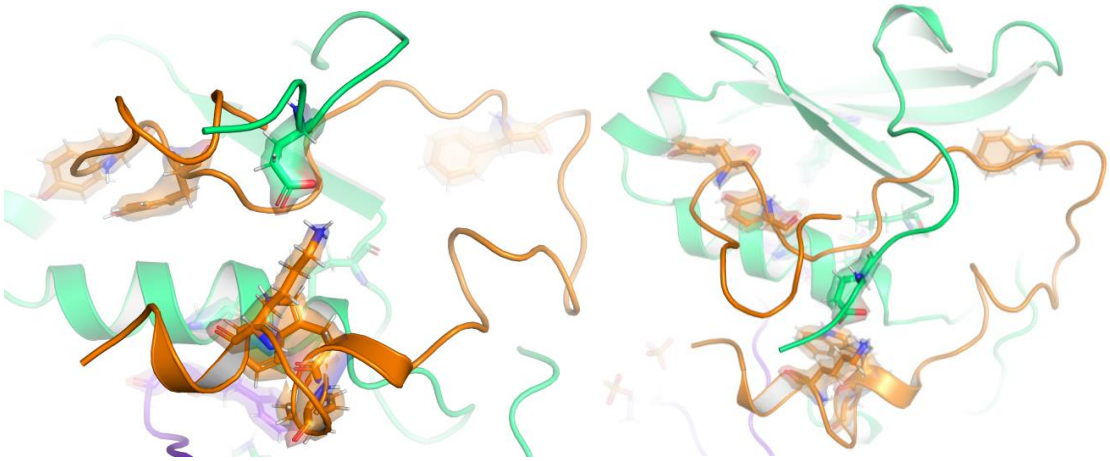


Figure 3-26 TPX2 position across clusters C0 and C1 output models of CF A AurA+TPX2

The left panel shows the C0 and right C1. The highlighted TPX2 residues are dynamic across the clusters output models of CF A AurA+TPX2. As these residues are critical to the AurA+TPX2 PPI its striking that this motion was seen.

In examining the pseudo b-factors for AurA, for both AurA Apo 1OL7 and the AurA+TPX2 C FA model, we see little difference between the two states. This is understandable given that both these models are illustrative of the canonical active conformation. The RMSD trace, (Figure 3-28), presents as a stabilised structure, as there is little variation in the RMSD after a stable conformer has been found, mid-way through the simulation (Figure 3-27).

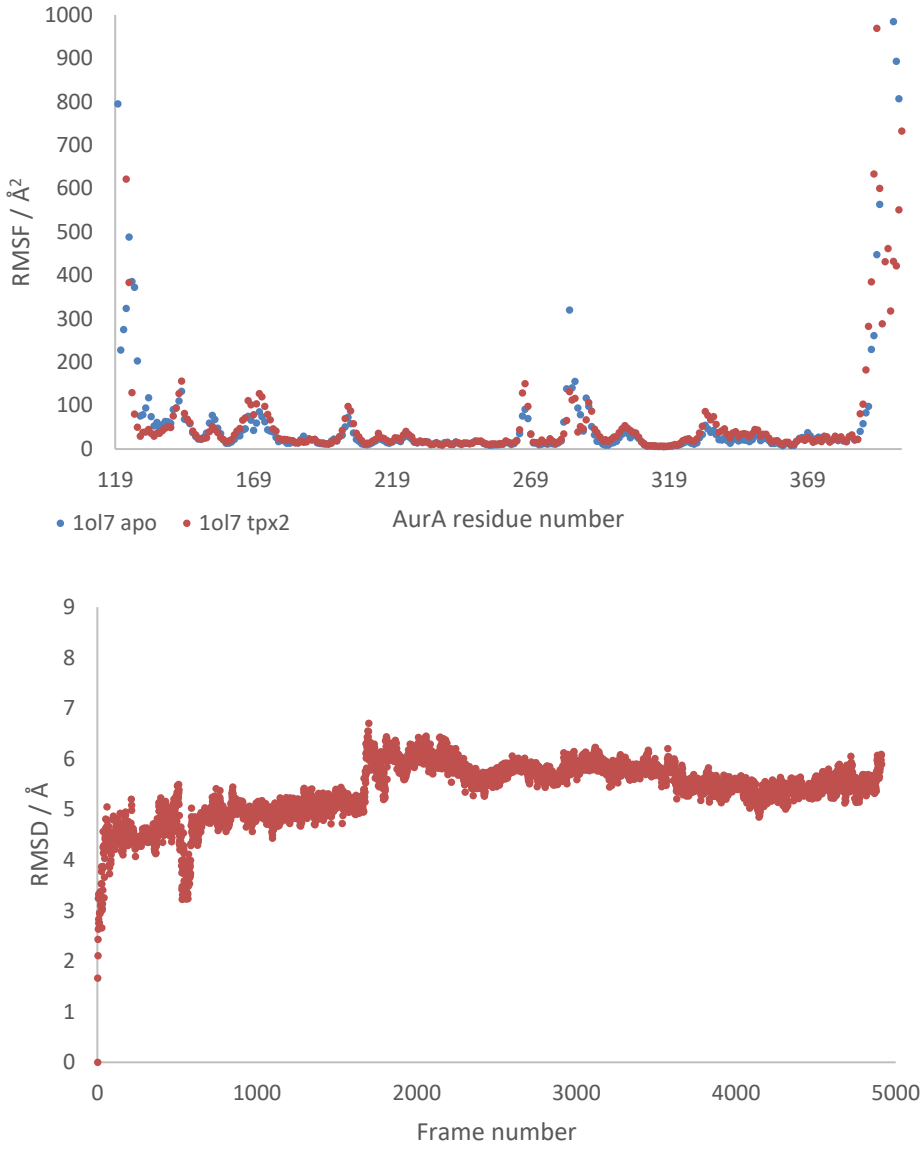


Figure 3-27 RSMD and pseudo b-factor trace from AurA+TPX2 MD length for AurA+TPX2 CF A model

The upper panel shows the pseudo b-factor (RMSF) values for both AurA Apo (1OL7 simulation, blue) and AurA+TPX2 (red). Lower) RSMD trace for the AurA+TPX2 CF A model simulation

3.4.2.4.3 Evaluation of the MD simulation stability of activation loop out, discontinuous helix, AurA KD conformer in the presence and absence of TPX2 1-43

3.4.2.4.3.1 Summary

The input structure was found with a Helix-D-activation loop distance of 51.9 Å and N-lobe/C-lobe hinge angle of 88.2º. However, the cluster found with the largest frames showed the Helix-D-activation loop distance shortening to 48.9 Å, but with an increase of the N-lobe/C-lobe hinge angle to 95.8º. The large angle seen was due to the extended hinging open supported by contacts between the lobes by the F-pocket.

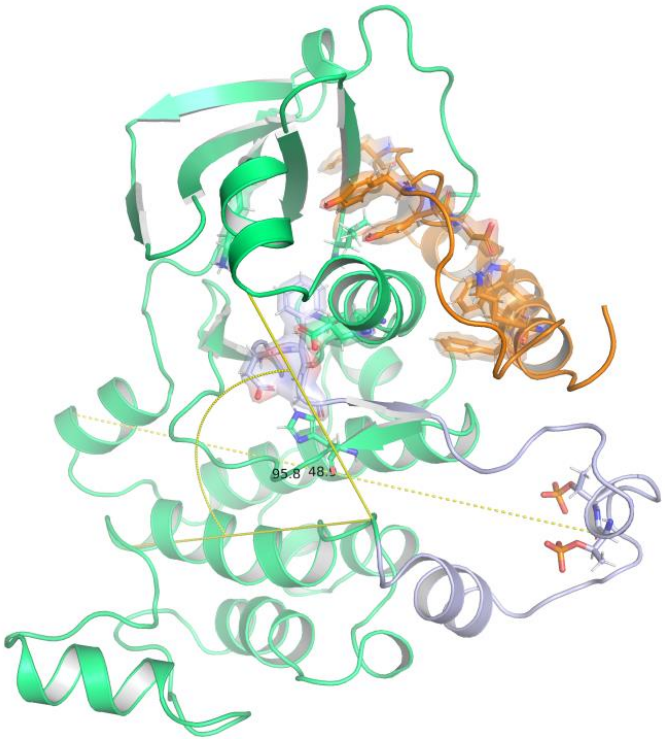


Figure 3-28 Cluster C0 output models of 6C2T AurA+TPX2

The C0 cluster of model CF A shows contact between the discontinuous helix of the activation loop of AurA and TPX2. Phosphorylation on the activation loop (silver) in solvent with a Lys224-Thr287 Cα-Cα distance of 48.2 Å and an angle formed between the Cα's of residues 172-308-324 of 95.8º.

For the simulation of the AurA+TPX2 model generated from the activation loop out, discontinuous helix, crystal structure, 6C2T scored well as its original AurA Apo model, (Figure 3-29). It was more stable with the addition of TPX2, as demonstrated by the reduction in residue fluctuations seen in the pseudo b-factor analysis, (Figure 3-15). For the simulation of AurA+TPX2 in its canonically active conformation, we saw no significant change in the pseudo b-factor fluctuations. However, TPX2 appeared to be more dynamic in this simulation than the 6C2T simulation.

In the 6C2T structure, it is noticeable that the C-lobe undergoes a large rotation around the hinge region to find a stable conformer. As the simulation progresses, it appears to be optimising packing of the structured subdomain of the activation loop, TPX2 W-pocket engagement, and the hinge rotation. Also in this simulation, we see a large change from the input structure (C0 cluster output from AURA Apo) appear, due to TPX2 supporting the

phosphorylated subdomain as a separated domain, with the structured subdomain of the activation loop packing into the C-lobe (Figure 3-30). This is shown by the pseudo b-factors being lower for the simulation with TPX2.

It appears that the AurA CTE is questing for a stable pose but does not find one during the simulation. This is to be expected given that this property is also seen in crystal structures.

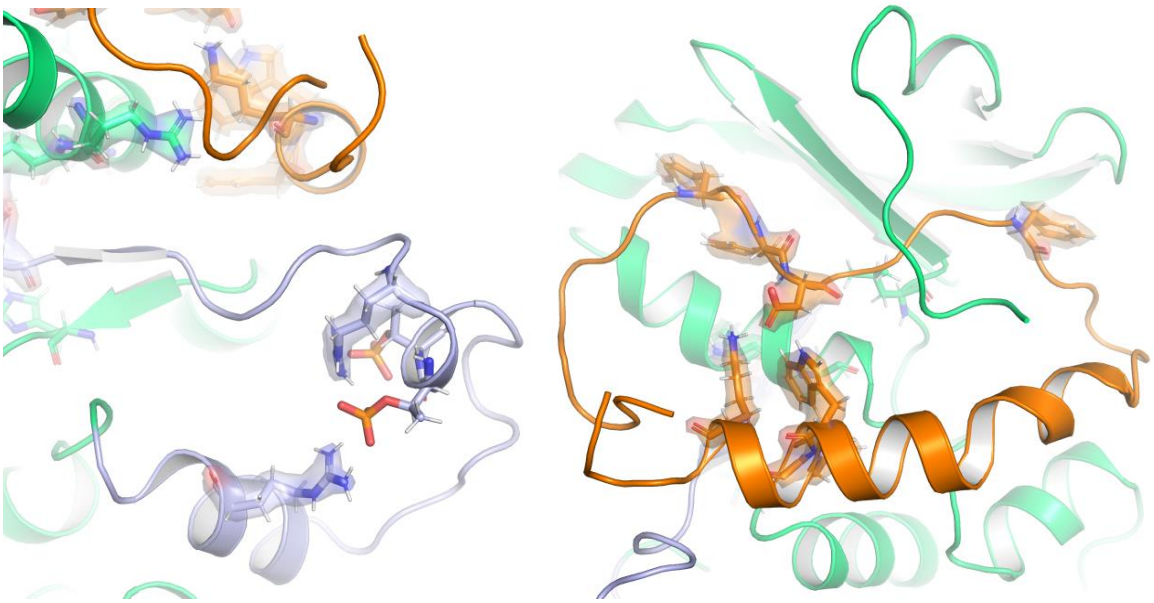


Figure 3-29 The Cluster C0 output models of 6C2T AurA+TPX2

Two alternatives views of the AurA+TPX2 interaction found in the C0 cluster. Left shows the phosphorylated residues of activation loop (silver) interacting with AurA Arg304. Right shows the AurA NTE bound into the cleft formed by TPX2 (orange).

3.4.2.4.3.2 MD data

Comparing the pseudo b-factor graphs of the AurA 6C2T conformer in the presence and absence of TPX2 suggests that TPX2 stabilises the novel structure, as we see a reduction in pseudo b-factor fluctuations for the activation loop and Helix G (Figure 3-31).

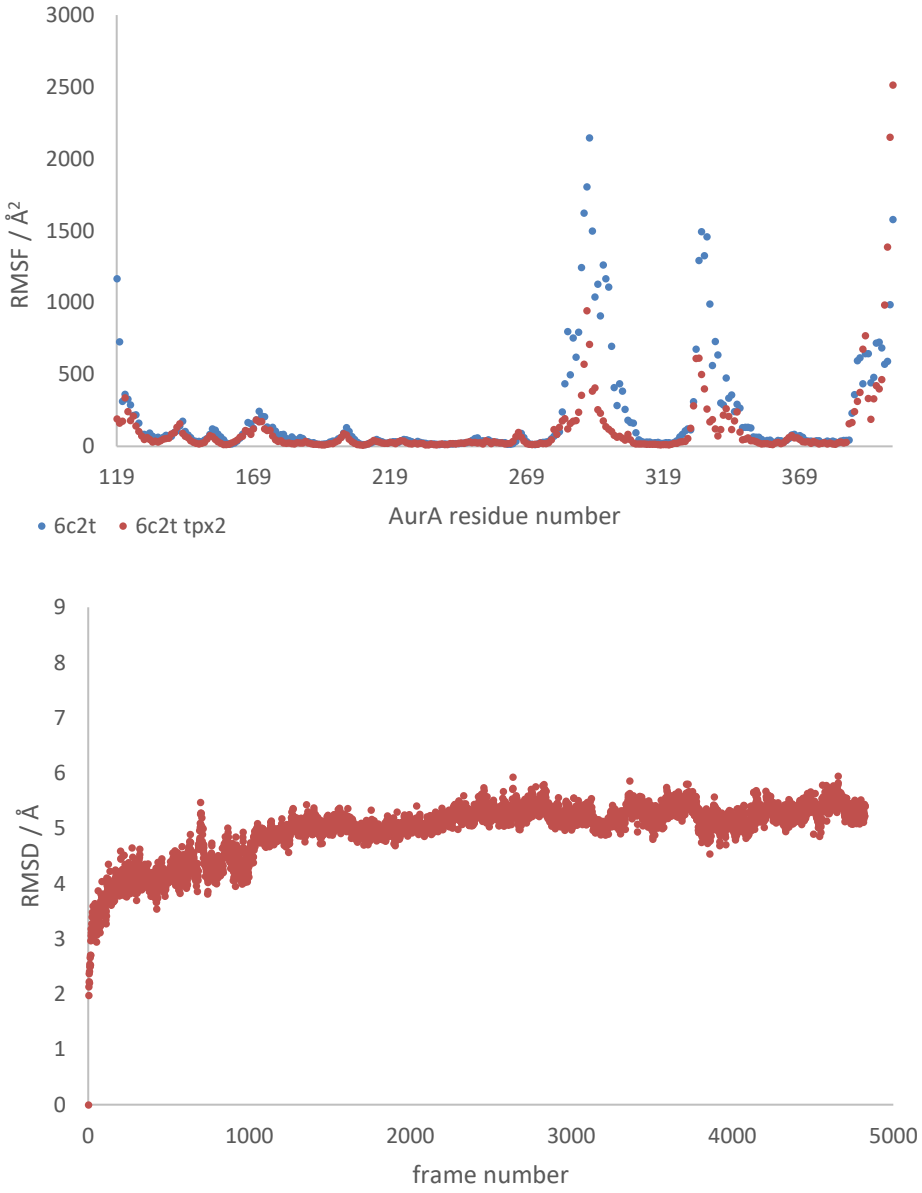


Figure 3-30 RSMD and pseudo b-factor trace from AurA+TPX2 MD length for AurA+TPX2 6C2T C0 model

The upper panel shows the pseudo b-factor (RMSF) values for both AurA Apo (6C2T simulation, blue) and AurA+TPX2 (red). The activation loop of AurA is significantly more with TPX2. Lower) RSMD trace for the AurA+TPX2 6C2T CF A model simulation

3.4.2.5 Evaluation of stable clustered Models by XLM tools and CCS prediction

3.4.2.5.1 Evaluation of output structures using n>1 crosslinks

Once compiled, the clustered structures with the most frames from the two simulations were used to generate histograms of the SASD (Figure 3-32). The clustered structures with the three largest percentages of the active AurA+TPX2 conformer and the activation loop open conformer simulations were taken forward for evaluation by XLM tools, and for comparison of predicted CCS values to those found in existing experimental data. Several different crosslink and monolink datasets were used to evaluate the models. Using all crosslinks and monolinks, on average, 35 % of the crosslinks were found to be longer than the Cα-Cα tolerance allows. Using all crosslinks and mono-links, an average 12.5 crosslinks matched (stdev 1.25) out of a possible 19. The average crosslinking (MNXL) score was -0.06, (stdev=0.18) and the average MoDS score was -69.8, (stdev=2.06). 6C2T C2 and CF A cluster C1 were the best scoring. The standard deviations across the five different conditions were very low, indicating that XLM tools is not successfully differentiating between the two models (Table 3-16). This is the case even when removing the AurA-TPX2 crosslinks, which will be due to the dynamic NTE of TPX2. This is seen in the simulations of the active conformer of AurA+TPX2. Each of the large clusters places the NTE of TPX2 interacting with a different part of AurA. As >90 % of AurA-AurA and >70 % of all crosslinks are within tolerance, the models generated convincingly accord with the experimental distance restraints and monolink coverage found. AurA Ser284-TPX2 Lys5 is within tolerance of both large clusters.

Table 3-13 Scoring of clustered output models from AurA+TPX2 MD

All output structures were subject to XLM tools analysis using AurA+TPX2 n>1 and n>1 with > 50 % normalization. For MoDs, the evaluation by monolink depth, the larger the better. MNXL is the evaluation by crosslink SASD, and again the larger the better. XLMO is the combined score, and the larger the score the better. 'Crosslinks over' is the number of crosslinks that exceed the 35 Å SASD cut-off; total crosslinks were 19 and 11 respectively. The average MoDs for n>1 was -83.7±1.1 and the average MNXL was -0.05±0.08. For n>1 with > 50 % is was -46.2±2.1 and -0.01±0.12 respectively.

PDB		**ColabFold A**			**6C2T**		
Cluster		C0	C1	C2	C0	C1	C2
CSM>1	MoDS	-83.7	-81.8	-83.7	-83.7	-85.4	-84.2
	MNXL	0.14	-0.12	-0.01	-0.03	0.15	-0.18
	XLMO	1.31	0.89	0.03	-0.1	-0.29	-1.84
	Crosslinks out of tolerance	4	5	5	3	3	4
	Place by XLMO	2nd	3rd	4th	5th	1st	6th
CSM>1 and >50 %	MoDS	-48.32	-48.51	-46.39	-43.51	-47.09	-43.1
	MNXL	0.04	-0.06	-0.05	-0.12	0.04	-0.17
	XLMO	0.29	-1.99	-0.01	0.53	1.28	-0.1
	Crosslinks out of tolerance	4	5	4	4	4	4
	Place by XLMO	3rd	6th	4th	2nd	1st	5th
Predicted CCS / nm^2		24.4	24.64	25.65	26.34	25.85	26.21
	Conformer	II-III	II-III	III	III	III	III

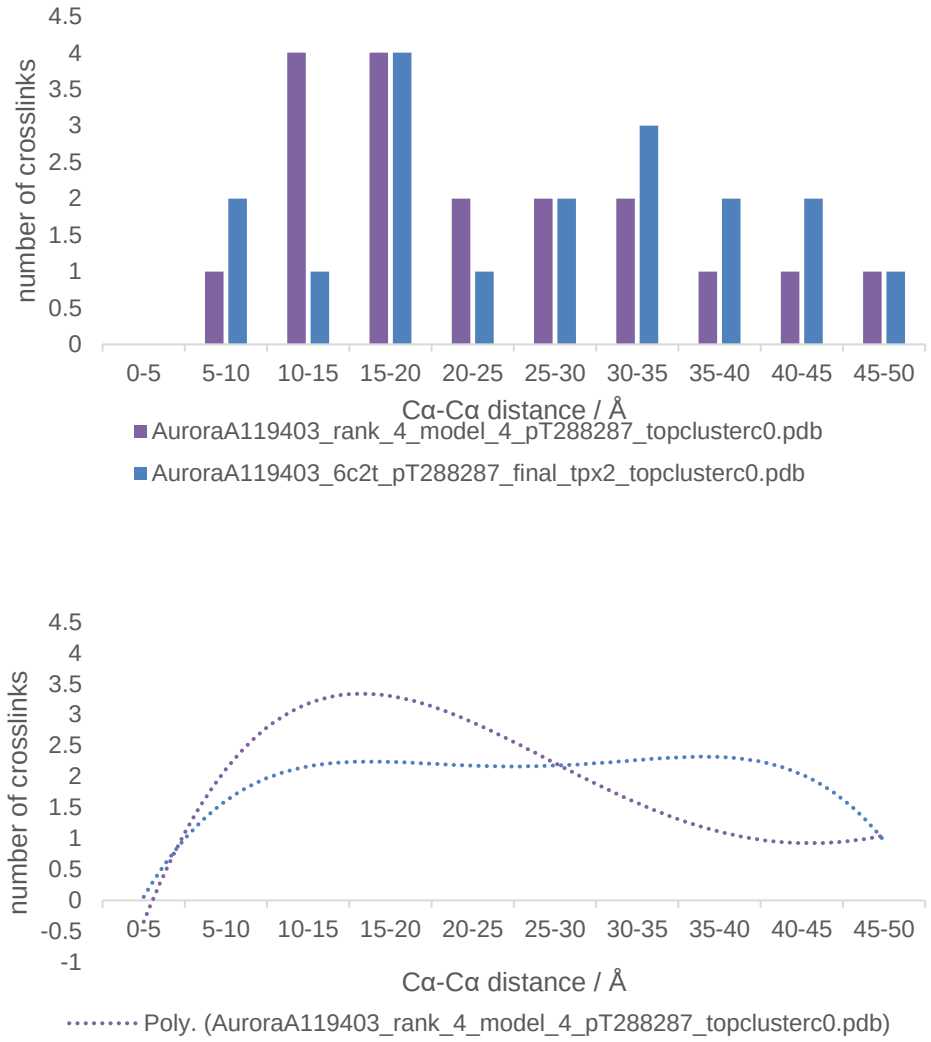


Figure 3-31 Histograms of of normalised Cα-Cα distances found for AurA+TPX2 crosslinks annotated on model structures

Histograms (Upper) using 5 Å bins and used to generate quadrinomial trendline (Lower). Comparison of crosslink distances generated from annotating the specified models with the crosslinks with more than a single CSM.

3.4.2.5.2 Evaluation of output structures using normalised crosslinks

We wanted to assess whether the 'most important' crosslinks and monolinks *actually* mattered. During this model evaluation process, it was unclear whether our models need to be within tolerance for all crosslinks, or a subsection of them. In the previous Chapter we used reactivity normalisation to remove crosslinks that were generated from overly reactive residues. A

comparison of the models scored by XLM tools shows little discrimination between our final pool of MD stabilised AurA+TPX2 conformations.

3.4.2.5.3 Examination of the over-long crosslinks for all high scoring models

Examining the over-long crosslinks for the largest cluster, we saw AurA N-lobe crosslinks to the TPX2 N-terminus and the Lys156-Lys162 crosslink, which have both been overlong for all previous models of AurA. Table 3-17 shows the overlong crosslinks found and their SASD distance on the C0 cluster for CF A and 6C2T. As the N-terminus of TPX2 is the most dynamic in the crystallographic structures and our MD simulations it is not unreasonable that the crosslinks are overlong. If it were the case that the N-terminus of TPX2 adopted a stable structure or was contributing to binding it would be concerning.

Table 3-14 Over-long crosslinks for all high scoring models

The table shows crosslinks that were found overlong by SASD. Δ gives the difference in SASD in Å for each crosslink that was overlong.

Type	Protein 1	Res 1	Protein 2	Res 2	CSM	Norm	topcluster C0 / Å		Δ / Å
							CF A	6C2T	
Interaprotein	AurA	156	TPX2	-2	6	0.14	48.1	58.3	10.2
Intraprotein	AurA	156	AurA	162	5	-1.00	44.3	44.1	-0.1
Interaprotein	AurA	143	TPX2	-2	2	-2.48	53.8	37.4	-16.4
Interaprotein	AurA	171	TPX2	-2	35	-11.70	37.1	35.1	-2.0

3.4.2.6 In summary

3.4.2.6.1 Active AurA conformer with TXP2 1-45

The use of MD generated structures explains the idiosyncrasies with the crosslinks of AurA and TPX2. The NTE of TPX2 is very dynamic, able to interact with both the phosphorylated residues of the activation loop, the NTE of AurA, and itself. This seems to explain why the NTE of TPX2 generated a large number of crosslinks. It was surprising to see the temporary slippage of TPX2 Tyr8 from the Y-pocket, which mediated this dynamic process.

We also formulated a potential hypothesis for why there were no Lys38 crosslinks observed: its engagement to the NTE as TPX2 rolls into fuller contact with the activation loop and alpha helix.

3.4.2.6.2 Structured subdomain out, discontinuous helix structure

The activation loop out structure that was unexpectedly high scoring in Chapter 2 was suggested to be stable by MD, and as it continued to

generate a majority of in tolerance crosslinks, despite its unexpected structure, and it seems likely that this may be of biological relevance. This continued even with the addition of TPX2. The AurA N-lobe-TPX2 PPI in the MD simulation was very stable. The lack of Lys38 crosslinks was justified by the Lys5 and Asp11 interaction blocking solvent access, whilst K38 formed H-bonds with the alpha helix of AurA.

Unlike the active pose, the N-/C-lobe of AurA further hinged open to the structured subdomain of the activation loop, to effectively pack against the C-lobe. The phosphorylated residues, supported by TPX2, were held offset from AurA, with a network of Arginine residues from the activation loop and the αC Helix.

3.5 Discussion

3.5.1 Introduction

In Chapter 2, we found that although kinase-IDR+IDP complexes could be crosslinked to give structural insights into the interactions and dynamics of these important targets for structural characterisation, the generated Cα-Cα distances lacked resolution. This reduced our confidence in the capacity of the crosslinking datasets to discriminate between different models. We hypothesised that this was due to the use of Cα-Cα distances as our chosen crosslink metric not accurately representing the distances that the in-solution crosslinker was spanning. In addition to this, the manually constructed portions of the structural models did not reflect the average lifetime conformation of the IDRs of AurA. To resolve these issues, we used MD simulations to generate models that we hoped would reflect the energy stabilised conformers of AurA, that could then be evaluated using the Topf group's crosslink and monolink evaluation software, XLM tools.[301]

After producing a range of AurA Apo conformers, using a large, quantified range of crystal structure adapted models, we subjected them to evaluation by XLM tools, which gave SASD distances and monolink depth predictions of each structure, and showed very striking results. Firstly, two models that were generated from a domain-swapped face-to-face dimeric crystal structure scored highly, and the canonically active structure scored poorly. This was remarkable, as the former structures appeared unstable as monomers, with the activation loop partially or fully ejected from the C-lobe and held solvated. Initially, we speculated that we were either observing a transitory structure that had been kinetically trapped out, that our crosslinks

lacked the coverage to accurately place the activation loop (a discussion point of the previous chapter), or that the scoring or models were biased towards open structures.

3.5.1.1 Observations on XLM tools

We found XLM tools to be an effective method to score large numbers of models with different crosslink/monolink lists. MoDs which was, at the time of writing, the only software available to score models using monolinks, allowed the direct quantitative use of monolinks to interpret data, whereas in Chapter 2, monolinks were relegated to comparative use. However, at times the monolinking scores appeared to conflict with the crosslinking scores.

A priori, a structure will score highly by monolink scoring if it is open, as the depths of residues will be reduced. A compacted structure, on the other hand, will score highly by crosslink scoring, as the surface distances are reduced. We reasoned that molecular dynamics could be used to generate better models of kinases with IDRs, as we could not confidently dismiss models with over-long crosslinks that could be justified in a stabilised position. We thus selected the canonically active structure and the two highest scoring to generate three structures which were reasonable approximations of the lifetime average of the conformers.

Based on the work presented here, the use of the lower resolution crosslink Cα-Cα distance as a metric for lower resolution partial homology or manually constructed kinase structures is acceptable. This is because the more precise SASD metric gives the misleading appearance of higher resolution interpretation. We found that as we did not have crosslinks from some regions of our target protein, we could not positively discriminate models even with the higher resolution data. However, using the patterns or themes of crosslinks, we were able to make qualitative judgements about what range of structures could be present.

The false impression is especially pertinent when the range of MoDS and NXML scores shows very little deviation. Thus reaching the ‘limit of detection’ for the crosslinking dataset, all models are of equal validity. To improve the confidence of the discrimination between models, either the coverage of crosslinks has to be improved or orthogonal methods have to be used.

The scoring of the AurA+TPX2 model was particularly frustrating, suggesting we had reached the limits for the data set. We saw no appreciable metric to discriminate between the models using XLM tools. Unexpected or

unimportant elements were scored highly. We reasoned that we were seeing two opposing stable monomers; this was born out by MD simulation, with the MD simulation of AurA+TPX2 with the highest cluster showing binding as a direct NTE contact, and the second held in contact to the activation loop. This leads to our recommendations. XLM tools is trying to generate structures that match the average of the crosslinks and monolinks found, just as XL-MS is also an averaging technique. Thus, it is important to use a range of experimental methods to determine the number of conformations within an ensemble.

3.5.1.2 MD simulations

The AurA±TPX2 interactions, especially changes to conformation of the activation loop, were extensively investigated by MD simulations. This was both to aid the interpretation of in-solution structural data and to explore previously observed phenomena. It was reasonable to use AMBER to generate stable versions of our models, because of the resolution of the XL-MS data. Of the two novel structures, one showed a collapsing A-Helix, and so we interpreted that model as unstable and discounted it from this study. Although AurA does not require the presence of a nucleotide, this may have been affecting the stability of our MD simulations. This has previously been shown by Muzavizadeh *et al*.[318] This does not reflect in-solution experimentation, as the addition of a nucleotide appears not to disrupt the equilibrium position when measured by FRET probes. [164]

We were surprised that TPX2 directly stabilised the conformer structure adapted from the crystallographic structure 6C2T that after MD, stabilised to a Structured subdomain out, discontinuous helix conformer. This was indicated by the reduction in pseudo b-factor for the residues of the activation loop, since the final cluster generated showed that TPX2 33-43 directly interacts with the flipped-out activation loop. This matches previous MD simulations, and the current biological dogma, which posits that TPX2 stabilises conformers that are active.[164]

3.5.2 The use of IMS-MS and MD simulations with XL-MS

We were fortunate that an IMS/MS data set was published showing the presence and CCS values of AurA Apo and AurA+TPX2 conformations.[308] This helped us to guide the models that we generated, as the conformation and apparent stability of the open conformer was unexpected. The use of CCS values and MD multiply the resolution gained through XL-MS. Although this approach has been shown by other groups on different target types, its

use with kinase+IDP complexes helps to disentangle the averaging effects caused by the crosslinking of dynamic complexes.[256 305] We thus do not recommend investigation of kinase-IDR+IDPs without some understanding of the conformational ensemble of the complex.

3.5.2.1.1 Analysis of conformations using IMS data

IMS analysis of active AurA KD (the same construct as used in XL-MS) with MD simulations reports three conformers with collision cross sections (CCSs) of 20.4, 22.5 and 24.8 nm^2, with a relative abundance of 11, 54 and 34 %. The smallest conformer was unassigned, conformer II was assigned as a DFG-in or DFG-out structure (see section 1.3.2.2), and conformer III was assigned as an activation loop open structure. AurA+TPX2 produced the same three conformer ensemble of 20.0, 22.4 and 26.6 nm^2 (no relative abundance reported). Crucially, the curve for AurA Apo was a continuous curve formed from the overlap of three Gaussian curves, whereas AurA+TPX2 was a discontinuous curve with no overlap with the smaller, active conformer.

The hypothesis generated using IR probe labelled AurA and FRET data is that TPX2, in binding to AurA, stabilises the active conformation and thus reduces the prevalence of the inactive activation loop pose, whilst simultaneously stabilising the active DFG pose. The discontinuous nature of the IMS-MS curve could be interpreted as TPX2 stabilising the active and our activation loop out structure, whilst disfavouring the intermediate activation loop positions.

Zorba *et al* showed that TPX2 stabilises the formation of a long-lived dimer, which they propose is necessary for AurA activation through auto-phosphorylation.[145] This is potentially seen in the changes to the stoichiometry of crosslinked species formed in the present and absence of TPX2. The addition of TPX2 removed any crosslinked oligomers in favour of the clear 1+1 dimer complex.

3.5.3 Biological impacts

3.5.3.1.1 XL-MS will produce an averaged picture of conformers

XL-MS will produce an averaged picture of the conformers and complexes found in-solution for the system being studied. However, modulation of the equilibrium by, for example, small molecules, mutations in the proteins studied, or the introduction of interactors, can be observed and rationalised.

3.5.3.1.2 Insights from Chapter 3

Chapter 2, using the qualitative evaluation of Euclidean Cα-Cα distances of crosslinks found in AurA±TPX2, showed that AurA NTE must be displaced by TPX2, and that our activation loop open structure was reasonably tolerated, in addition to the canonically active structure. It was expected that an MLN-like inactive structure would be observed in the dataset but, potentially due to the protocol used, crosslinks involving the activation loop were not seen.

In this chapter, MD simulations of hypothetical AurA models were used to produce clustered structures that it was inferred were stable. XLM tools software was used to score these structures using SASD and monolink depth, instead of the SLD. This approach was anticipated to improve confidence in the analysis by reducing the hypothetical search space. An energy minimised model of a dynamic complex, measured using distance restraints that respect the complex surfaces found in kinases, would produce distance measures of the experimental crosslinks found that did not require the addition of a post-hoc tolerance factor. This was one improvement, the other being was the ability to utilise the monolink data quantitatively.

3.5.3.1.3 Unexpected open conformation of AurA performed well

It was surprising that a monomeric structure generated from a domain swapped crystal structure could satisfactorily rationalise the XL-MS data, whereas a monomeric inactive structure could not.

Several groups have used time-resolved FRET of AurA KD with the activation loop and D-Helix to propose that the activation loop is in dynamic equilibrium between an MSP1-like pose and an active pose. It has also been shown previously that the equilibrium is modulated by the addition of TPX2 or allosteric inhibitors; TPX2 shifts AurA from a ratio of 23 %/73 % inactive/active to 12 %/88 %. The inhibitor MLN8054 is thought to be inactivating, because it stabilises the inactive conformation. This model was nuanced by the use of a nucleotide pocket IR probe that was modelled and showed that there was an autoinhibitory activation loop in the conformer that formed part of the active conformer seen in the FRET probe. This autoinhibited ‘active’ state was shown to be caused by the AurA sequence catalytic site differing from other AGC kinases that do not require additional stabilisation of the catalytic site by allosteric interactors. Ruff *et al* used DEER spectroscopy to confirm the AurA special active-like inactive conformer, calling it an autoinhibitory DFG-in conformer, which TPX2 then destabilises in favour of the active conformer.[164] An inactive AurA Apo

crystallographic structure has yet to be presented. The MSP1-like structure is only seen with AurA complexed with inhibitors.

The domain swapped dimer that is required for auto-phosphorylation was not modelled into the ensemble, and it is reasonably inferred from the crystal structures that a domain swapped dimer would produce the same FRET distance. However, the dimerization mechanism has previously been shown to be dependent on the cysteine residue in the activation loop. It is therefore possible that C290A reduces the relative abundance of the dimer.

Tomlinson *et al* generated CCS data of AurA KD±TPX2 1-43 that was interpreted to show two distinct conformers.[308] A canonical active conformer, and a second conformer larger than this, were consistent with their experimental data. The predicted CCS values of an activation loop open model generated from 4C3P was in line with this larger conformer. Tomlinson et al. also struggled to differentiate between MLN-like structures and active-like structures, as these classes were predicted to have overlapping CCS values.

As part of this present study, the active AurA KD structure generated from the largest portion of the 1OL7 MD simulation was predicted to have a CCS value of 22.9 nm^2, matching the experimental value of 22.5 nm^2. This model scored well for crosslinks but poorly for monolinks and was shown to be very stable by pseudo b-factor analysis and RMSD over time. It was also consistent with the original crystallographic structure, with the addition of AurA NTE bound into the F-pocket.

The manually generated 4C3P and 6C2T structures scored well by the output structures, but 6C2T scored better and was more stable. 6C2T was predicted to have a CCS value of 24.67 nm^2, which again matches the experimental value of 24.8 nm^2. The Tomlinson *et al* study predicts, taking averages of all frames of the simulation, 24.3-25.5 nm^2, which is consistent with the 23.53 nm^2 found for the largest cluster of our 4C3P model.[145] 4C3P scored much worse, and the alpha Helix snapped mid-simulation suggesting that it is not stable. The CCS predicted value for 4C3P of 23.53 nm^2 does not match the experimental value. If it had scored well, we might be looking at a novel inactive structure, but it did not score well, and so is dropped.

It is of note that Tomlinson *et al* do not report the presence of a dimer in their ensemble.

As the 6C2T clustered structure from the most frames scored highly by our crosslinking datasets and with a CCS value that corresponds to the

experimental, further study into the biological relevance of this conformation is warranted.

3.5.3.1.4 The activation loop out structure has an identical DFG and R-spine pose to the autoinhibited DFG-in structure

Serendipitously, the DFG motif and R-spine of the clustered output of the 6C2T model aligns convincingly with those found in AurA structure 5LK8. 5LK8 is representative of the auto-inhibitory DFG-in inactive structure that Ruff *et al* reported by the IR spectroscopy of DTNB labelled Q185C AurA KD and MD simulation.[164] This is even without ADP and Y-pocket interactors present, or the phosphorylation of the activation loop subdomain. The alignment improves for the AurA+TPX2 activation loop out conformer.

3.5.3.1.5 How does the presence of activation loop out conformer behave with TPX2?

The addition of TPX2 to the 6C2T conformer was shown to stabilise the activation loop during MD, and the resulting clustered model scored highly against the crosslinking dataset. Additionally, we were surprised to see that its predicted CCS value corresponded to the experimental value of a minor conformer of the identical construct as was used for crosslinking. We noticed that the IMS/MS curves for AurA+TPX2 showed two distinct curves, whereas the AurA Apo had a continuous single curve. We interpreted TPX2 breaking the IMS curve as binary stabilisation of dynamic activation loop. This effect can also be seen in other in-solution techniques. Thus, it is predicted that TPX2 is stabilising the canonically active structure and our extended novel structure. This fits the current model of TPX2 activation of AurA.

3.5.3.1.6 Could N-lobe/C-lobe hinging motions account for the CCS mismatch?

A constant frustration of AurA crosslinking datasets is that the AurA Lys156-Lys162 (Beta1 to the catalytic lysine) could not be reconciled with any model of AurA (Table 3-7, Section 3.4.1.1.3.1). By SASD measurement it was found to be ≈5-10 Å over the upper bound of crosslink distance for DSBU. Two hypothesises present themselves: either the dynamic motions of the loop that bears Lys156 on the N-lobe allows the closing of the distance between these residues, or the C-lobe, and hinge region, pivots open allowing easier access to between the crosslinked residues (Figure 3-33).

Akamine *et al* presented an apoenzyme PKA crystallographic structure that lacked a nucleotide and has the C-lobe hinged open.[332] Cyphers *et al* performed EPR of AurA and showed that addition of ADP caused a

shortening of the mean distance between the αC-helix and the activation loop.[156] This suggests a hinging shut of the N- and the C-lobe upon addition of a nucleotide.

The crosslinking and measurement of crystallographic and wildtype constructs of AurA in the presence and absence of nucleotides should be a priority for any future work.

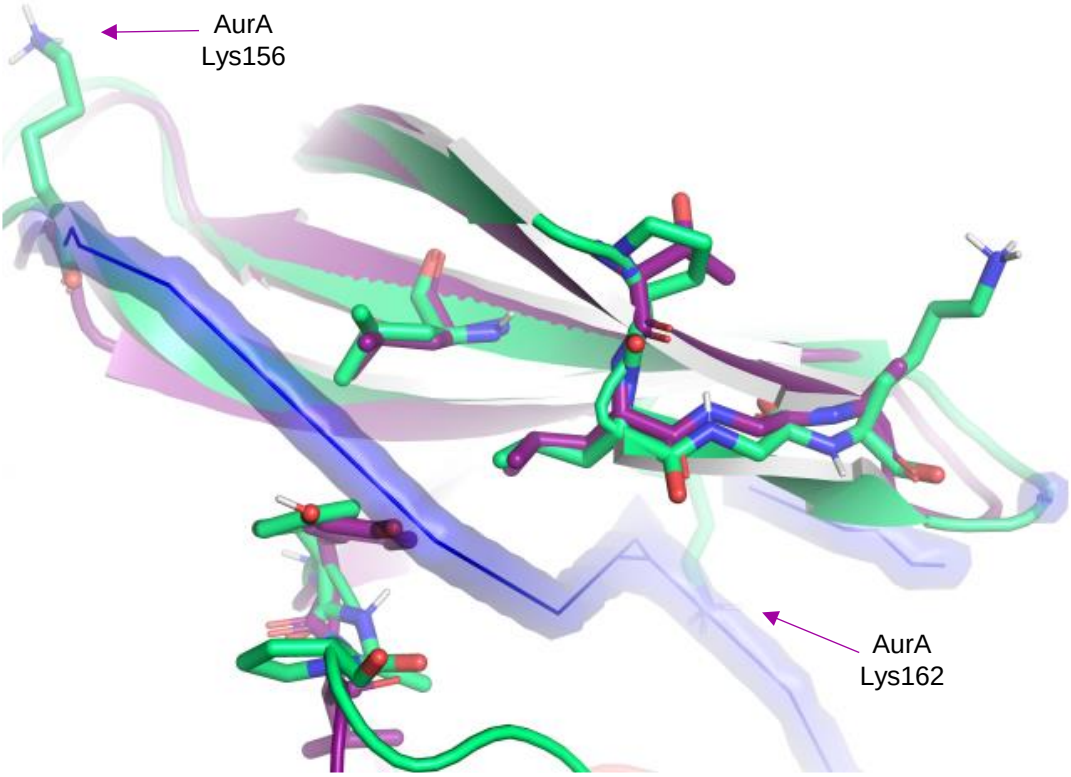


Figure 3-32 Lys156-Lys162 crosslink annotated to Cluster C0 6C2T

Cluster C0 6C2T in green aligned to apoenzyme PKA in purple (PBD 1J3H) with crosslink Lys156-Lys162 highlighted with occluding residues shown.

3.5.3.2 Potential modifications to the activation model of AurA by TPX2

Thus we submit that, as we saw in Chapter 2, the role of AurA dimerization and this novel conformation in activation needs further exploration. Although unexpectedly stable by MD, we found that an open conformer of AurA generated from a domain swapped dimeric structure scored highly by crosslinks and monolinks using XLM tools, in addition to having a predicted CCS value that matched a recently discovered larger conformer of AurA. We found that this conformer was stabilised by the addition of TPX2, and could represent an important step on activation pathway of AurA. As kinase inhibitor campaigns focus on the conformational control of AurA,

understanding if and where this conformer fits into the ensemble of AurA is important to ongoing drug discovery work.

3.1 Conclusion

AurA is a kinase whose activation and regulatory control are vital to orderly and consistent mitotic progression. For the structural biologist it has posed several interconnected challenges. Firstly, AurA is conceptualised as a PKA or AGC kinase, whose control are usually mediated by intra-protein interactions, but control of AurA is mediated by IDR interactors. Secondly, these IDR interactors have resisted both crystallographic and NMR attempts to characterise them. Thirdly, AurA's 'broken' regulatory surfaces allow it a wider range of dynamic movement that is modulated by these interactors. Finally, AurA's conformational and stoichiometric landscape is also modulated by a redox dependent mechanism.

Our study, with its deep exploration of the literature surrounding the conformational control of AurA and its interactors, found, using XL-MS and MD simulations, that a dimer-like activation loop out conformer with a misconfigured DFG-in subdomain may bridge the gap between activatory dimerization and the canonical DFG-in active conformer. Difficulties with our XL-MS protocol did however result in several unresolved questions.

4 Conclusion and Outlook

4.1 Have the posed biological research questions been answered?

Initially, this PhD was tasked with validating the use of XL-MS in exploring the interaction of AurA with N-Myc. We were hesitant that the disordered protein N-Myc would generate reliable and functional distance data, and so we used the well-characterized AurA+TPX2 complex to validate and explore the boundaries of this approach. During this time, the understanding of the dynamics of AurA+TPX2 were changing, due to the advances that the insights of TR-FRET data was generating.

We found that the canonically inactive MSP1-like conformer could not be rationalised by our crosslinks and that there existed a novel open conformer. We found, unexpectedly, that the AurA NTE was held close to the N-Lobe and not permanently solvated and held away from the N-lobe. Thus, it required displacement in order to bind TPX2. Due to the limits of time, refinement of this model through further optimisation of the protocol could not be performed. This novel structured subdomain out, discontinuous helix conformer could not therefore be rationalised into the current model of TPX2 mediated AurA activation. We were initially surprised with the persistent generation of dimeric and oligomeric species, and assumed they were artefacts of the crosslinking process. However, this appears not to be the case, and the appearance of the structured subdomain out, discontinuous helix conformer of AurA remains elusive.

It was unfortunate that AurA+N-Myc was more problematic and generated a more complex heterogenous crosslinked mixture that refused to yield useful AurA+N-Myc crosslinks, despite the many changes made to the crosslinking protocol. This prevented any advances in the understanding of the AurA+N-Myc structure, apart from the insights into the potential stoichiometry; we also validated that N-Myc binds across the TPX2 binding site of AurA. It was striking that in this case, the AurA NTE was not displaced, unlike with the binding of TPX2.

4.1.1 Impact on AurA biology

Comparison of the kinase domain of AurA to other AGC kinases revealed that the activation loop lacks a key hydrophobic residue that anchors the activation loop to the C-lobe. Additionally, the conserved APE motif found in the phosphorylated loop of AGC kinases is present as an PPE motif in AurA.

Hence we can speculate that the dynamics of the activation loop of AurA leads to the structured subdomain out, discontinuous helix conformer, unlike other AGC kinases. The formation of a transient dimer is likely to be required for activation/auto-phosphorylation. Throughout this study, it was expected that the MSP1-like conformer of AurA would be corroborated by our data but no direct evidence for it was found. Further investigation, using the MLN series of inhibitors and AurA mutants that stabilise this conformation is needed to rule in or out its presence by XL-MS given the majority of literature supports its role in AurA activation.

The finding that the crosslinking reaction of AurA generates dimer and oligomers in a manner modulated by the presence of binding partners is an intriguing link to the activation pathway of AurA that other structural methods did not find. This demonstrates the power of XL-MS as an approach, as it doesn't require labelled protein to understand conformation changes modulated by stimuli.

Our hypothesis that the NTE of AurA (at least in this construct) gatekeeps access to the F-pocket of AurA will affect how the interactors engage with AurA. We were able to explain why small molecules target the F-pocket.

Although we weren't able to fully explore the molecular basis of the N-Myc TAD and AurA binding, XL-MS did support the investigation of the stoichiometry of the interaction. Critically, understanding if N-Myc displaces the NTE in the F-pocket would question the binding model proposed by E Lake and others.[326] Our work further indicates that N-Myc TAD, like TACC3 and TPX2, uses avidity with a combination of pockets of AurA to bind and that reduction in the length of N-Myc when studying the interaction may disturb the balance of multiple weak interactions with AurA.

4.1.2 Potential improvements to the XL-MS of kinase with IDRs

Herein, we have shown that kinases with IDRs that interact with IDPs can successfully be studied by XL-MS. However, as we failed to generate crosslinking conditions that gave a single species, the necessary use of in-gel digestion reduced the quality of data observed. We found that both members of the complex under investigation had dimeric conformers as part of their in-solution equilibrium ensembles. This makes the generation of a homogeneous crosslinked mixture extremely hard to produce. The next step, following on from the insights of this thesis, would be to investigate the composition of species crosslinking using native ESI- MS, to more reliably gauge how the crosslinking conditions alter the species produced.

We found normalisation and comparison of the post-hoc monolink changes in reactivity to be a compelling additional use of DSBU crosslinking. Direct comparison with NHS-acetate to calibrate the reactivity of residues would be the next logical step, following on from our work here. The use of CSMs in place of Area Under the Curve of the TIC trace analysis from which to normalise from was surprisingly effective. A direct comparison of these methods needs to be performed before more widespread adoption of the approach.

Finally, we propose that it may be beneficial to deploy strategies to temper the reactivities of residues that are over-presented in existing data, alongside the titration of the IDP under investigation, to generate quantitative XL-MS data capable of investigating dynamic kinases with IDRs or similar targets that bind IDPs.[289]

As to the scholarship question of whether XL-MS validating X-ray-structures and homology models or are X-ray-structures and homology models validating the distance restraints found by XL-MS, we found that XL-MS used in combination with orthogonal MS techniques and MD yielded the most insights into our system. Here conformers of AurA that were generated from crystallographic structures were subject to MD simulations to understand their stability and subsequently scored by XL-MS datasets and assessed against CSS datasets. Development of XL-MS for this system led to a combinatorial approach to kinase dynamics and PPI should aid in the elucidation of the effect small molecule inhibitors have on the conformational ensembles of kinases.

5 Appendices

5.1 For the introduction

5.1.1 Species Alignment for consensus sequences for AurA interactors

To generate the aligment of AurA interactors with PKA NTE/NTE, the multiple sequence aligments used previously were directly adapted and included TACC3.[88] Homo sapiens, Mus musculus, Rattus norvegicus, Xenopus laevis, Xenopus Tropicalis, Drosophila melanogaster, and Caenorhabdtitis elegans were used by Tavernier *et al* however the species and protein Uniprot IDs are detailed in the table below.

Lane	Species	Taxonomy ID	Entry TPX2	Entry Bora	Entry TACC3
1	Oryctolagus cuniculus (Rabbit)	9986	G1SJ85	A0A5F9DMN0	A0A5F9DVK4
2	Caenorhabditis elegans	6239	not included	not included	not included
3	Drosophila melanogaster (Fruit fly)	7227	not included	not included	not included
4	Mus musculus (Mouse)	10090	A2APB8	Q8BS90	Q9JJ11
5	Rattus norvegicus (Rat)	10116	D3ZA78	Q5M864	A0A8I6ATD8
6	Xenopus laevis (African clawed frog)	8355	Q6NUF4	Q6DJL7	Q9PTG8
7	Xenopus tropicalis (Western clawed frog)	8364	A4IH24	A0A6I8PVN9	A0A8J1J635
8	Ovis aries (Sheep)	9940	A0A6P7ENX0	W5Q102	W5Q353
9	Gallus gallus (Chicken)	9031	A0A8V0YGC2	A0A3Q2U088	A0A8V0YWT0
10	Sus scrofa (Pig)	9823	A0A287A0D0	A0A287BCZ6	A0A287AT98
11	Danio rerio (Zebrafish) (Brachydanio rerio)	7955	E9QE51	Q5U3U6	Q6Y8G7
12	Equus caballus (Horse)	9796	F7A3W9	A0A3Q2HQR3	F7AEJ6
13	Felis catus (Cat) (Felis silvestris catus)	9685	A0A2I2UW03	M3WCM8	A0A5F5XZB5
14	Bos taurus (Bovine)	9913	A6H6Z7	F1N3C9	A6QL93
15	Vulpes vulpes (Red fox)	9627	A0A3Q7SJ67	A0A3Q7T8T1	A0A3Q7THG5
NA	Homo sapiens (human)	9606	Q9ULW0	Q6PGQ7	Q9Y6A5

5.1.2 Bora Multiple protein sequence alignment

Aligned using https://www.uniprot.org/align and Consensus sequence generated using Clustal Omega

```
sp|Q6PGQ7|BORA_HUMAN            MQITPETPGRIPVLNPFESPSDYSNLHEQTLASPSVFKSTKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQALYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSQCH 126
sp|Q8BS90|BORA_MOUSE            MQITPETPGRIPVLNPFESPSDYSNLHEQTLASPSIFKSTKL--PTPGKFRWSIDQLAIINPVEIDPEEIHRQASYLRLSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHDGKQPSELH 126
tr|A0A5F9DMN0|A0A5F9DMN0_RABIT MQISPETPGRIPVLNPFESPSDYCNLHEQTLASPSVFKSAKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQASYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSEYH 126
sp|Q5M864|BORA_RAT             MQITPATPGRIPVLNPFESPSDYSNLHEQTLASPSVFKSTKL--PTPGEFRWSIDQLAVINPVEIDPEEIHRQASYLSLSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHDGKQPSELH 126
sp|Q6DJL7|BORA_XENLA            TQLTPETPGRAAILNPFESPSDYCSLHEPFVSSPSVFKPSKS-AATPQQFRWSIDQLAAINPIEIDAEDIHRQALYLSRAKTDKETEERRQKAIEEFFTKRTIVPSPWTQHEGKPAAPFH 124
tr|A0A6I8PVN9|A0A6I8PVN9_XENTR TQLTPETPGRAAILNPFESPSDYCSLHEPFVSSPSIFKPTKS-AVTPQQFRWSIDQLAAINPVEIDPEDIHRQALFLSRAKTDKETEERRQKAIEEFFTKRTIVPSPWTQHEGKQAAPFH 124
tr|W5Q102|W5Q102_SHEEP          MQITPETPGRIPILNPFESPGDYSNLHEQTVSSP-VFKSTKL--PTPGEFRWSIDQLAVINPVEIDPEDIHRQALYLSRSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSEYH 187
tr|A0A3Q2U088|A0A3Q2U088_CHICK MQISPETPGRVAVLNPFESPNDYCTLQEQIVSSPSVFKSTKS-SSTPGKFRWSIDQLALINPVEIDSEDIRRQAMYLSHARIDKETEERRQKAIEEFFTKSLIVPSPWTEHEGKQVSQFN 127
tr|A0A287BCZ6|A0A287BCZ6_PIG    MQITPETPGRIPVLNPFESPGDYSSLHEQTLSSPSVFKSTKL--PTPGEFRWSIDQLAVINPVEIDPEDIHRQALYLSRSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSESH 126
sp|Q5U3U6|BORA_DANRE            IHITPETPARPTILNPFESPNDYHRLHESVVPSPSVFKSCKSSSATPAKFKWSIDEMANLLPVEIDPEDIHRQAVFFSQARADSEIEEKRQHAIEQFFTKGAIVPSPWGPLTSKQSMKLP 126
tr|M3WCM8|M3WCM8_FELCA          MQITPETPGRIPVLNPFESPGDYSNLHEQTLSSPSVFKSTKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQALYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSEYH 177
tr|A0A3Q2HQR3|A0A3Q2HQR3_HORSE MQITPETPGRIPVLNPFESPGDYSNLHEQTLSSPSVFKSTKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQALYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSDYH 238
tr|F1N3C9|F1N3C9_BOVIN          MQITPETPGRIPVLNPFESPGDYSNLHEQTVSSPSVFKSTKL--PTPGEFRWSIDQLAVINPVEIDPEDIHRQALYLSRSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSEYH 126
tr|A0A3Q7T8T1|A0A3Q7T8T1_VULVU MQITPETPGRVPVLNPFESPGDYSNLHEQTLSSPSVFKSRKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQALYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSEYH 177

sp|Q6PGQ7|BORA_HUMAN            MQITPETPGRIPVLNPFESPSDYSNLHEQTLASPSVFKSTKL--PTPGKFRWSIDQLAVINPVEIDPEDIHRQALYLSHSRIDKDVEDKRQKAIEEFFTKDVIVPSPWTDHEGKQLSQCH 126
Consensus                       MQITPETPGR PVLNPFESPSDY  LHEQTLASPSVFKSTK --PTPGKFRWSIDQLAVINPVEIDPEDIHRQA YLSHSRIDKDV DKRQKAIEEFFTKDVIVPSPWTDHEGKQ S
```

5.1.3 Tacc3 Multiple protein sequence alignment

Aligned using https://www.uniprot.org/align and Consensus sequence generated using Clustal Omega

```
sp|Q9Y6A5|TACC3_HUMAN          --------------------------EPVP------THQQGQP----AL-ELKEESFRDPAEV----LGTGAEVDYLEQFGTSSFK-ESALRKQSLYLKFDPLLRDSPGRPVP--VATET 581
tr|A0A5F9DVK4|A0A5F9DVK4_RABIT VISAVVHATDEEKLAVTNQKWTCMTVDLEA------DKQDFPQPSDLST-FVNETKFSSPTEE--LDYRNSYEIEYMEKIGTSLPQDDDAPKKQALYLMFDTSQ-ESPVKSPPARVSD-- 272
sp|Q9JJ11|TACC3_MOUSE          -----VT--------------PPIEPLLEP------SHQGLEP----VV-DLKEESFRDPSEV----LGTGAEVDYLEQFGTSSFK-ESAWRKQSLYVKFDPLLKDSPLRPMP--VAPIT 370
tr|A0A8I6ATD8|A0A8I6ATD8_RAT   -----VT--------------LPVEPVPES------SHQGPEP----LL-DPKEESFRDPAEV----LGAGAEVDYLEQFGTSSFK-ESAWRKQSLYLKFDPLLKESPLRSMPVPVAPVT 330
sp|Q9PTG8|TACC3_XENLA          ------------------------------EFKLATEADFLL--------AADMDFKPASEIFSEGFRQPVEIDYLENFGTNSFK-ESVLRKQSLYLKFDPLLRESPKKSAA-----GI 646
tr|A0A8J1J635|A0A8J1J635_XENTR ------------------------------AFNLPTDADFFP--------AADMDFKPASEIFPGGFGQPIEIDYLENFGSNSFK-ESALRKQSLYLKFDPLLRESPKKNAA-----GN 597
tr|W5Q353|W5Q353_SHEEP         -----VH--------------QCAQLRTTI------GGRGWVT----VG-RDRAEETGSLSEV----LGAGAEVDYLEQFGASSFK-ESALRKQSLYLNFDPLLQDSPRRLAP--SRPGG 523
tr|A0A8V0YWT0|A0A8V0YWT0_CHICK L----VSAD--------SGKRSMPAEEIKPTSHKNEITADQQT----AS-AEAEELFRPSAEV----LGMG--IDYLEQFGTSSFK-ESALRKQSLYLKFDPLLRDSPRKQPIFGTIEAN 599
tr|A0A287AT98|A0A287AT98_PIG   ------------------------------P------VPCSPGP----AP-DLSGEQFRDPADV----LGTGAEVDYLEQFGASSFK-DSALRKQSLYLKFDPLLQDSPRRPAP--TGPKP 218
tr|Q6Y8G7|Q6Y8G7_DANRE         M----------------------AVPESI-----ADIQ--NQTCDLGVAASNEDEFVPG-ALFMPGGDFDGQIDYLEQFGSSNFK-ESALRKQSLYLKFDPLLKESPKKNGGEGSNFG- 664
tr|F7AEJ6|F7AEJ6_HORSE         -----TA---------------SPTKPASP------PLQGPEP----AS-DLSDEHFRDPAEV----LGAGAEVDYLEQFGTCLFK-ESALRKQSLYLKFDPLLMDSPQRPAP--VVPAT 548
tr|A0A5F5XZB5|A0A5F5XZB5_FELCA -----GR---------------ASASAPPP------PPHGPEP----PS-DLDEEHFRDPAEV----LGTGADVDYLEQFGSSSFK-ESALRKQSLYLKFDPLLKDSPQRPVP--AAPDA 545
tr|A6QL93|A6QL93_BOVIN         -----PT--------------A-PTNPTPP------TERGPEP----TL-DLNGEQFRDPAEV----LGAGAELDYLEQFGAPSFK-ESALRKQSLYLNFDPLLQDSPQGLTP--SSSG- 522
tr|A0A3Q7THG5|A0A3Q7THG5_VULVU ------S---------------APAGSLGG------PPQDPEP----AS-DPKEECFRDPTEV----LGTGAEVDYLEQFGSSLFK-ESALRKQSLYLKFDPLLKDSPRRPVP--VAPET 528

sp|Q9Y6A5|TACC3_HUMAN          --------------------------EPVP------THQQGQP----AL-ELKEESFRDPAEV----LGTGAEVDYLEQFGTSSFK-ESALRKQSLYLKFDPLLRDSPGRPVP--VATET 581
Consensus                                                                 EE FRDP EV----LG GAE DYLEQFG SSFK-ESALRKQSLYLKFDPLL  SPG   P
```

5.1.4 TPX2 Multiple protein sequence alignment

Aligned using https://www.uniprot.org/align and Consensus sequence generated using Clustal Omega

```
sp|Q9ULW0|TPX2_HUMAN           ----------------------------------MSQVKSSYSYDAPSDFINFSSLDDEGDTQNIDSWFEEKANLENKLLGKNGTGGLFQGKTPLRKAN--------LQQAIVTPLKPVD 78
sp|A2APB8|TPX2_MOUSE           ----------------------------------MSQVPTTYSFDAPTDFINFSSLDAEEDTENIDSWFDEKANLENKFLRQRGIGEPFQGKNSLRKAK--------LQQGFVTPLKAVD 78
tr|G1SJ85|G1SJ85_RABIT         MFNQISEPFLCTEGNPLGKVIQILGKICWSALRKMSQVQSSYTFDAPTDFINFTSLNDEEDSENVDLWFEEKANLENKFPGKNGTGELLHGKTTLRKAN--------LQQAGVTPLRQVD 112
tr|D3ZA78|D3ZA78_RAT           ----------------------------------MAQVPTAYSFDAPTDFINFSSLDAEEDTENIDSWFDEKANLENKFLRQSGIGEPFQGKTSLRKAK--------LQQGCVTPLKAVD 78
sp|Q6NUF4|TPX2A_XENLA          ----------------------------------MEDTQDTYSYDAPSI-FNFSSFHE---DHNADSWFDQVTNAENIPPDQRRLSET-SVNT--------------EQNSKVEPVQ--- 64
sp|A4IH24|TPX2_XENTR           ----------------------------------MAEAQDLYTFDAPSTFINFTAFHE---DHGADSWFDKVTNAENIPPDQRRLSEIPAVNA--------------EQNGMVEPEE--- 66
tr|A0A6P7ENX0|A0A6P7ENX0_SHEEP ----------------------------------MSQVTTSYSYDAPTDFINFSSLTDEEDTQHIDSWFDEKANLENKFPGKDGTGGLFQGKTPLRKAN--------LHQ-DVTPLRPVD 77
tr|A0A8V0YGC2|A0A8V0YGC2_CHICK ------------------------------------MSRYSFDVPNPCINFATLSD-DDVHNADSWFDQQANLENIPPAENLA-KVLQNSPAFSKPA--------LVLSSVTSQGMTK 72
tr|A0A287A0D0|A0A287A0D0_PIG   ----------------------------------MSHVTTSYSYDAPMDFINFTSLDDEEDTQNIDSWFEEKANLENKFPGKNGTGGLFQDKTPLRKAN--------LQQ-DFTPLRPVD 77
tr|E9QE51|E9QE51_DANRE         ----------------------------------MDGGSTYSYAWDAPTHVLESIDDFALDPQDNADKWFGLQAASESGVMEQ--------LTTPLRTKPFDEPVDTNKPKAMVSPLVRQ- 78
tr|F7A3W9|F7A3W9_HORSE         ----------------------------------MSQVKTSYSYDAPMDFINFTSLDDEEDTQNIDSWFEEKANLENKFPGKNGTRGLFQGKTPLRKAN--------PQQAIVTPLRPVD 78
tr|A0A2I2UW03|A0A2I2UW03_FELCA ----------------------------------MSQVKTSYSYDAPTDFINFTSLDDEEDTQNIDSWFEEKANLENKFPGKNGIGGLFQGKTPLRKPN--------PQQAIVTPLRPVD 78
sp|A6H6Z7|TPX2_BOVIN           ----------------------------------MSQVTTSYSYDAPTDFINFSSLTDEEDMQHIDSWFDEKANLENKFTGKDGTGGLYQGKTPLRKAN--------LHR-DVTPLRPVD 77
tr|A0A3Q7SJ67|A0A3Q7SJ67_VULVU ----------------------------------MSQVKTSYSYDAPTDFINFTSLDDEEDAQNIDSWFEEKANLENKFSGKNGIAGLFQDKTPLRKPN--------PQQAIVTPLRPVE 78

sp|Q9ULW0|TPX2_HUMAN           ----------------------------------MSQVKSSYSYDAPSDFINFSSLDDEGDTQNIDSWFEEKANLENKLLGKNGTGGLFQGKTPLRKAN--------LQQAIVTPLKPVD 78
Consensus                      ----------------------------------MS V SSYSYDAPS FINFSSLDDEGD   IDSWFEEKANLENKLL     GGLF GKTPLRKA --------L    VTPLKPVD
```

5.2 For Chapter 2

5.2.1 Protein sequences used in MeroX searches

>TPX2_HUMAN-1-43
GAMAQVKSSYSYDAPSDFINFSSLDDEGDTQNIDSWFEEKANLEN
>His N-Myc-1-137
MHHHHHHSSGLVPRGSGMKETAAAKFEENLYFQGAMMPSCSTSTMPGMICKNPDLEFDSLQPCFYP
DEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAF
GLGGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG
>3xFLAG-N-myc-1-137
MDYKDHDGDYKDHDIDYKDDDDKAAMVPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGP
DSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTP
NPVILQDCMWSGFSAREKLERAVSEKLQHG
>Tagless_N-Myc_22-89
GAMDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLEN
ELWG
>Tagless_N-Myc_1-137
GAMMPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPPLSPS
RGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERAV
SEKLQHG
>Aurora-A_C290A_C393A-122-403
GAMESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKAQLEKAGVEHQLRREVEIQS
HLRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRD
IKPENLLLGSAGELKIADFGWSVHAPSSRRTTLAGTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEFLV
GKPPFEANTYQETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPSN
AQNKESASKQS

5.2.2 Detected crosslinks

5.2.2.1 Single in-solution digested AurA KD + Flag tagged N-Myc 1-137

Monolinks and crosslinks shown in separate tables. XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=df363314e5117b4b0221e2e1877efd0ca8f0d0d5 and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=f9332711e277846aca736f3b6f9601c9f1d7f5ea.

Type	Protein 1	Residue 1	Top Score	Total CSMs
monolink	AurA	156	163	3
monolink	AurA	224	168	3
monolink	AurA	246	146	2
monolink	AurA	249	218	4
monolink	AurA	250	208	2
monolink	AurA	283	148	2
monolink	AurA	284	200	3
monolink	AurA	365	176	10
monolink	AurA	369	131	3
monolink	AurA	396	165	5
monolink	NMyc	-16	165	7
monolink	NMyc	-15	163	3
monolink	NMyc	-9	185	6
monolink	NMyc	-8	179	1

Source	Type	Protein 1	Residue 1	Protein 2	Residue 2	Top Score	Total CSMs
AurA_n1_flag1137_insol	intrapeptidal	AurA	141	AurA	143	157	3
AurA_n1_flag1137_insol	intrapeptidal	AurA	365	AurA	369	88	1
AurA_n1_flag1137_insol	Intraprotein	AurA	118	AurA	156	100	1
AurA_n1_flag1137_insol	Intraprotein	AurA	118	AurA	162	74	3
AurA_n1_flag1137_insol	Intraprotein	AurA	118	AurA	226	69	1
AurA_n1_flag1137_insol	Intraprotein	AurA	118	AurA	249	140	1
AurA_n1_flag1137_insol	Intraprotein	AurA	118	AurA	250	80	2
AurA_n1_flag1137_insol	Intraprotein	AurA	123	AurA	125	81	1
AurA_n1_flag1137_insol	Intraprotein	AurA	124	AurA	171	110	3
AurA_n1_flag1137_insol	Intraprotein	AurA	124	AurA	250	138	3
AurA_n1_flag1137_insol	Homotypic	AurA	141	AurA	141	201	1
AurA_n1_flag1137_insol	Intraprotein	AurA	141	AurA	148	66	1
AurA_n1_flag1137_insol	Intraprotein	AurA	141	AurA	171	84	2
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	162	86	2
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	166	55	1
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	171	185	6
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	227	63	1
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	258	87	1
AurA_n1_flag1137_insol	Intraprotein	AurA	143	AurA	284	77	1
AurA_n1_flag1137_insol	Intraprotein	AurA	162	AurA	284	73	1
AurA_n1_flag1137_insol	Homotypic	AurA	171	AurA	171	234	4
AurA_n1_flag1137_insol	Intraprotein	AurA	171	AurA	227	175	1
AurA_n1_flag1137_insol	Intraprotein	AurA	171	AurA	284	179	1
AurA_n1_flag1137_insol	Intraprotein	AurA	227	AurA	326	89	2
AurA_n1_flag1137_insol	intrapeptidal	NMyc	-15	NMyc	-9	128	2
AurA_n1_flag1137_insol	intrapeptidal	NMyc	-15	NMyc	-8	173	2
AurA_n1_flag1137_insol	Intraprotein	NMyc	-15	NMyc	125	78	1
AurA_n1_flag1137_insol	intrapeptidal	NMyc	-8	NMyc	-3	92	1
AurA_n1_flag1137_insol	intraprotein	AurA	118	NMyc	125	74	1
AurA_n1_flag1137_insol	intraprotein	AurA	124	NMyc	72	51	1
AurA_n1_flag1137_insol	intraprotein	AurA	125	NMyc	125	109	1
AurA_n1_flag1137_insol	intraprotein	AurA	141	NMyc	64	88	1
AurA_n1_flag1137_insol	intraprotein	AurA	143	NMyc	133	81	1
AurA_n1_flag1137_insol	intraprotein	AurA	171	NMyc	72	51	1
AurA_n1_flag1137_insol	intraprotein	AurA	171	NMyc	125	74	1
AurA_n1_flag1137_insol	intraprotein	AurA	171	NMyc	133	180	4
AurA_n1_flag1137_insol	intraprotein	AurA	227	NMyc	125	57	1
AurA_n1_flag1137_insol	intraprotein	AurA	227	NMyc	133	152	4
AurA_n1_flag1137_insol	intraprotein	AurA	250	NMyc	125	70	1
AurA_n1_flag1137_insol	intraprotein	AurA	250	NMyc	133	176	2
AurA_n1_flag1137_insol	intraprotein	AurA	369	NMyc	133	67	1

5.2.2.2 Duplicate in-gel flag digested AurA KD + Flag tagged N-Myc 1-137

Monolinks and crosslinks shown in the same table. XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=791135385adec8b41cfd714a2f801de59fa1ab6e and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=4ff6903d6981145b7e41bb62b302d86643e87aa2

AurA + N-Myc band

protein1	res1	protein2	res2	CSM
AurA	143	AurA	171	4
AurA	124	AurA	162	2
AurA	124	AurA	171	2
AurA	124	AurA	250	2
AurA	125	NMyc	125	10
AurA	171	NMyc	125	5
AurA	143	NMyc	125	3
AurA	227	NMyc	125	3
AurA	227	NMyc	133	3
AurA	284	NMyc	125	3
AurA	224	NMyc	125	2
AurA	224	NMyc	133	2
AurA	227	NMyc	52	2

protein1	res1	protein2	res2	CSM
AurA	365	monolink	H2O	6
AurA	125	monolink	H2O	4
AurA	171	monolink	H2O	4
AurA	391	monolink	H2O	3
AurA	143	monolink	H2O	2
AurA	156	monolink	H2O	2
AurA	224	monolink	H2O	2
AurA	250	monolink	H2O	2
AurA	258	monolink	H2O	2
AurA	396	monolink	H2O	2

XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=a978e92c3aac66778280a17473908558b552dfcf and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=08724e5003afc7fd870189f8c5d5c6ffe9a4a1ac

AurA band

protein1	res1	protein2	res2	CSM
AurA	143	AurA	171	26
AurA	118	AurA	250	6
AurA	118	AurA	156	5
AurA	250	AurA	250	3
AurA	365	AurA	369	10
AurA	143	AurA	156	2
AurA	143	AurA	258	2
AurA	143	AurA	224	3
AurA	143	AurA	166	2
AurA	143	AurA	284	2
AurA	233	AurA	250	3
AurA	125	AurA	141	2
AurA	166	AurA	171	2
AurA	171	AurA	284	2
AurA	118	AurA	249	3
AurA	124	AurA	156	3
AurA	326	AurA	339	3
AurA	124	AurA	171	3
AurA	238	AurA	249	2
AurA	141	AurA	143	4
AurA	143	AurA	227	4
AurA	141	AurA	171	6
AurA	124	AurA	162	2
AurA	124	AurA	250	2
AurA	125	AurA	250	4

protein1	res1	protein2	res2	CSM
AurA	125	monolink	H2O	9
AurA	143	monolink	H2O	8
AurA	156	monolink	H2O	3
AurA	162	monolink	H2O	4
AurA	171	monolink	H2O	10
AurA	224	monolink	H2O	6
AurA	249	monolink	H2O	3
AurA	250	monolink	H2O	4
AurA	337	monolink	H2O	3
AurA	365	monolink	H2O	27
AurA	369	monolink	H2O	7
AurA	389	monolink	H2O	2
AurA	396	monolink	H2O	11
AurA	401	monolink	H2O	5

5.2.2.3 Triplicate in-gel digested AurA KD Apo

Monolinks and crosslinks shown in the same table. XiNet diagram for crosslinks/monolinks with two or more CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=ee3e33f20b419abc3df761f1342c730856500125

res1	res2	CSM	z-score	>50 %
365	Mono	215	4.44	1
171	Mono	81	1.37	1
309	Mono	34	0.29	1
227	Mono	26	0.11	1
166	Mono	19	-0.05	1
224	Mono	19	-0.05	1
283	Mono	17	-0.10	1
258	Mono	14	-0.17	1
162	Mono	12	-0.21	1
278	Mono	11	-0.23	1
369	Mono	11	-0.23	1
249	Mono	9	-0.28	0
326	Mono	9	-0.28	0
401	Mono	8	-0.30	0
250	Mono	7	-0.33	0
398	Mono	4	-0.39	0
148	Mono	2	-0.44	0
284	Mono	2	-0.44	0

type	res1	res2	CSM	norm	>50 %
Intraprotein	162	258	5	7.40	1
Intraprotein	143	162	6	5.54	1
Intraprotein	156	162	6	5.54	1
Intraprotein	227	326	4	1.60	1
Intraprotein	124	156	5	1.10	1
Intraprotein	309	365	7	0.67	1
Intraprotein	118	250	4	-0.31	1
Intraprotein	124	250	4	-0.31	1
intrapep	365	369	7	-0.84	1
intrapep	118	125	3	-1.52	1
Intraprotein	118	156	2	-2.84	0
Intraprotein	162	171	7	-3.02	0
intrapep	389	391	2	-3.25	0
intrapep	226	227	5	-4.86	0
Intraprotein	143	171	16	-5.49	0
Intraprotein	124	166	2	-6.26	0
intrapep	166	171	6	-8.15	0
Intraprotein	309	369	6	-8.24	0

5.2.2.4 Triplicate in-gel digested AurA KD + TPX2 1-43

XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=bb777a1de9ee694f904104bf22b2afc0b96fb8bc and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=f6989a32bfa06f5d55325d85f5aed59910dc9ac4

res1	res2	CSM	z-score	>50 %	type	protein1	res1	protein1	res2	CSM	norm	>50 %
365	Mono	174	4.19	1	Interaprotein	AurA	118	TPX2	-2	28	14.51	1
171	Mono	43	0.71	1	intrapep	AurA	226	AurA	227	4	1.49	1
227	Mono	30	0.36	1	intrapep	AurA	166	AurA	171	4	0.76	1
309	Mono	12	-0.12	1	intrapep	AurA	365	AurA	369	4	0.37	1
369	Mono	11	-0.14	1	Interaprotein	AurA	156	TPX2	-2	6	0.14	1
401	Mono	11	-0.14	1	Interaprotein	AurA	118	TPX2	5	5	-0.52	1
224	Mono	10	-0.17	1	Intraprotein	AurA	156	AurA	162	5	-1	1
162	Mono	8	-0.22	1	Intraprotein	AurA	124	AurA	156	4	-1.17	1
250	Mono	4	-0.33	1	Interaprotein	AurA	171	TPX2	5	9	-1.28	1
278	Mono	4	-0.33	0	Intraprotein	AurA	143	AurA	171	10	-1.69	1
258	Mono	3	-0.36	0	Interaprotein	AurA	124	TPX2	-2	3	-1.82	0
283	Mono	3	-0.36	0	Interaprotein	AurA	124	TPX2	5	3	-1.82	0
326	Mono	3	-0.36	0	intrapep	AurA	389	AurA	391	3	-1.93	0
396	Mono	3	-0.36	0	intrapep	AurA	118	AurA	125	2	-2.48	0
249	Mono	2	-0.38	0	Interaprotein	AurA	143	TPX2	-2	2	-2.48	0
388	Mono	2	-0.38	0	Interaprotein	AurA	284	TPX2	5	2	-2.48	0
398	Mono	2	-0.38	0	Interaprotein	AurA	166	TPX2	-2	2	-2.62	0
					Intraprotein	AurA	143	AurA	162	3	-3.54	0
					Interaprotein	AurA	171	TPX2	-2	35	-11.7	0

Table of Cα-Cα crosslink distances on selected structures

type	protein1	res1	protein2	res2	CSM	norm	1ol5 / Å	4c3p / Å	Change / Å
Interaprotein	AurA	118	TPX2	-2	28	14.51	14.2	13.6	-0.6
intrapep	AurA	226	AurA	227	4	1.49	8.6	8.3	-0.3
intrapep	AurA	166	AurA	171	4	0.76	8.6	8.4	-0.2
intrapep	AurA	365	AurA	369	4	0.37	12.2	8.7	-3.5
Interaprotein	AurA	156	TPX2	-2	6	0.14	46.8	47.4	0.6
Interaprotein	AurA	118	TPX2	5	5	-0.52	37.9	45.5	7.6
Intraprotein	AurA	156	AurA	162	5	-1.00	17.9	17.9	0
Intraprotein	AurA	124	AurA	156	4	-1.17	6.8	6.6	-0.2
Interaprotein	AurA	171	TPX2	5	9	-1.28	19.1	19.2	0.1
Intraprotein	AurA	143	AurA	171	10	-1.69	24.5	14.6	-9.9
Interaprotein	AurA	124	TPX2	-2	3	-1.82	25.6	33.5	7.9
Interaprotein	AurA	124	TPX2	5	3	-1.82	11.7	10.9	-0.8
intrapep	AurA	389	AurA	391	3	-1.93	3.9	3.8	-0.1
intrapep	AurA	118	AurA	125	2	-2.48	10.8	11.3	0.5
Interaprotein	AurA	143	TPX2	-2	2	-2.48	32.8	32.8	0
Interaprotein	AurA	284	TPX2	5	2	-2.48	44.8	43.4	-1.4
Interaprotein	AurA	166	TPX2	-2	2	-2.62	26.9	25.9	-1
Intraprotein	AurA	143	AurA	162	3	-3.54	12.8	12.4	-0.4
Interaprotein	AurA	171	TPX2	-2	35	-11.7	28.9	33.1	4.2

Table of SASD crosslink distances on selected structures

source	type	protein1	res1	protein2	res2	CSM	norm	1ol5	4c3p
AurA_N3_txp2	Interaprotein	AurA	118	TPX2	-2	28	14.51	44.8	43.4
AurA_N3_txp2	intrapep	AurA	226	AurA	227	4	1.49	3.9	3.8
AurA_N3_txp2	intrapep	AurA	166	AurA	171	4	0.76	12.2	8.7
AurA_N3_txp2	intrapep	AurA	365	AurA	369	4	0.37	8.6	8.3
AurA_N3_txp2	Interaprotein	AurA	156	TPX2	-2	6	0.14	19.1	19.2
AurA_N3_txp2	Interaprotein	AurA	118	TPX2	5	5	-0.52	28.9	33.1
AurA_N3_txp2	Intraprotein	AurA	156	AurA	162	5	-1.00	8.6	8.4
AurA_N3_txp2	Intraprotein	AurA	124	AurA	156	4	-1.17	17.9	17.9
AurA_N3_txp2	Interaprotein	AurA	171	TPX2	5	9	-1.28	26.9	25.9
AurA_N3_txp2	Intraprotein	AurA	143	AurA	171	10	-1.69	24.5	14.6
AurA_N3_txp2	Interaprotein	AurA	124	TPX2	-2	3	-1.82	32.8	32.8
AurA_N3_txp2	Interaprotein	AurA	124	TPX2	5	3	-1.82	46.8	47.4
AurA_N3_txp2	intrapep	AurA	389	AurA	391	3	-1.93	12.8	12.4
AurA_N3_txp2	intrapep	AurA	118	AurA	125	2	-2.48	6.8	6.6
AurA_N3_txp2	Interaprotein	AurA	143	TPX2	-2	2	-2.48	25.6	33.5
AurA_N3_txp2	Interaprotein	AurA	284	TPX2	5	2	-2.48	11.7	10.9
AurA_N3_txp2	Interaprotein	AurA	166	TPX2	-2	2	-2.62	37.9	45.5
AurA_N3_txp2	Intraprotein	AurA	143	AurA	162	3	-3.54	10.8	11.3
AurA_N3_txp2	Interaprotein	AurA	171	TPX2	-2	35	-11.70	14.2	13.6

5.2.2.5 Triplicate in-gel digested AurA KD + taggless N-Myc 1-137

XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=bb777a1de9ee694f904104bf22b2afc0b96fb8bc and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=f6989a32bfa06f5d55325d85f5aed59910dc9ac4

type	protein1	res1	res2	CSM	z-score
Mono	AurA	365	Mono	104	4.36
Mono	AurA	171	Mono	52	1.91
Mono	AurA	227	Mono	34	1.07
Mono	AurA	369	Mono	20	0.41
Mono	AurA	309	Mono	12	0.03
Mono	AurA	162	Mono	7	-0.20
Mono	AurA	224	Mono	6	-0.25
Mono	AurA	283	Mono	6	-0.25
Mono	AurA	401	Mono	6	-0.25
Mono	AurA	278	Mono	5	-0.29
Mono	AurA	326	Mono	5	-0.29
Mono	AurA	398	Mono	5	-0.29
Mono	AurA	166	Mono	4	-0.34
Mono	AurA	258	Mono	4	-0.34
Mono	AurA	249	Mono	3	-0.39
Mono	AurA	250	Mono	3	-0.39
Mono	AurA	389	Mono	2	-0.43
Mono	N-Myc	133	Mono	10	-0.06
Mono	N-Myc	52	Mono	6	-0.25
Mono	N-Myc	-3	Mono	3	-0.39

type	protein1	res1	protein2	res2	CSM	z-score	norm
Interaprotein	AurA	143	AurA	171	6	2.26	-2.26
Interaprotein	AurA	124	AurA	156	2	0.00	0.00
Interaprotein	AurA	156	AurA	162	2	0.00	0.00
intrapep	AurA	389	AurA	391	2	0.00	0.00
Interaprotein	AurA	171	N-Myc	125	7	2.83	-2.82
Interaprotein	AurA	124	N-Myc	125	2	0.00	0.00
Interaprotein	AurA	227	N-Myc	125	2	0.00	0.00

5.2.2.6 Triplicate in-gel digested AurA KD + taggless N-Myc 22-89

XiNet diagram for all crosslinks and monolinks can be found at https://crosslinkviewer.org/uploaded.php?uid=a12a3f68b72d6e5f34a9c388831b2288aa27a670 and for CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=ac63a4bb35331f3de8e787524a6d1423672335e4

type	protein1	res1	res2	CSM	norm	>50%
Mono	AurA	365	Mono	108	3.86	1
Mono	AurA	171	Mono	45	1.29	1
Mono	AurA	227	Mono	24	0.44	1
Mono	AurA	309	Mono	12	-0.05	1
Mono	AurA	369	Mono	9	-0.17	1
Mono	AurA	278	Mono	8	-0.21	1
Mono	AurA	162	Mono	7	-0.25	1
Mono	AurA	401	Mono	7	-0.25	1
Mono	AurA	283	Mono	6	-0.29	1
Mono	AurA	326	Mono	5	-0.33	
Mono	AurA	166	Mono	4	-0.38	
Mono	AurA	396	Mono	4	-0.38	
Mono	AurA	249	Mono	2	-0.46	
Mono	AurA	250	Mono	2	-0.46	
Mono	AurA	258	Mono	2	-0.46	
Mono	AurA	398	Mono	2	-0.46	
Mono	N-Myc	52	Mono	2	-0.46	

type	protein1	res1	protein1	res2	CSM	norm	>50%
Interaprotein	AurA	156	AurA	162	3	4.9	1
Interaprotein	AurA	118	AurA	250	2	0.0	1
Interaprotein	AurA	124	AurA	156	2	0.0	1
Interaprotein	AurA	162	AurA	258	2	0.0	1
intrapep	AurA	166	AurA	171	2	0.0	1
Interaprotein	AurA	162	AurA	171	3	-2.0	
intrapep	AurA	226	AurA	227	3	-2.8	
Interaprotein	AurA	143	AurA	171	7	-4.8	

5.2.2.7 Triplicate in-gel digested N-Myc 1-137 Apo

Monolinks and crosslinks shown in the same table. XiNet diagram for crosslinks and monolinks with two or more CSMs greater than 2 can be found at https://crosslinkviewer.org/uploaded.php?uid=afb0c994b7e16fc763aa15a2c5a92709d17f1e90

protein 1	res1	protein 2	res2	CSM
N-Myc	52	N-Myc	58	4
N-Myc	52	N-Myc	64	2
N-Myc	52	N-Myc	64	2
N-Myc	58	N-Myc	64	2
N-Myc	118	N-Myc	125	4
N-Myc	125	N-Myc	125	12
N-Myc	125	N-Myc	133	3
N-Myc	133	N-Myc	133	7

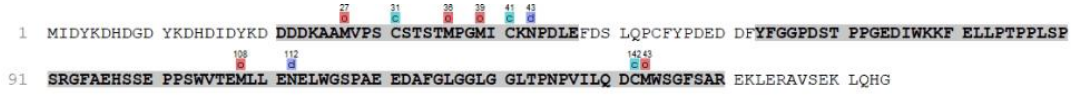
protein 1	res1	protein 2	res 2	CSM
N-Myc	-3	monolink	H2O	38
N-Myc	5	monolink	H2O	7
N-Myc	6	monolink	H2O	4
N-Myc	15	monolink	H2O	3
N-Myc	52	monolink	H2O	39
N-Myc	58	monolink	H2O	2
N-Myc	62	monolink	H2O	14
N-Myc	64	monolink	H2O	19
N-Myc	71	monolink	H2O	8
N-Myc	72	monolink	H2O	5
N-Myc	118	monolink	H2O	2
N-Myc	121	monolink	H2O	3
N-Myc	125	monolink	H2O	5

5.2.3 Non-crosslink peptide coverage searches

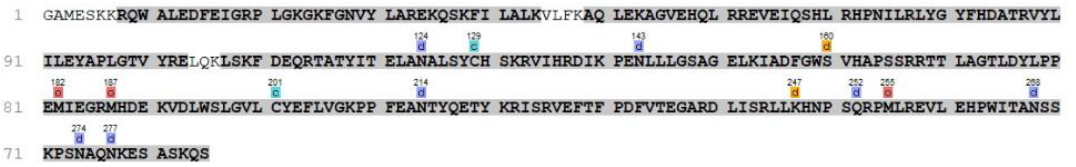
5.2.3.1 Single in-solution digested AurA+ N-Myc

From the measurement referred to in section 2.4.1.3.1. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

FLAG tag N-Myc 1-137 coverage



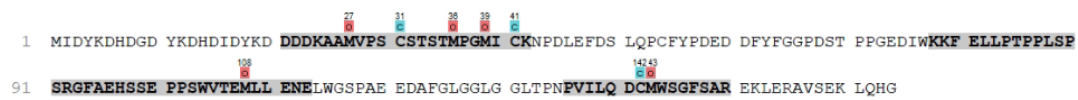
AurA 119-403 coverage



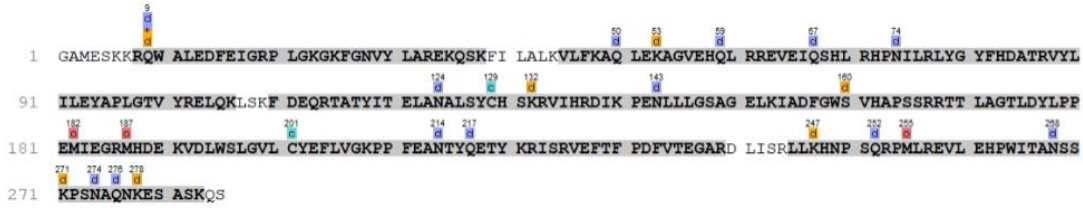
5.2.3.2 In-gel digested AurA+ N-Myc

From the measurement referred to in section 2.4.1.3.2. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

FLAG tag N-Myc 1-137 coverage



AurA 119-403 coverage



5.2.3.3 In-gel digested AurA Apo

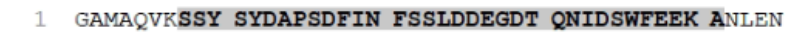
From the measurement referred to in section 2.4.2.2.1. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

AurA 119-403 coverage

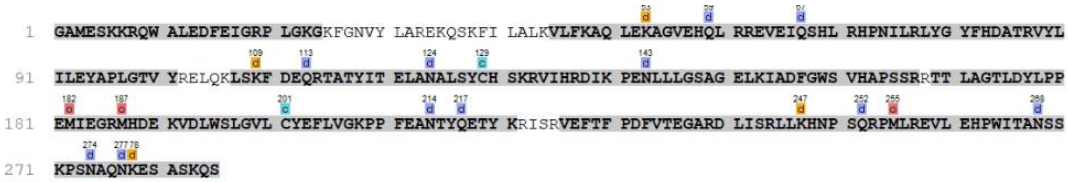
5.2.3.4 In-gel digested AurA+TPX2

From the measurement referred to in section 2.4.2.3.1. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

TPX2 3-43 coverage



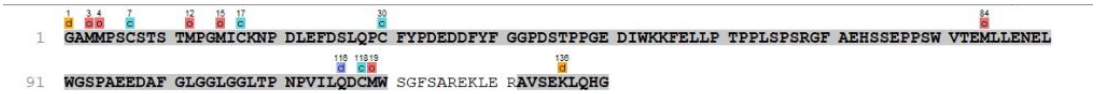
AurA 119-403 coverage



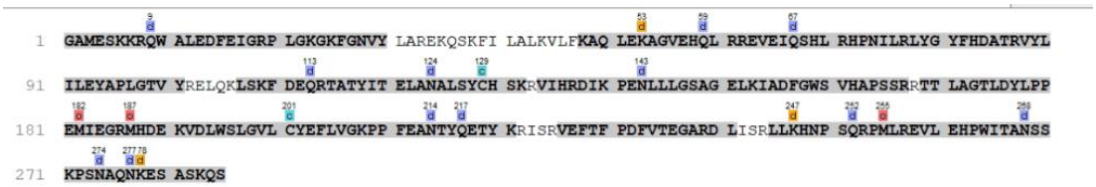
5.2.3.5 In-gel digested AurA+N-Myc 1-137

From the measurement referred to in section 2.4.3.1.1. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

Tagless N-Myc 1-137



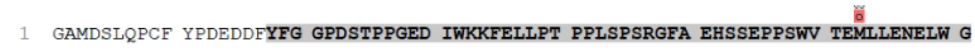
AurA 119-403 coverage



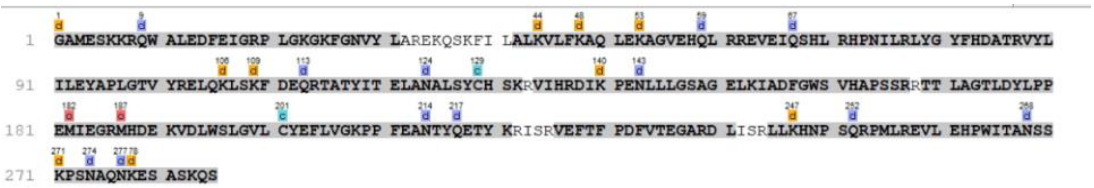
5.2.3.6 In-gel digested AurA+N-Myc22-89

From the measurement referred to in section 2.4.3.1.2. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

Tagless N-Myc 22-89



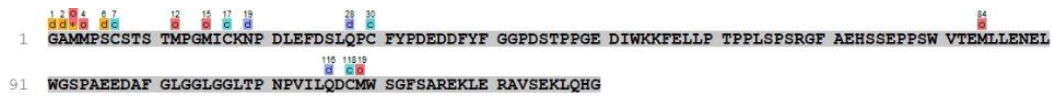
AurA 119-403 coverage



5.2.3.7 In-solution digestest Tagless N-Myc 1-137 apo

From the measurement referred to in section 2.4.4. Diagrams show residues found from peptides from a 1 % FDR PEAKs search.

Tagless N-Myc 1-137



5.2.4 Crosslink distances generated and processed

5.2.4.1 Aurora A

5.2.4.1.1 AurA Apo

Crosslinked table with Cα-Cα distance measured using crosslinks generated from the measurement referred to in section 2.4.2.2.1.

Residue 1	Residue 2	count	CA distance
118	250	4	36.6
118	156	2	34.4
124	156	5	25.6
124	250	4	25.5
124	171	1	21.3
309	391	1	19.1
162	171	7	18.6
141	171	1	18.5
162	258	5	18.1
156	162	6	17.8
124	166	2	15.1
143	171	16	12.8
143	162	6	11.6
309	365	7	10
227	326	4	8.5
224	227	1	8.4
309	369	6	7.4
224	226	1	5.6

monolinked	count
365	86
171	39
309	22
166	17
283	13
278	11
162	10
227	10
258	10
249	7
250	7
369	7
224	5
401	4
284	2
141	1
326	1
339	1
389	1
391	1

5.2.4.1.2 With TPX2

Crosslinked table with Cα-Cα distance measured constructed 1OL5 and 4CEP models referred to in section 2.4.2.3.5.

protein 1	res 1	protein 2	res 2	count	CA distance 1ol5	CA distance 4cep	change
AurA	156	TPX2	1	6	44.8	47.4	2.6
AurA	224	AurA	401	1	36.3	37	0.7
AurA	124	TPX2	1	3	32.5	45.5	13
AurA	283	TPX2	1	1	32	39.8	7.8
AurA	171	AurA	342	2	31.6	33.9	2.3
AurA	283	TPX2	5	2	30.7	30.2	-0.5
AurA	119	TPX2	1	28	29	43.4	14.4
AurA	143	TPX2	1	2	27.9	25.9	-2
AurA	124	TPX2	5	3	25.2	33.5	8.3
AurA	124	AurA	156	4	24.3	14.6	-9.7
AurA	119	TPX2	5	5	23.7	32.8	9.1
AurA	166	TPX2	1	2	18.9	19.2	0.3
AurA	156	AurA	162	5	18	17.9	-0.1
AurA	171	TPX2	1	35	15.6	13.6	-2
AurA	143	AurA	171	10	13.1	12.4	-0.7
AurA	143	AurA	162	3	10.7	11.3	0.6
AurA	171	TPX2	5	9	9.6	10.9	1.3
AurA	224	AurA	227	1	8.9	8.3	-0.6
AurA	224	AurA	226	1	5.7	5.5	-0.2

protein 1	res 1	protein 2	res 2	count
AurA	365	monolink	H2O	58
AurA	171	monolink	H2O	20
AurA	227	monolink	H2O	15
AurA	309	monolink	H2O	7
AurA	401	monolink	H2O	7
AurA	162	monolink	H2O	5
AurA	278	monolink	H2O	4
AurA	250	monolink	H2O	3
AurA	283	monolink	H2O	3
AurA	369	monolink	H2O	3
AurA	224	monolink	H2O	2
AurA	258	monolink	H2O	2
AurA	249	monolink	H2O	1
AurA	396	monolink	H2O	1

Table showing the changes in monolink count between crosslinking measurement AurA APO and AurA + TPX2 1-43 after alignment

		AurA APO	AurA + TPX2	
protein1	res1	actual	actual	change
AurA	162	10	5	-5
AurA	166	17	0	-17
AurA	171	39	20	-19
AurA	224	5	2	-3
AurA	227	10	15	5
AurA	249	7	0	-7
AurA	250	7	0	-7
AurA	258	10	2	-8
AurA	278	11	4	-7
AurA	283	13	3	-10
AurA	309	22	7	-15
AurA	365	86	58	-28
AurA	369	7	3	-4
AurA	401	4	7	3

Table showing the changes in crosslinking count between crosslinking measurement AurA APO and AurA + TPX2 1-43 after alignment

				AurA APO	AurA + TPX2	
protein1	res1	protein2	res2	actual	actual	change
AurA	119	AurA	250	4	0	-4
AurA	124	AurA	156	5	4	-1
AurA	124	AurA	250	4	0	-4
AurA	143	AurA	162	6	3	-3
AurA	143	AurA	171	16	10	-6
AurA	156	AurA	162	6	5	-1
AurA	162	AurA	171	7	0	-7
AurA	162	AurA	258	5	0	-5
AurA	171	AurA	342	0	2	2
AurA	227	AurA	326	4	0	-4
AurA	309	AurA	365	7	0	-7
AurA	309	AurA	369	6	0	-6
AurA	119	TPX2	1	0	28	28
AurA	119	TPX2	5	0	5	5
AurA	124	TPX2	1	0	3	3
AurA	124	TPX2	5	0	3	3
AurA	143	TPX2	1	0	2	2
AurA	156	TPX2	1	0	6	6
AurA	166	TPX2	1	0	2	2
AurA	171	TPX2	1	0	35	35
AurA	171	TPX2	5	0	9	9
AurA	283	TPX2	5	0	2	2

5.2.4.1.3 With N-Myc 22-89

Crosslinked table with Cα-Cα distance measured from constructed 5g1x model from which crosslinking experiment referred to in section 2.4.3.1.2.

protein1	res1	protein2	res2	count	CA distance
AurA	119	AurA	250	2	37.4
AurA	124	AurA	171	1	33.2
AurA	171	AurA	284	1	29.6
AurA	124	AurA	250	1	28.2
AurA	124	AurA	166	1	25.6
AurA	162	AurA	258	2	20

protein1	res1	protein2	res2	count
AurA	365	monolink	H2O	51
AurA	171	monolink	H2O	24
AurA	227	monolink	H2O	19
AurA	309	monolink	H2O	9
AurA	278	monolink	H2O	8
AurA	283	monolink	H2O	6

AurA	124	AurA	156	2	19.5
AurA	162	AurA	171	3	18.8
AurA	156	AurA	162	3	17.8
AurA	143	AurA	171	7	12.8
AurA	143	AurA	162	1	11.1
AurA	226	AurA	224	1	5.7

AurA	369	monolink	H2O	6
AurA	162	monolink	H2O	5
AurA	401	monolink	H2O	5
AurA	166	monolink	H2O	4
AurA	326	monolink	H2O	3
AurA	250	monolink	H2O	2
AurA	258	monolink	H2O	2
NMyc	52	monolink	H2O	2
AurA	249	monolink	H2O	1
AurA	389	monolink	H2O	1
NMyc	62	monolink	H2O	1

5.2.4.1.4 With N-Myc 1-137

Crosslinked table with Cα-Cα distance measured from constructed 5g1x model from which crosslinking experiment referred to in section 2.4.3.1.1.

protein1	res1	protein2	res2	count	CA distance
AurA	124	NMyc	125	2	90.7
AurA	162	NMyc	125	1	71.6
AurA	227	NMyc	125	2	68
AurA	226	NMyc	125	1	65.9
AurA	171	NMyc	125	7	57.9
AurA	119	AurA	250	1	37.4
AurA	124	AurA	156	2	19.5
AurA	162	AurA	171	1	18.8
AurA	156	AurA	162	2	17.8
AurA	143	AurA	171	6	12.8
AurA	143	AurA	162	1	11.1
AurA	309	AurA	369	1	7.4

protein1	res1	protein2	res2	count
AurA	365	monolink	H2O	41
AurA	171	monolink	H2O	26
AurA	227	monolink	H2O	25
AurA	309	monolink	H2O	9
AurA	224	monolink	H2O	6
AurA	283	monolink	H2O	6
AurA	278	monolink	H2O	5
AurA	162	monolink	H2O	4
AurA	166	monolink	H2O	4
AurA	369	monolink	H2O	4
AurA	250	monolink	H2O	3
AurA	258	monolink	H2O	3
AurA	401	monolink	H2O	3
AurA	249	monolink	H2O	2
AurA	141	monolink	H2O	1
AurA	326	monolink	H2O	1
AurA	388	monolink	H2O	1
AurA	389	monolink	H2O	1
AurA	391	monolink	H2O	1
AurA	398	monolink	H2O	1
NMyc	133	monolink	H2O	10
NMyc	52	monolink	H2O	4
NMyc	1	monolink	H2O	1
NMyc	58	monolink	H2O	1
NMyc	64	monolink	H2O	1

Table showing the changes in monolink count between crosslinking measurement AurA APO, AurA + N-Myc 1-137 and AurA + N-Myc 22-89 after alignment

		AurA APO		Nmyc TAD APO		AurA + 1-137		AurA + 22-89	
protein1	res1	change	actual	change	actual	change	actual	change	actual
AurA	162		10			6	4	5	5
AurA	166		17			13	4	13	4
AurA	171		39			13	26	15	24
AurA	224		5			-1	6	5	
AurA	227		10			-15	25	-9	19
AurA	249		7			5	2	7	
AurA	250		7			4	3	7	
AurA	258		10			7	3	10	
AurA	278		11			6	5	3	8
AurA	283		13			7	6	7	6
AurA	309		22			13	9	13	9
AurA	326					0		-3	3
AurA	365		86			45	41	35	51

AurA	369		7			3	4	1	6
AurA	401		4			1	3	-1	5
NMyc	-3				21				
NMyc	5				6				
NMyc	6				3				
NMyc	15				2				
NMyc	52				22		4		2
NMyc	62				4				
NMyc	64				5				
NMyc	71				2				
NMyc	72				2				
NMyc	121				2				
NMyc	125				5				
NMyc	133						10		

Table showing the changes in crosslink count between crosslinking measurement AurA APO, AurA + N-Myc 1-137 and AurA + N-Myc 22-89 after alignment

				AurA APO		Nmyc TAD APO		AurA + 1-137		AurA + 22-89	
protein1	res1	protein2	res2	change	actual	change	actual	change	actual	change	actual
AurA	119	AurA	250		2			2		-2	4
AurA	124	AurA	156		5			3	2	3	2
AurA	124	AurA	250		4			4		4	
AurA	143	AurA	162		6			6		6	
AurA	143	AurA	171		16			10	6	9	7
AurA	156	AurA	162		6			4	2	3	3
AurA	162	AurA	171		7			7		4	3
AurA	162	AurA	258		5			5		3	2
AurA	171	NMyc	125					-7	7	0	
AurA	227	AurA	326		4			4		4	
AurA	227	NMyc	125					-2	2	0	
AurA	309	AurA	365		7			7		7	
AurA	309	AurA	369		6			6		6	
NMyc	52	NMyc	64				2				
NMyc	58	NMyc	64				2				
NMyc	118	NMyc	125				4				
NMyc	125	NMyc	125				12				
NMyc	125	NMyc	133				3				
NMyc	133	NMyc	133				7				

5.2.4.2 N-Myc

5.2.4.2.1 Tagless N-Myc 1-137 Apo

Crosslinked table with Ca-Ca distance measured from alphafold2 structure referenced in section 2.4.4.

protein1	res1	protein2	res2	count	CA distance
NMyc	64	NMyc	131	1	47.9
NMyc	1	NMyc	15	1	39.2
NMyc	52	NMyc	64	2	28.5
NMyc	58	NMyc	64	2	16.4
NMyc	118	NMyc	125	4	14.2
NMyc	125	NMyc	133	3	12.3
NMyc	133	NMyc	133	7	0
NMyc	125	NMyc	125	12	0

protein1	res1	protein2	res2	count
NMyc	52	monolink	H2O	22
NMyc	1	monolink	H2O	21
NMyc	5	monolink	H2O	6
NMyc	64	monolink	H2O	5
NMyc	125	monolink	H2O	5
NMyc	62	monolink	H2O	4
NMyc	6	monolink	H2O	3
NMyc	15	monolink	H2O	2
NMyc	71	monolink	H2O	2
NMyc	72	monolink	H2O	2
NMyc	121	monolink	H2O	2
NMyc	7	monolink	H2O	1
NMyc	58	monolink	H2O	1
NMyc	131	monolink	H2O	1
NMyc	133	monolink	H2O	1

protein1	res1	protein2	res2	count	CSM
NMyc	-3	-	Mono	21	2.40
NMyc	5	-	Mono	6	0.12
NMyc	6	-	Mono	3	-0.33
NMyc	7	-	Mono	1	-0.64
NMyc	15	-	Mono	2	-0.49
NMyc	52	-	Mono	22	2.55
NMyc	58	-	Mono	1	-0.64
NMyc	62	-	Mono	4	-0.18
NMyc	64	-	Mono	5	-0.03
NMyc	71	-	Mono	2	-0.49
NMyc	72	-	Mono	2	-0.49
NMyc	121	-	Mono	2	-0.49
NMyc	125	-	Mono	5	-0.03
NMyc	131	-	Mono	1	-0.64
NMyc	133	-	Mono	1	-0.64

source	type	protein1	res1	protein2	res2	CSM	z-score	norm
nmyc_N3_1137	Interaprotein	NMyc	-3	NMyc	15	1	-2.32	1.99
nmyc_N3_1137	Interaprotein	NMyc	52	NMyc	64	2	-2.30	29.64
nmyc_N3_1137	Interaprotein	NMyc	58	NMyc	64	2	-2.30	-19.76
nmyc_N3_1137	Interaprotein	NMyc	64	NMyc	131	1	-2.32	-119.79
nmyc_N3_1137	Interaprotein	NMyc	118	NMyc	125	4	-2.25	-95.62
nmyc_N3_1137	Interaprotein	NMyc	125	NMyc	125	12	-2.06	-2232.54
nmyc_N3_1137	Interaprotein	NMyc	125	NMyc	133	3	-2.27	-117.34
nmyc_N3_1137	Interaprotein	NMyc	133	NMyc	133	7	-2.18	-5.35

5.3 For Chapter 3

5.3.1 Data Tables generated from XLM tools scoring

5.3.1.1 AurA apo

5.3.1.1.1 Input structure scores

PDB	**3UOK**	**Ahanger**	**6hjk**	**2WTV**	**5l8k**	**5EW9**	**1ol7**	**4c3p**	**6c2t**
Parent Construct	123–390	119-403 C290A C393A	119-403 L210C, C290A, and C393A	119-403 L215R T217E R220K	119-403 C290A C393A	123-390	119-403 C290A C393A	122-403	122-403
Crystallographic ligand(s)	Type I SM	ADP nVAR cysteine reactive covalent SM	Type II SM	Type II SM	ADP nVAR	Type I½ SM	ADP	AMPPCP TPX2 1-43	Type I SM
224-287 CαCα / Å	16.5	18.3	20.6	21.2	32.0	35.7	35.8	49.2	52.6
DFG category	DFGinter	DFGin	DFGout	DFGinter	DFGinter	DFGinter	DFGin	DFGin	DFGinter
Dunbrack category	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	BLAminus	Unassigned	Unassigned
Rspine? (275 + 254 + 185 + 196)	Broken, Phe275 and His254 flipped	Broken, Phe275 and His254 rotated around ca	Broken, Phe275 flipped	Broken, Phe275 and His254 flipped	Broken, Phe275 flipped	Broken, Phe275 and His254 flipped	formed	formed	Broken, Phe275 flipped
Salt bridge (162 + 181)	Broken, Phe275 bisects residues	Broken, Trp277 bisects residues	Broken, Lys 162 in nt pocket	Broken, Phe275 bisects residues	Broken, Phe275 bisects residues	Broken, Phe275 bisects residues	formed	formed	Broken, Glu181 held between backbond CO of 275 and 177
172-308-324 Angle / ^{o}	86.1	82.3	87.7	84.7	88.5	73.2	85.0	80.7	88.5
Impact CCS / nm^2	23.16	23.79	23.66	22.95	22.54	22.90	22.56	25.27	25.47
XLMO score	-0.48	-0.80	-0.84	-1.69	0.29	0.22	0.39	1.52	1.40
MNXL score	0.17	-0.07	0.22	-0.02	0.23	0.10	0.35	0.26	0.20
MoDS score	-97.97	-92.80	-100.85	-98.14	-96.14	-92.97	-99.21	-91.64	-90.27
Place by XLMO	6^{th}	7^{th}	8^{th}	9^{th}	4^{th}	5^{th}	3^{rd}	2^{nd}	1^{st}
PDB	3UOK	Syed	6hjk	2WTV	5l8k	5EW9	1ol7	4c3p	6c2t
Taken forward for MD sim	No	No	No	No	No	No	Yes	Yes	yes

5.3.1.1.2 Crosslink lengths

Residue 1	Residue 2	count	4c3p SASD / Å	4c3p Euclidean / Å	1ol7 SASD / Å	1ol7 Euclidean / Å	6c2t SASD / Å	6c2t Euclidean / Å
143	171	16	14.3	12.5	14.9	12.9	11.6	11.2
162	171	7	30	18.8	29.7	18.7	28.5	18.8
309	365	7	23.8	10.4	25.4	10.5	26.2	11.6
143	162	6	12.7	11.2	14	10.7	16	12.1
156	162	6	40.1	17.9	42.6	18.2	45.7	18.2
309	369	6	9.8	7.2	9.2	7.8	6.9	6.2
124	156	5	27.6	18	25	17.6	30.4	18.3
162	258	5	22.8	19.9	22.1	18	21.5	19.1
119	250	4	21.7	18.8	27.1	20.5	22.1	19.8
124	250	4	20.9	18.3	20.8	18.8	21.9	19.2
227	326	4	16.8	8.4	14.8	8.7	16.5	8.7
119	156	2	19.4	15.4	20	15.5	20.1	15.5
124	166	2	27	22.1	27.1	21.1	28.1	21.2
124	171	1	36.9	28.7	37.9	27.6	38.3	28.2
141	171	1	23.5	18.4	22.9	19.1	23.1	18.3
224	226	1	17.5	8.4	13.6	8.5	14.6	8.7
224	227	1	11.2	5.6	7.5	5.7	7.5	5.9
309	391	1	52.5	30.5	23.4	19.1	37.8	29.5

Scoring using all the AurA Apo crosslinks

Model	Matched	NoV	NoNA	MoDS	MNXL	XLMO
AuroraA119403_4c3p_Matt_pT288287_cap_hydrogens	15	3	0	-91.64	0.36	1.52
AuroraA119403_6c2t_pT288287_cap_hydrogens	15	3	0	-90.27	0.30	1.40
AuroraA119403_1ol7_Matt_pT288287_cap_hydrogens	16	2	0	-99.21	0.45	0.39
AuroraA119403_5l8k_pT288287_cap_hydrogens	15	3	0	-96.14	0.33	0.29
AuroraA119403_5ew9_Matt_cap_hydrogens	14	4	0	-92.97	0.20	0.22
AuroraA119403_3UOK__pT288287_cap_hydrogens	15	3	0	-97.97	0.27	-0.48
AuroraA119403_nomyc_syed_pT288287_fixed_cap_hydrogens3	13	5	0	-92.8	0.03	-0.80
AuroraA119403_6hjk_pT288287_cap_hydrogens	16	2	0	-100.85	0.32	-0.84
AuroraA119403_2wtv_b_Matt_pT288287_cap_hydrogens	13	5	0	-98.14	0.08	-1.69

Scoring using AurA Apo crosslinks with more than 2 CSMs

Model	MNXL	Matched	NoV	NoNA	MoDS	XLMO
AuroraA119403_6c2t_pT288287_cap_hydrogens	0.29647	10	1	0	-64.4	1.733
AuroraA119403_4c3p_Matt_pT288287_cap_hydrogens	0.32866	10	1	0	-67.3	1.413
AuroraA119403_5ew9_Matt_cap_hydrogens	0.17856	9	2	0	-68	0.209
AuroraA119403_1ol7_Matt_pT288287_cap_hydrogens	0.3058	10	1	0	-73.5	0.063
AuroraA119403_5l8k_pT288287_cap_hydrogens	0.15018	9	2	0	-68.2	-0.03
AuroraA119403_nomyc_syed_pT288287_fixed_cap_hydrogens3	0.04332	8	3	0	-64.5	-0.08
AuroraA119403_3UOK__pT288287_cap_hydrogens	0.14779	9	2	0	-71.9	-0.76
AuroraA119403_6hjk_pT288287_cap_hydrogens	0.18081	10	1	0	-74.5	-1.03
AuroraA119403_2wtv_b_Matt_pT288287_cap_hydrogens	0.06132	8	3	0	-72.7	-1.52

5.3.1.2 AurA+TPX2

5.3.1.2.1 Input structures scores

PDB	AuroraA119403_1ol7_Matt_pT288287_final_topclusterc0_tpx2.pdb	AuroraA119403_3UOK_pT288287_hydro_old_TPX2.pdb	AuroraA119403_4c3p_Matt_pT288287_final_topclusterc0_tpx2.pdb	unrelaxed_rank_3_model_2_edit_H2.pdb	unrelaxed_rank_4_model_4_edit_H.pdb	AuroraA119403_6vPJ_pT288287_H_TPX2.pdb	AuroraA119403_6c2t_pT288287_final_topclusterc0_tpx2.pdb	AuroraA119403_4c3p_Matt_pT288287_final_topclusterc7_tpx2.pdb
Crystallographic ligand(s)	Type I SM	ADP nVAR cysteine reactive covalent SM	Type II SM	Type II SM	ADP nVAR	Type I½ SM	ADP	AMPPCP TPX2 1-43
224-287 CαCa / Å	16.5	18.3	20.6	21.2	32.0	35.7	35.8	49.2
Impact CCS / nm^2	26.35	26.52	26.65	26.83	27.25	27.75	27.95	28.06
MoDS score	-70.44	-71.79	-77.77	-72.31	-70.33	-67.12	-66.49	-64.41
MNXL score	-0.339	0.007	-0.591	-0.001	-0.073	0.015	-0.018	-0.344
XLMO score	-0.533	0.229	-2.362	0.125	0.227	0.966	0.969	0.377
Place by XLMO using all	7th	4th	8th	6th	5th	2nd	1st	3rd
Place by XLMO using n/3	7th	3rd	8th	2nd	6th	4th	1st	5th
PDB	1ol7 cluster c0	3uok	4c3p cluster c0	Colab 32	Colab 44	6vPJ	6c2t cluster c0	4c3p cluster c7
Taken forward for MD sim					yes		yes	

5.3.1.2.2 Output

Scoring of the output AurA+TPX2 structures using all the AurA+TPX2

Model	Matched	NoV	NoNA	MoDS	XLMO
AuroraA119403_6vPJ_pT288287_H_TPX2	12	7	0	-66.60	0.98
AuroraA119403_6c2t_pT288287_final_topclusterc0_tpx2	12	7	0	-66.23	0.94
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc7_tpx2	10	9	0	-64.14	0.33
Tagless_TPX_AuroraA_C290A_C393A12240_72926_unrelaxed_rank_4_model_4_edit_H	12	7	0	-70.26	0.24
Tagless_TPX_AuroraA_C290A_C393A12240_72926_unrelaxed_rank_3_model_2_edit_H2	12	7	0	-71.80	0.23
AuroraA119403_3UOK_pT288287_hydro_old_TPX2	10	9	0	-71.99	0.23
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc0_tpx2	10	9	0	-70.87	-0.57
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc0_tpx2	9	10	0	-78.94	-2.36

5.3.1.2.1 Output structures scores

PDB	Manual C0	C1	C2	Colab c-	C1	C2	3uok c0	C1	C2
Frames	Manual_chatp2_AA_pT288287_hydro_old_NMyc22-89_cluster0	Manual_chatp2_AA_pT288287_hydro_old_NMyc22-89_cluster1	Manual_chatp2_AA_pT288287_hydro_old_NMyc22-89_cluster2	Aura_122403_nmyc_19_89_Colabfold_model4nohydro_cluster_c0	Aura_122403_nmyc_19_89_Colabfold_model4nohydro_cluster_c1	Aura_122403_nmyc_19_89_Colabfold_model4nohydro_cluster_c2	3UOK_pT288287_hydro_old_NMyc22-89_cluster0	3UOK_pT288287_hydro_old_NMyc22-89_cluster1	3UOK_pT288287_hydro_old_NMyc22-89_cluster2
Fraction	0.679	0.131	0.115	0.421	0.335	0.078	0.514	0.383	0.042
stdev	0.278	0.291	0.301	0.325	0.325	0.395	0.281	0.28	0.393
Central frame	3596	1143	631	4326	2195	880	1538	3945	231
224-287 CαCα / Å	36.7	36.8	37.3	36.7	36.7	36.6	13.5	12.7	14.5
Angle	81.1	79.2	83.4	83.4	84.3	89.2	86.1	86.4	84.2
DFG category	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGinter	DFGinter	DFGinter
Dunbrack category	BLAminus	BLAminus	BLAminus	ABAminus	ABAminus	ABAminus	Unassigned	Unassigned	Unassigned
Impact CCS / nm^2	26.15	26.54	26.49	25.22	25.50	25.37	26.16	26.59	26.66
MoDS score	-90.95	-88.02	-85.09	-85.01	-83.63	-84.44	-97.4	-102.39	-94.26
MNXL score	-0.16	-0.14	-0.12	-0.25	-0.24	-0.27	-0.58	-0.39	-0.48
XLMO score	0.41	0.74	1.05	0.61	0.76	0.56	-1.68	-1.41	-1.04
Place by XLMO	6^{th}	3^{rd}	1^{st}	4^{th}	2^{nd}	5^{th}	9^{th}	8^{th}	7^{th}

5.3.2 Data Tables MD simulation for AurA Apo

5.3.2.1 Pseudo B-factor Table

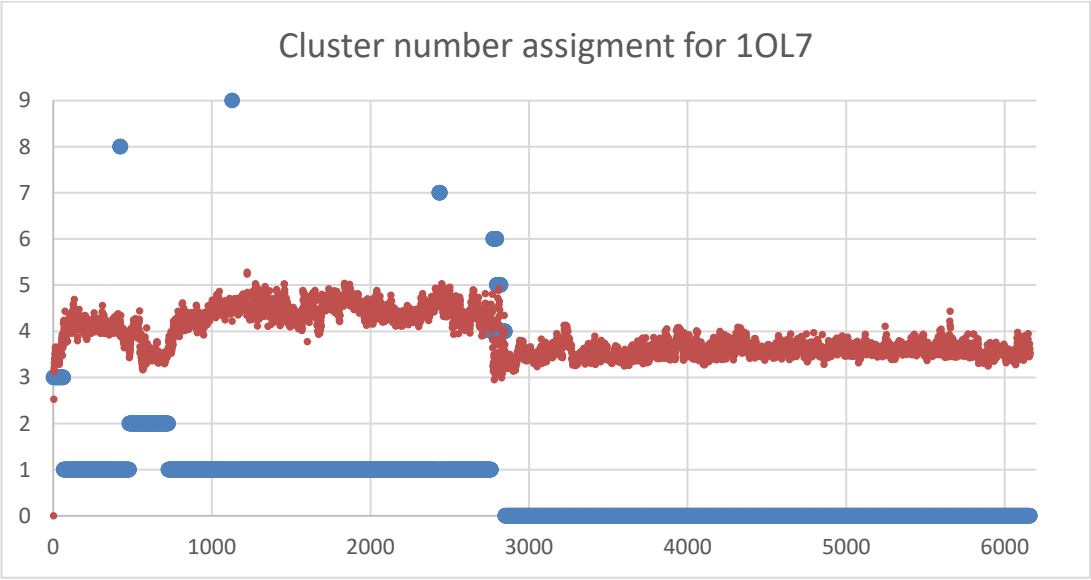
Table showing the Pseudo B-factors for AurA MD simulations

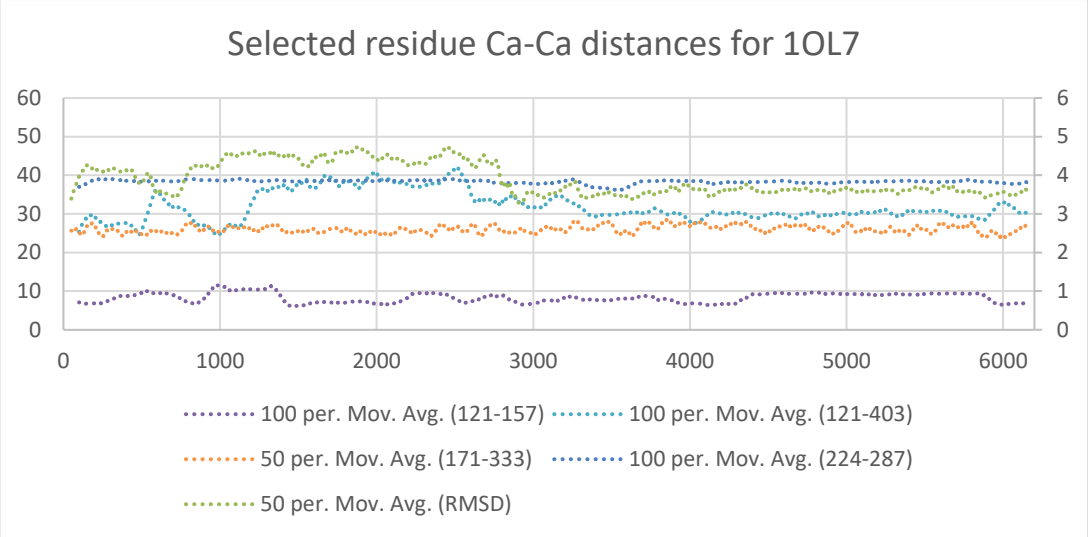
Residue	4c3p	1ol7	6c2t
119	875	1439	1168
120	498	796	729
121	330	228	315
143	385	133	176
203	310	74	133
290	796	99	1625
291	1390	52	1807
292	1288	33	2146
293	874	18	1499
294	1431	17	1040
295	1557	11	1130
296	981	10	910
297	902	9	1263
298	555	14	1168
299	528	15	1111
331	244	33	677
332	324	55	1295
333	399	47	1496
334	392	39	1328
335	475	43	1459
396	608	1018	481
397	668	1754	719
400	997	985	573
401	1414	893	594
402	1596	807	987
403	2144	1104	1579

5.3.2.2 1OL7

Cluster scores for 1OL7 simulation

Cluster	1	2	3	4	5	6	7	8	9	0
Frames	2437	247	64	40	26	22	6	4	2	3311
Fraction	0.396	0.04	0.01	0.006	0.004	0.004	0.001	0.001	0	0.538
stdev	0.294	0.345	0.429	0.537	0.593	0.556	0.527	0.573	0	0.289
Central frame	322	615	26	2824	2805	2781	2435	422	1127	3687
224-287 CαCα / Å	38.5	38.6	36.1	37.3	38.8	38.0	39.5	38.0	31.1	37.2
DFG category	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin
Dunbrack category	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus	BLAminus
Rspine (275 + 254 + 185 + 196)	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed
Salt bridge (162 + 181)	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed	Formed
172-308-324 Angle / °	80.4	78.9	78.5	72.6	75.0	74.0	75.0	78.4	76.2	79.4
Impact CCS / nm^2	22.70	22.92	22.64	23.28	24.11	23.20	22.96	22.98	23.46	22.92
Conformer match	2	2	2	2-3	2-3	2-3	2	2	2-3	2
XLMO score	-0.050	-0.053	0.082	0.043	0.162	0.181	-0.181	0.048	-0.052	-0.26
MNXL score	-1.13	-0.49	-0.36	-0.74	0.30	0.04	-1.54	-1.10	-0.34	0.060
MoDS score	-97.8	-93.9	-97.1	-98.2	-95.5	-97.6	-96.4	-100.5	-93.0	-95.8

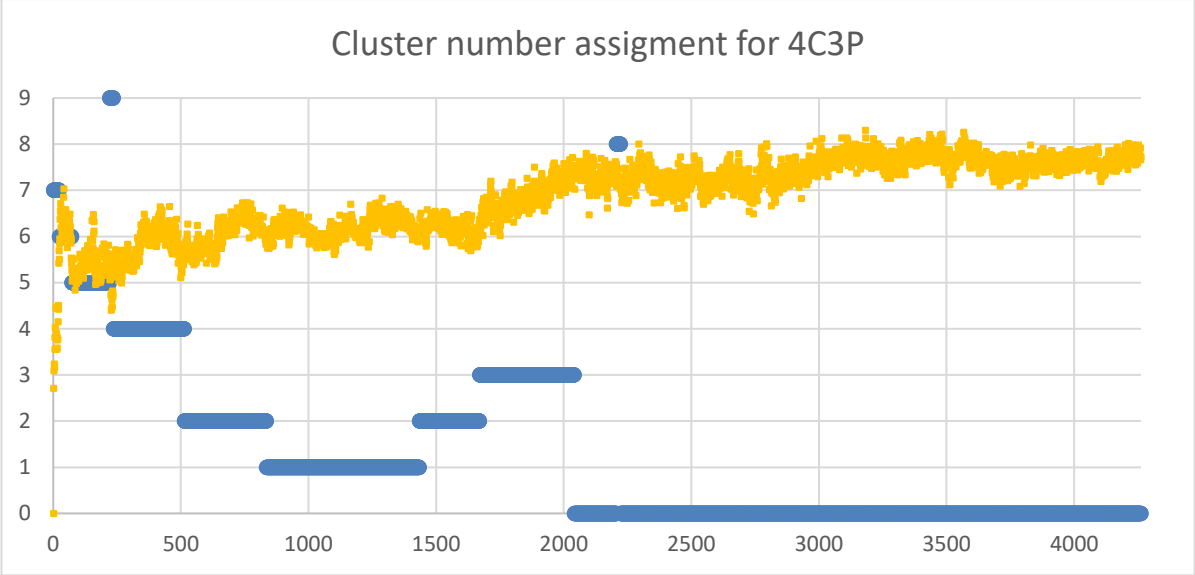


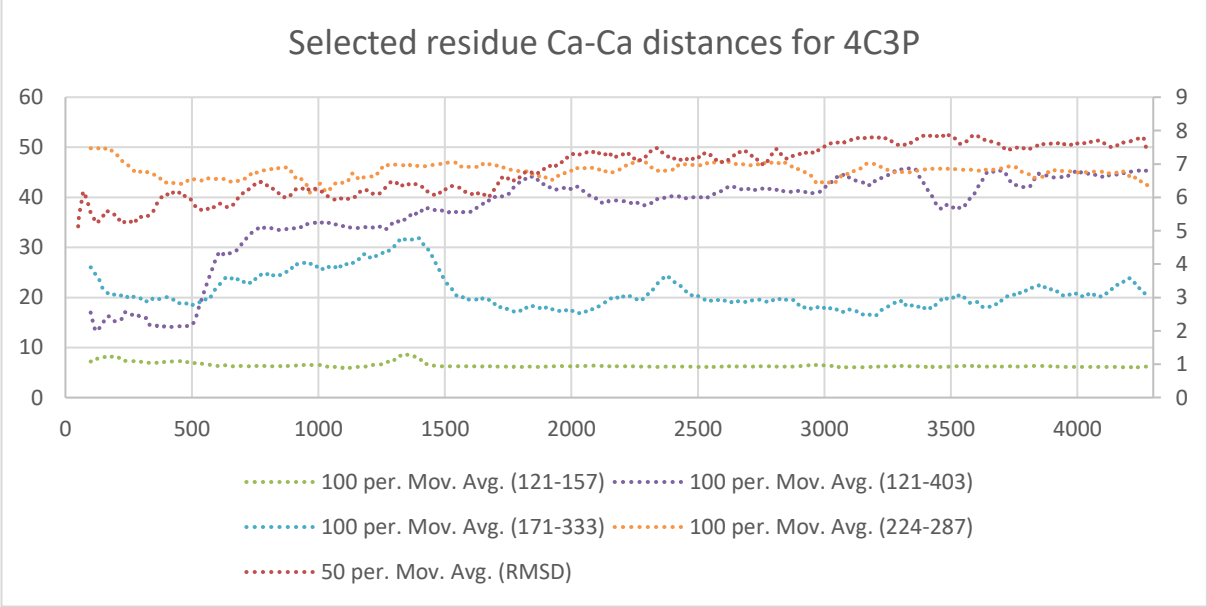


5.3.2.3 4C3P

Cluster scores for 4C3P simulation

Cluster	1	2	3	4	5	6	7	8	9	0
Frames	599	558	371	278	148	49	23	16	15	2205
Fraction	0.141	0.131	0.087	0.065	0.035	0.011	0.005	0.004	0.004	0.517
stdev	0.341	0.356	0.359	0.379	0.432	0.603	0.772	0.627	0.783	0.315
Central frame	1202	750	1898	489	166	46	10	2216	226	3296
224 287 CαCα / Å	47.3	46.6	46.8	45.5	50.5	49.8	50.9	47.8	44.8	44.4
DFG category	?	?	?	DFGin	DFGin	DFGin	DFGin	?	DFGin	?
Dunbrack category	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned
Rspine? (275 + 254 + 185 + 196)	Broken in half, 275+254 rotated down and pi-pi	Broken in half, 275+254 rotated down and pi-pi	Broken in half, 275+254 rotated down and pi-pi	Formed	Formed	Formed, 196 slightly out	Formed	Broken in half, 275+254 rotated down and pi-pi	Formed	Broken in half, 275+254 rotated down and pi-pi
Salt bridge (162 + 181)	Gln177 replaced Glu181 with 181 rotated in solvent	Gln177 replaced Glu181 with 181 rotated in solvent	Gln177 replaced Glu181 with 181 rotated in solvent	Broken, 181 rotated in solvent, 177-276	Broken, 181 rotated in solvent, 177-276	Broken, 181 rotated in solvent, 177-276	Formed	Gln177 replaced Glu181 with 181 rotated in solvent	Gln177 replaced Glu181 with 181 rotated in solvent	Gln177 replaced Glu181 with 181 rotated in solvent
172-308-324 Angle / °	91.8	97.9	104.4	74.6	80.3	90.3	90.2	98.6	91.4	88.9
Impact CCS / nm^2	23.99	24.22	23.63	23.72	24.13	24.58	25.57	24.21	24.59	23.53
Conformer match	2-3	2-3	2-3	2-3	2-3	3	3	2-3	2-3	2-3
XLMO score	0.17	0.26	-1.13	-0.13	1.67	0.34	1.86	-0.84	1.68	-1.68
MNXL score	0.130	0.167	-0.085	0.094	0.262	-0.047	0.272	0.041	0.263	0.018
MoDS score	-95.3	-95.9	-96.7	-96.1	-90.2	-89.0	-89.4	-98.8	-90.2	-103.1



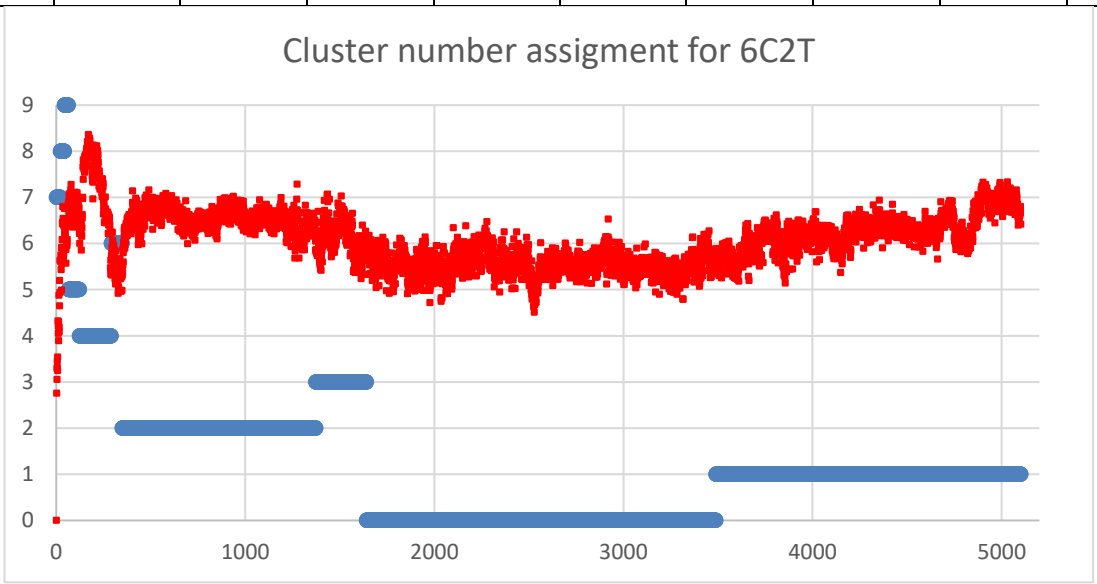


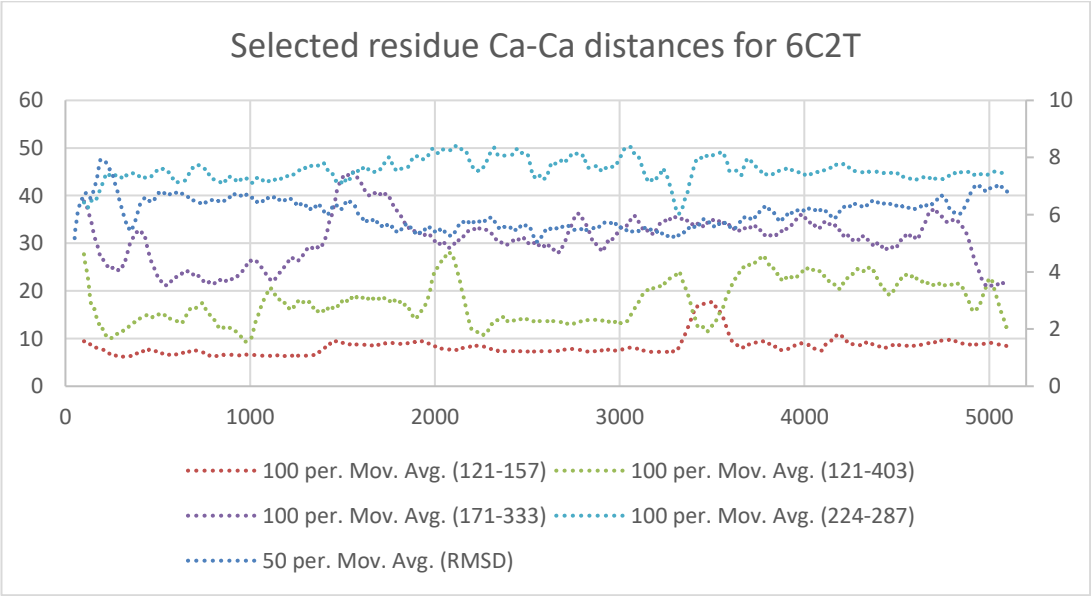
5.3.2.4 6C2T

Cluster scores for 6C2T simulation

Cluster	1	2	3	4	5	6	7	8	9	0
Frames	1613	1024	268	168	59	58	21	21	21	1848
Fraction	0.316	0.201	0.053	0.033	0.012	0.011	0.004	0.004	0.004	0.362
stdev	0.377	0.337	0.404	0.499	0.571	0.535	0.786	0.783	0.724	0.336
Central frame	4323	776	1587	189	107	310	13	35	51	2345
224 287 CαCα / Å	46.0	44.9	48.1	45.7	47.9	44.4	51.0	45.5	47.3	51.8
DFG category	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin	DFGin
Dunbrack category	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned	Unassigned
Rspine? (275 + 254 + 185 + 196)	Broken, close binding of 254-274	Broken	Broken	Broken, Phe275 flipped and 186 H bond to 194	Broken, Phe275 flipped and 186 H bond to 194	Broken, Phe275 flipped and 186 H bond to 194	Broken, Phe275 flipped	Broken, Phe275 flipped	Broken, Phe275 flipped	Broken, Phe275 flipped

Salt bridge (162 + 181)	Broken	Broken	Broken	Broken, Glu181 in solvent	Broken	Broken, Glu181 held between back bone CO of 275 and 177	Broken, Glu181 held between back bone CO of 275 and 177	Broken, Glu181 in solvent	Broken, Glu181 in solvent	Broken, Glu181 held between back bone CO of 275 and 177
172-308-324 Angle / °	91.8	97.9	104.4	74.6	80.3	90.3	90.2	98.6	91.4	88.9
Impact CCS / nm^2	24.16	23.79	24.59	24.18	25.27	24.85	25.92	26.48	25.25	24.67
Conformer match	2-3	2-3	2-3	2-3	3	3	>3	>3	3	3
XLMO score	-0.25	0.48	-1.33	-0.11	1.08	0.82	1.97	-0.55	0.78	-2.03
MNXL score	0.083	0.190	-0.123	0.098	0.151	0.056	0.303	0.085	0.103	-0.071
MoDS score	-96.4	-95.3	-96.8	-96.1	-90.5	-89.2	-89.7	-98.3	-90.8	-102.6





.

5.3.1 AurA Apo Output clusters scores

5.3.1.1 Using All Crosslinks and monolinks

Model	MNXL	Matched	NoV	NoNA	MoDS	XLMO	zMNXL	zMoDS	Place
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc7	0.60	17	1	0	-93.69	2.40	1.70	1.47	1
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc9	0.66	17	1	0	-96.13	2.28	2.18	0.84	2
AuroraA119403_4c3p_Matt_pT288287_cap_hydrogens	0.59	17	1	0	-97.52	1.55	1.58	0.48	3
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc5	0.57	17	1	0	-97.07	1.53	1.43	0.59	4
AuroraA119403_6c2t_pT288287_final_topclusterc7	0.43	16	2	0	-92.76	1.42	0.17	1.71	5
AuroraA119403_6c2t_pT288287_cap_hydrogens	0.53	17	1	0	-96.46	1.34	1.02	0.75	6
AuroraA119403_6c2t_pT288287_final_topclusterc0	0.47	16	2	0	-96.39	0.97	0.52	0.77	7
AuroraA119403_6c2t_pT288287_final_topclusterc2	0.44	16	2	0	-96.46	0.75	0.24	0.75	8
AuroraA119403_6c2t_pT288287_final_topclusterc4	0.42	16	2	0	-96.88	0.50	0.01	0.64	9
AuroraA119403_6c2t_pT288287_final_topclusterc5	0.39	16	2	0	-96.28	0.43	-0.23	0.80	10
AuroraA119403_5l8k_pT288287_cap_hydrogens	0.53	17	1	0	-101.47	0.36	1.02	-0.55	11
AuroraA119403_nomyc_syed_pT288287_fixed_cap_hydrogens3	0.36	15	3	0	-95.85	0.34	-0.46	0.91	12
AuroraA119403_6c2t_pT288287_final_topclusterc1	0.43	16	2	0	-98.32	0.30	0.12	0.27	13
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc0	0.48	16	2	0	-100.5	0.21	0.57	-0.30	14
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc1	0.52	16	2	0	-102.41	0.11	0.93	-0.79	15
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc6	0.32	15	3	0	-95.71	0.08	-0.84	0.95	16
AuroraA119403_1ol7_Matt_pT288287_cap_hydrogens	0.57	17	1	0	-104.29	0.06	1.36	-1.28	17
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc5	0.42	16	2	0	-99.09	0.06	0.01	0.07	18
AuroraA119403_6c2t_pT288287_final_topclusterc8	0.25	15	3	0	-93.45	0.05	-1.46	1.53	19
AuroraA119403_6c2t_pT288287_final_topclusterc3	0.31	15	3	0	-96.01	-0.03	-0.91	0.87	20
AuroraA119403_6c2t_pT288287_final_topclusterc6	0.22	15	3	0	-93.14	-0.08	-1.72	1.61	21
AuroraA119403_3UOK__pT288287_cap_hydrogens	0.50	16	2	0	-102.74	-0.10	0.74	-0.88	22
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc3	0.46	16	2	0	-101.88	-0.15	0.45	-0.65	23
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc2	0.44	16	2	0	-101.18	-0.20	0.21	-0.47	24
AuroraA119403_5ew9_Matt_cap_hydrogens	0.34	15	3	0	-97.89	-0.23	-0.69	0.38	25
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc4	0.43	16	2	0	-101.57	-0.32	0.14	-0.57	26
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc9	0.31	15	3	0	-98.12	-0.44	-0.91	0.32	27
AuroraA119403_2wtv_b_Matt_pT288287_cap_hydrogens	0.45	16	2	0	-103.01	-0.47	0.33	-0.95	28
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc6	0.43	16	2	0	-102.52	-0.51	0.14	-0.82	29
AuroraA119403_6c2t_pT288287_final_topclusterc9	0.16	14	4	0	-93.76	-0.63	-2.29	1.45	30
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc2	0.33	15	3	0	-99.9	-0.67	-0.74	-0.14	31
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc3	0.34	15	3	0	-102.39	-1.10	-0.67	-0.79	32
AuroraA119403_6hjk_pT288287_cap_hydrogens	0.43	17	1	0	-105.8	-1.13	0.17	-1.67	33

AuroraA119403_1ol7_Matt_pT288287_final_topclusterc4	0.33	15	3	0	-102.37	-1.13	-0.72	-0.78	34
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc0	0.45	16	2	0	-107.55	-1.35	0.33	-2.12	35
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc8	0.41	16	2	0	-106.22	-1.38	-0.05	-1.78	36
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc8	0.30	15	3	0	-102.7	-1.42	-1.01	-0.87	37
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc1	0.32	15	3	0	-103.44	-1.43	-0.83	-1.06	38
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc7	0.20	14	4	0	-102.04	-1.95	-1.88	-0.70	39

5.3.1.2 Using crosslinks and monolinks with more than 2 CSMs and normalised

Model	MNXL	Matched	NoV	NoNA	MoDS	XLMO	zMNXL	zMoDS	Place
AuroraA119403_6c2t_pT288287_final_topclusterc0	0.32	9	1	0	-54.27	1.62	0.93	1.32	1
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc7	0.31	9	1	0	-54.69	1.48	0.84	1.20	2
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc9	0.35	9	1	0	-56.19	1.46	1.22	0.80	3
AuroraA119403_6c2t_pT288287_final_topclusterc1	0.29	9	1	0	-54.27	1.33	0.53	1.32	4
AuroraA119403_6c2t_pT288287_final_topclusterc2	0.29	9	1	0	-54.93	1.26	0.61	1.14	5
AuroraA119403_6c2t_pT288287_cap_hydrogens	0.31	9	1	0	-56.08	1.16	0.78	0.83	6
AuroraA119403_6c2t_pT288287_final_topclusterc4	0.32	9	1	0	-57.61	0.97	0.93	0.41	7
AuroraA119403_4c3p_Matt_pT288287_cap_hydrogens	0.33	9	1	0	-58.22	0.95	1.07	0.25	8
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc1	0.36	9	1	0	-59.4	0.91	1.34	-0.07	9
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc5	0.29	9	1	0	-57.16	0.77	0.53	0.53	10
AuroraA119403_6c2t_pT288287_final_topclusterc5	0.26	9	1	0	-56.4	0.72	0.26	0.74	11
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc9	0.33	9	1	0	-59.62	0.67	1.06	-0.13	12
AuroraA119403_6c2t_pT288287_final_topclusterc3	0.18	8	2	0	-54.41	0.43	-0.68	1.28	13
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc4	0.30	9	1	0	-60.05	0.30	0.66	-0.25	14
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc6	0.17	8	2	0	-54.7	0.27	-0.82	1.20	15
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc2	0.28	9	1	0	-59.49	0.27	0.47	-0.10	16
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc0	0.31	9	1	0	-60.97	0.25	0.84	-0.50	17
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc5	0.31	9	1	0	-60.95	0.24	0.82	-0.49	18
AuroraA119403_nomyc_syed_pT288287_fixed_cap_hydrogens3	0.21	8	2	0	-56.94	0.17	-0.36	0.59	19
AuroraA119403_6c2t_pT288287_final_topclusterc7	0.13	8	2	0	-53.66	0.15	-1.27	1.48	20
AuroraA119403_2wtv_b_Matt_pT288287_cap_hydrogens	0.32	9	1	0	-62.18	0.07	0.92	-0.82	21
AuroraA119403_5l8k_pT288287_cap_hydrogens	0.26	9	1	0	-59.67	0.07	0.24	-0.14	22
AuroraA119403_6c2t_pT288287_final_topclusterc6	0.13	8	2	0	-54.23	0.04	-1.27	1.33	23
AuroraA119403_6c2t_pT288287_final_topclusterc9	0.13	8	2	0	-54.39	-0.02	-1.31	1.28	24
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc3	0.32	9	1	0	-63.29	-0.11	0.97	-1.12	25
AuroraA119403_3UOK__pT288287_cap_hydrogens	0.32	9	1	0	-63.17	-0.15	0.89	-1.09	26
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc0	0.26	9	1	0	-61.87	-0.39	0.20	-0.74	27

AuroraA119403_5ew9_Matt_cap_hydrogens	0.19	8	2	0	-59.05	-0.44	-0.63	0.02	28
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc6	0.27	9	1	0	-63.98	-0.67	0.38	-1.31	29
AuroraA119403_6c2t_pT288287_final_topclusterc8	0.03	7	3	0	-54.37	-0.84	-2.46	1.29	30
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc2	0.19	8	2	0	-61.37	-0.87	-0.60	-0.60	31
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc8	0.18	8	2	0	-61.24	-0.94	-0.74	-0.57	32
AuroraA119403_1ol7_Matt_pT288287_cap_hydrogens	0.29	9	1	0	-66.08	-0.97	0.54	-1.88	33
AuroraA119403_4c3p_Matt_pT288287_final_topclusterc3	0.16	8	2	0	-61.06	-1.12	-1.03	-0.52	34
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc4	0.18	8	2	0	-64.03	-1.49	-0.74	-1.32	35
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc1	0.16	8	2	0	-63.52	-1.56	-0.97	-1.19	36
AuroraA119403_6hjk_pT288287_cap_hydrogens	0.18	9	1	0	-65.34	-1.70	-0.68	-1.68	37
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc8	0.17	8	2	0	-64.91	-1.76	-0.88	-1.56	38
AuroraA119403_1ol7_Matt_pT288287_final_topclusterc7	0.02	7	3	0	-62.52	-2.55	-2.63	-0.92	39

6 List of References

1. Courtheoux, T., Diallo, A., Damodaran, A. P., Reboutier, D., Watrin, E., and Prigent, C. 2018. Aurora A kinase activity is required to maintain an active spindle assembly checkpoint during prometaphase. *Journal of Cell Science* **131**(7), article number: 191353 [no pagination].
2. Crane, R., Gadea, B., Littlepage, L., Wu, H., and Ruderman, J. V. 2004. Aurora A, meiosis and mitosis. *Biology of the Cell* **96**(3), pp. 215–29.
3. Hannah, E., Kirkham, M., Hyman, A. A., and Oegema, K. 2001. Aurora-A kinase is required for centrosome maturation in Caenorhabditis elegans. *The Journal Cell Biology* **155**(7), pp. 1109–1116.
4. Darumoto, T., Honda, S., Hara, T., Nitta, M., Hirota, T., Kohmura, E., and Saya, H. 2003. Aurora-A kinase maintains the fidelity of early and late mitotic events in HeLa cells. *Journal of Biological Chemistry* **278**(51), pp. 51786–95.
5. Otto, T., Horn, S., Brockman, M., Eilers, U., Schüttrumpf, L., Popov, N., Kenney, A., Schulte, J. H., Beijersbergen, R., Christiansen, H., Berwanger, B., and Eilers, M. 2009. Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer cell* **15**(1), pp. 67–78.
6. Yang, S., He, S., Zhou, X., Liu, M., Zhu, H., Wang, Y., Zhang, W., Yan, S., Quan, L., Bai, J., and Xu, N. 2010. Suppression of Aurora-A oncogenic potential by c-Myc downregulation. *Experimental & molecular medicine* **42**(11), pp. 759–767.
7. Büchel, G., Carstensen, A., Mak, K., Rosechert, I., Leen, E., Sumara, O., Hofstetter, J., Herold, S., Kalb, J., Baluapuri, A., Poon, E., Kwok, C., Chesler, L., Maric, H. M., Rickman, D. S., Wolf, E., Bayliss, R., Walz, S., and Eilers, M., et al. 2017. Association with Aurora-A controls N-MYC-dependent promoter escape and pause release of RNA Polymerase II during the cell cycle. *Cell reports* **21**(12), pp. 3483–3497.
8. Shang, X., Burlingame, S. M., Okcu, M. F., Ge, N., Russell, H. V., Egler, R. A., David, R. D., Vasudevan, S. A., Yang, J., and Nuchtern, J. G. 2009. Aurora A is a negative prognostic factor and a new therapeutic target in human neuroblastoma. *Molecular Cancer Therapeutics* **8**(8), pp. 2461–2469.
9. Hu, W., Kavanagh, J. J., Deaver, M., Johnston, D. A., Freedman, R. S., Verschraegen, C. F., and Sen, S., 2005. Frequent overexpression of STK15/Aurora-A/BTAK and chromosomal instability in tumorigenic cell cultures derived from human ovarian cancer. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* **15**(1), pp. 49–57.
10. Burum-Auensen, E., De Angelis, P. M., Schjølberg, A. R., Kravik, K. L., Aure, M., and Clausen, O. P. F. 2007. Subcellular localization of the spindle proteins Aurora A, Mad2, and BUBR1 assessed by immunohistochemistry. *Journal of Histochemistry & Cytochemistry* **55**(5), pp. 477–486.
11. Gustafson, W. C., Meyerowitz, J. G., Nekritz, E. A., Chen, J., Benes, C., Charron, E., Simonds, E. F., Seeger, R., Matthay, K. K., Hertz, N. T., Eilers, M., Shokat, K. M., and Weiss, W. A. 2014. Drugging MYCN through an allosteric transition in Aurora kinase A. *Cancer Cell* **26**(3), pp. 414–427.
12. Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J., and Toretsky, J. A. 2018. Targeting intrinsically disordered transcription factors: changing the paradigm. *Journal of Molecular Biology* **430**(16), pp. 2321–2341.
13. Beaulieu, M., Castillo, F., and Soucek, L. 2020. Structural and biophysical insights into the function of the intrinsically disordered Myc oncoprotein. *Cells* **9**(4), article number: 1038 [no pagination].
14. Dingar, D., Kalkat, M., Chan, P., Srikumar, T., Bailey, S., Tu, W. B., Coyaud, Ponzielli, R., Kolyar, M., Jurisica, I., Huang, A., Lupien, M., Penn, L. Z., and Raught, B. 2015. BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors. *Journal of Proteomics* **118**, pp. 95–111.
15. Chen, H., Liu, H. and Qing, G., 2018. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal transduction and targeted therapy*, **3**(1), p.5.
16. Torres-Ayuso, P. and Brognard, J., 2019. Combing the cancer genome for novel kinase drivers and new therapeutic targets. *Cancers*, **11**(12), p.1972.
17. Reguart, N. and Remon, J., 2015. Common EGFR-mutated subgroups (Del19/L858R) in advanced non-small-cell lung cancer: chasing better outcomes with tyrosine kinase inhibitors. *Future Oncology*, **11**(8), pp.1245-1257.
18. Rana, A., Hussain Shah, S., Rehman, N. and Ali Sh, M.A.G., 2011. Chronic myeloid leukemia: attributes of break point cluster region-abelson (BCR-ABL). *J Cancer Res Exp Oncol*, **3**(6), pp.62-66
19. Zhu, G., Yin, J. and Cuny, G.D., 2021. Current status and challenges in drug discovery against the globally important zoonotic cryptosporidiosis. *Animal Diseases*, **1**(1), pp.1-10.
20. Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A. and Brunak, S., 2010. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signalling*, **3**(104), pp.ra3-ra3
21. Prêtre, V., and Wicki, A. 2018. Inhibition of Akt and other AGC kinases: A target for clinical cancer therapy? *Seminars in Cancer Biology* **48**, pp. 70–77.
22. Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Lo Muzio, L. 2017. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine* **40**(2), pp. 271–280.
23. Pearce, L. R., Komander, D., and Alessi, D. R. 2010. The nuts and bolts of AGC protein kinases. *Nature Reviews Molecular Cell biology* **11**(1), pp. 9–22.
24. Nishi, H., Shaytan, A. and Panchenko, A.R., 2014. Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in genetics*, **5**, p.270

25. Nardozzi, J.D., Lott, K. and Cingolani, G., 2010. Phosphorylation meets nuclear import: a review. Cell Communication and Signaling, 8, pp.1-17.
26. Miller, C.J. and Turk, B.E., 2018. Homing in: mechanisms of substrate targeting by protein kinases. *Trends in biochemical sciences*, **43**(5), pp.380-394
27. Bórquez, D.A. and González-Billault, C., 2016. Bioinformatics Approaches for Predicting Kinase–Substrate Relationships. *Bioinformatics-Updated Features and Applications*, pp.1-21.
28. López-Palacios, T.P. and Andersen, J.L., 2022. Kinase regulation by liquid–liquid phase separation. *Trends in Cell Biology.*
29. Ubersax, J.A. and Ferrell Jr, J.E., 2007. Mechanisms of specificity in protein phosphorylation. Nature reviews Molecular cell biology, 8(7), pp.530-541.
30. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. 2002. The protein kinase complement of the human genome. *Science* **298**(5600), pp. 1912–1934.
31. Wilson, L. J., Linley, A., Hammond, D. E., Hood, F. E., Coulson, J. M., MacEwan, D. J., Ross, S. J., Slupsky, J. R., Smith, P. D., Eyers, P. A., and Prior, I. A. 2018. New Perspectives, Opportunities, and Challenges in Exploring the Human Protein Kinome. *Cancer Research* **78**(1), pp. 15–29.
32. Leroux, A. E., Schulze, J. O., and Biondi, R. M. 2018. AGC kinases, mechanisms of regulation and innovative drug development. Seminars in *Cancer Biology* **48**, pp. 1–17.
33. Engel, M. 2013. The PIF Pocket of AGC Kinases: A Target Site for Allosteric Modulators and Protein–Protein Interaction Inhibitors, In: Dömling, A. ed, *Protein-Protein Interactions in Drug Discovery*. Weinheim: Wiley-VCH.
34. Biondi, R. M., Cheung, P. C. F., Casamayor, A., Deak, M., Currie, R. A., and Alessi, D. R. 2000. Identification of a pocket in the PDK1 kinase domain that interacts with PIF and the C-terminal residues of PKA. *The Embo Journal* **19**(5), pp. 979–988.
35. Schulze, J. O., Saladino, G., Busschots, K., Neimanis, S., Süss, E., Odadzic D., Zeuzem, S., Hindie, V., Herbrand, A. K., Lisa, M., Alzari, P. M., Gervasio, F. L., and Biondi, R. M. 2016. Bidirectional allosteric communication between the ATP-binding site and the regulatory PIF pocket in PDK1 protein kinase. *Cell Chemical Biology*, **23**(10), pp. 1193–1205.
36. Chan, T.O., Armen, R.S., Yadav, S., Shah, S., Zhang, J., Tiegs, B.C., Keny, N., Blumhof, B., Deshpande, D.A., Rodeck, U. and Penn, R.B., 2020. A tripartite cooperative mechanism confers resistance of the protein kinase A catalytic subunit to dephosphorylation. *Journal of Biological Chemistry*, **295**(10), pp.3316-3329.
37. Taylor, S.S., Ilouz, R., Zhang, P. and Kornev, A.P., 2012. Assembly of allosteric macromolecular switches: lessons from PKA. *Nature reviews Molecular cell biology*, **13**(10), pp.646-658
38. Narayana, N., Cox, S., Xuong, N.H., Ten Eyck, L.F. and Taylor, S.S., 1997. A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility. *Structure*, **5**(7), pp.921-935.
39. Endicott, J.A., Noble, M.E. and Johnson, L.N., 2012. The structural basis for control of eukaryotic protein kinases. *Annual review of biochemistry*, **81**, pp.587-613.
40. Steichen, J.M., Kuchinskas, M., Keshwani, M.M., Yang, J., Adams, J.A. and Taylor, S.S., 2012. Structural basis for the regulation of protein kinase A by activation loop phosphorylation. *Journal of Biological Chemistry*, **287**(18), pp.14672-14680
41. Miller, C. J., and Turk, B. E. 2018. Homing in: mechanisms of substrate targeting by protein kinases. *Trends in Biochemical Sciences* **43**(5), pp. 380–394.
42. Lučić, I., Rathinaswamy, M. K., Truebestein, L., Hamelin, D. J., Burke, J. E., and Leonard, T. A. 2018. Conformational sampling of membranes by Akt controls its activation and inactivation. *PNAS* **115**(17), pp. 3940–3949.
43. Bayliss, R., Hay, T., and Yeoh, S. 2015. The Ys and wherefores of protein kinase autoinhibition. Biochimica et biophysica acta 1854, pp.1596–1594.
44. Reikhardt, B.A. and Shabanov, P.D., 2020. Catalytic subunit of PKA as a prototype of the eukaryotic protein kinase family. *Biochemistry (Moscow),* **85**, pp.409-424.
45. Kornev, A.P., Haste, N.M., Taylor, S.S. and Ten Eyck, L.F., 2006. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences*, **103**(47), pp.17783-17788.
46. Johnson, L.N., Noble, M.E. and Owen, D.J., 1996. Active and inactive protein kinases: structural basis for regulation. *Cell*, **85**(2), pp.149-158.
47. Dodson, Charlotte A., et al. 2013. The structural mechanisms that underpin mitotic kinase activation. *Biochemical Society Transactions* **41**(4) pp.1037-1041.
48. Bayliss, R., Fry, A., Haq, T. and Yeoh, S., 2012. On the molecular mechanisms of mitotic kinase activation. Open biology, 2(11), p.120136.
49. Bayliss, R., Burgess, S. G., and McIntyre, P. J. 2017. Switching Aurora-A kinase on and off at an allosteric site. *The FEBS Journal* **284**(18), pp. 2947–2954.
50. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W. and Ausio, J., 2001. Intrinsically disordered protein. *Journal of molecular graphics and modelling*, **19**(1), pp.26-59
51. Ruan, H., Sun, Q., Zhang, W., Liu, Y. and Lai, L., 2019. Targeting intrinsically disordered proteins at the edge of chaos. *Drug discovery today*, **24**(1), pp.217-227.
52. Miskei, M., Horvath, A., Vendruscolo, M. and Fuxreiter, M., 2020. Sequence-based prediction of fuzzy protein interactions. *Journal of molecular biology*, **432**(7), pp.2289-2303.
53. Boehr, D.D., Nussinov, R. and Wright, P.E., 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, **5**(11), pp.789-796

54. Wei, G., Xi, W., Nussinov, R. and Ma, B., 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. *Chemical reviews*, **116**(11), pp.6516-6551
55. Deiana, A., Forcelloni, S., Porrello, A. and Giansanti, A., 2019. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One*, **14**(8), p.e0217889
56. Gsponer, J., Futschik, M.E., Teichmann, S.A. and Babu, M.M., 2008. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*, **322**(5906), pp.1365-1368.
57. Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N. and Dunker, A.K., 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. *Journal of proteome research*, **6**(6), pp.2351-2366.
58. Coletta, A., Pinney, J.W., Solís, D.Y.W., Marsh, J., Pettifer, S.R. and Attwood, T.K., 2010. Low-complexity regions within protein sequences have position-dependent roles. *BMC systems biology*, **4,** pp.1-13.
59. Morris, O.M., Torpey, J.H. and Isaacson, R.L., 2021. Intrinsically disordered proteins: modes of binding with emphasis on disordered domains. *Open Biology*, **11**(10), p.210222.
60. Olsen, J.G., Teilum, K. and Kragelund, B.B., 2017. Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. *Cellular and Molecular Life Sciences*, **74**, pp.3175-3183.
61. Fuxreiter, M. and Tompa, P., 2012. Fuzzy complexes: a more stochastic view of protein function. *Fuzziness: Structural disorder in protein complexes*, pp.1-14.
62. Weng, J. and Wang, W., 2020. Dynamic multivalent interactions of intrinsically disordered proteins. *Current Opinion in Structural Biology*, **62**, pp.9-13
63. Gopal, S.M., Wingbermühle, S., Schnatwinkel, J., Juber, S., Herrmann, C. and Schäfer, L.V., 2020. Conformational preferences of an intrinsically disordered protein domain: A case study for modern force fields. *The Journal of Physical Chemistry B*, **125**(1), pp.24-35.
64. Henzler-Wildman, K. and Kern, D., 2007. Dynamic personalities of proteins. *Nature*, **450**(7172), pp.964-972
65. Wei, G., Xi, W., Nussinov, R. and Ma, B., 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. *Chemical reviews*, **116**(11), pp.6516-6551.
66. Celik, L., Lund, J.D.D. and Schiøtt, B., 2007. Conformational dynamics of the estrogen receptor α: molecular dynamics simulations of the influence of binding site structure on protein dynamics. *Biochemistry*, **46**(7), pp.1743-1758
67. Kar, G., Keskin, O., Gursoy, A. and Nussinov, R., 2010. Allostery and population shift in drug discovery. *Current opinion in pharmacology*, **10**(6), pp.715-722.
68. Vijayan, R. S. K., He, P., Modi, V., Duong-Ly, K. C., Ma, H., Peterson, J. R., Dunbrack Jr, R. L., and Levy, R. M. 2015. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *Journal of Medicinal Chemistry* **58**(1), pp. 466–479.
69. Hsu, C.C., Buehler, M.J. and Tarakanova, A., 2020. The order-disorder continuum: Linking predictions of protein structure and disorder through molecular simulation. *Scientific reports*, **10**(1), pp.1-14
70. Arbesú, M. and Pons, M., 2019. Integrating disorder in globular multidomain proteins: Fuzzy sensors and the role of SH3 domains. *Archives of biochemistry and biophysics*, **677**, p.108161.
71. Ung, P. M., Rahman, R., and Schlessinger, A. 2018. Redefining the protein kinase conformational space with machine learning. *Cell Chemical Biology* **25**(7), pp. 916–924.
72. Beenstock, J., Mooshayef, N. and Engelberg, D., 2016. How do protein kinases take a selfie (autophosphorylate)?. *Trends in biochemical sciences*, **41**(11), pp.938-953.
73. Lochhead, P.A., 2009. Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation. *Science signaling*, **2**(54), pp.pe4-pe4.
74. Julien, M., Bouguechtouli, C., Alik, A., Ghouil, R., Zinn-Justin, S., Theillet, F. 2020. Multiple Site-Specific Phosphorylation of IDPs Monitored by NMR. *Methods in Molecular Biology* **2141**, pp. 793–817.
75. Gógl, G., Kornev, A. P., Reményi, A., and Taylor, S. S. 2019. Disordered protein kinase regions in regulations of kinase domain cores. *Trends in Biochemical Sciences*. **44**(4), pp. 300–311.
76. Sen, S., Zhou, H. and White, R.A., 1997. A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene*, **14**(18), pp.2195-2200
77. Carmena, M., and Earnshaw, W. C. 2003. The cellular geography of Aurora kinases. *Nature Reviews Molecular Cell Biology* **4**, pp. 842–854.
78. Seeger, R. C., Brodeur, G. M., Sather, H., Dalton, A., Siegel, S. E., Wong, K. Y., and Hammond, D. 1985. Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *New England Journal of Medicine* **313**(18), pp. 1111–1116.
79. Schwab, M. 2004. MYCN in neuronal tumours. *Cancer Letters* **204**(2), pp. 179–187.
80. Naso, F.D., Boi, D., Ascanelli, C., Pamfil, G., Lindon, C., Paiardini, A. and Guarguaglini, G., 2021. Nuclear localisation of Aurora-A: its regulation and significance for Aurora-A functions in cancer. *Oncogene*, **40**(23), pp.3917-3928.
81. Vader, G., and Lens, S. 2008. The Aurora kinase family in cell division and cancer. *Biochimica et biophysica acta* **1786**(1), pp. 60-72.
82. Willems, E., Dedobbelerr, M., Digregorio, M., Lombard, A., Lumapat, P. N., Rogister, B. 2018. The functional diversity of Aurora kinases: a comprehensive review. *Cell Division* **18**, article number: 7 [no pagination].
83. Ahmed, A., Shamsi, A., Mohammad, T., Hasan, G. M., Islam, A., and Hassan, M. I. 2021. Aurora B kinase: A potential drug target for cancer therapy. *Journal of Cancer Research and Clinical Oncology* **147**(8), pp. 2187–2198.
84. Bejar, J. F., DiSanza, Z., and Quartuccio, S. M. 2021. The oncogenic role of meiosis-specific Aurora kinase C in mitotic cells. *Experimental Cell Research* **407**(2), article number: 112803 [no pagination].

85. Galetta, D., and Cortes-Dericks, L. 2020. Promising therapy in lung cancer: spotlight on Aurora kinases. *Cancers* **12**(11), article number: 3371 [no pagination].
86. McIntyre, P. J., Collins, P. M., Vrzal, L., Birchall, K., Arnold, L. H., Mpamhanga, C., Coombs, P. J., Burgess, S. G., Richards, M. W., Winter, A., Veverka, V., von Delft, F., Merritt, A., and Bayliss, R. 2017. Characterization of three druggable hot-spots in the Aurora-A/TPX2 interaction using biochemical, biophysical, and fragment-based approaches. *ASC Chemical Biology* **12**(11), 2906–2914.
87. Lochhead, P. A. 2009. Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation. *Science signalling* **2**(54), article number: 254pe4 [no pagination].
88. Tavernier, N., Thomas, Y., Vigneron, S., Maisonneuve, P., Orlicky, S., Mader, P., Regmi, S. G., Van Hove, L., Levinson, N. M., Gasami-Seabrook, Joly, N., Poteau, M., Velez-Aguilera, G., Gavet, O., Castro, A., Dasso, M., Lorca, T., Sicheri, F., and Pintard, L. 2021. Bora phosphorylation substitutes in trans for T-loop phosphorylation in Aurora A to promote mitotic entry. *Nature Communications* **12**, article number: 1899 [no pagination].
89. Pitsawong, W., Buosi, V., Otten, R., Agafonov, R.V., Zorba, A., Kern, N., Kutter, S., Kern, G., Pádua, R.A., Meniche, X. and Kern, D., 2018. Dynamics of human protein kinase Aurora A linked to drug selectivity. *Elife*, **7**, p.e36656.
90. Tavernier, N., Sicheri, F., Pintard, L. 2021. Aurora A kinase activation: different means to different ends. *Journal of Cell Biology* **220**(9), article number: 202106128 [no pagination].
91. Stenoien, D.L., Sen, S., Mancini, M.A. and Brinkley, B.R., 2003. Dynamic association of a tumor amplified kinase, Aurora-A, with the centrosome and mitotic spindle. *Cell motility and the cytoskeleton*, **55**(2), pp.134-146.
92. Eyers, P.A., Erikson, E., Chen, L.G. and Maller, J.L., 2003. A novel mechanism for activation of the protein kinase Aurora A. *Current Biology*, **13**(8), pp.691-697.
93. Liu, Q. and Ruderman, J.V., 2006. Aurora A, mitotic entry, and spindle bipolarity. *Proceedings of the National Academy of Sciences*, **103**(15), pp.5811-5816.
94. Li, S., Deng, Z., Fu, J., Xu, C., Xin, G., Wu, Z., Luo, J., Wang, G., Zhang, S., Zhang, B. and Zou, F., 2015. Spatial compartmentalization specializes the function of Aurora A and Aurora B. *Journal of Biological Chemistry*, **290**(28), pp.17546-17558
95. Bai, M., Ni, J., Shen, S., Wu, J., Huang, Q., Le, Y. and Yu, L., 2014. Two newly identified sites in the N-terminal regulatory domain of Aurora-A are essential for auto-inhibition. *Biotechnology letters*, **36**, pp.1595-1604.
96. Taguchi, S.I., Honda, K., Sugiura, K., Yamaguchi, A., Furukawa, K. and Urano, T., 2002. Degradation of human Aurora-A protein kinase is mediated by hCdh1. *FEBS letters*, **519**(1-3), pp.59-65.
97. Littlepage, L.E. and Ruderman, J.V., 2002. Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. *Genes & development*, **16**(17), pp.2274-2285.
98. Arlot-Bonnemains, Y., Klotzbucher, A., Giet, R., Uzbekov, R., Bihan, R. and Prigent, C., 2001. Identification of a functional destruction box in the Xenopus laevis aurora-A kinase pEg2. *FEBS letters*, **508**(1), pp.149-152.
99. Horn, V., Thélu, J., Garcia, A., Albiges-Rizo, C., Block, M.R. and Viallet, J., 2007. Functional interaction of Aurora-A and PP2A during mitosis. *Molecular biology of the cell*, **18**(4), pp.1233-1241.
100. Abdelbaki, A., Ascanelli, C., Okoye, C.N., Akman, H.B., Janson, G., Min, M., Marcozzi, C., Hagting, A., Grant, R., De Luca, M. and Asteriti, I.A., 2023. Revisiting degron motifs in human AURKA required for its targeting by APC/Cfzr1. *Life Science Alliance*, **6**(2).
101. Giubettini, M., Asteriti, I.A., Scrofani, J., De Luca, M., Lindon, C., Lavia, P. and Guarguaglini, G., 2011. Control of Aurora-A stability through interaction with TPX2. *Journal of cell science*, **124**(1), pp.113-122.
102. Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Honda, S., Kobayashi, O., Hatakeyama, K., Ushio, Y., Saya, H. and Hirota, T., 2003. CENP-A phosphorylation by Aurora-A in prophase is required for enrichment of Aurora-B at inner centromeres and for kinetochore function. *Developmental cell*, **5**(6), pp.853-864.
103. Kettenbach, A.N., Schweppe, D.K., Faherty, B.K., Pechenick, D., Pletnev, A.A. and Gerber, S.A., 2011. Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. *Science signalling*, **4**(179), pp.rs5-rs5
104. Szimler, T., Gráczer, É., Györffy, D., Végh, B., Szilágyi, A., Hajdú, I., Závodszky, P. and Vas, M., 2019. New type of interaction between the SARAH domain of the tumour suppressor RASSF1A and its mitotic kinase Aurora A. *Scientific Reports*, **9**(1), p.5550.
105. Nikonova, A.S., Astsaturov, I., Serebriiskii, I.G., Dunbrack, R.L. and Golemis, E.A., 2013. Aurora A kinase (AURKA) in normal and pathological cell division. *Cellular and Molecular Life Sciences*, **70**, pp.661-687.
106. Scrittori, L., Skoufias, D.A., Hans, F., Gerson, V., Sassone-Corsi, P., Dimitrov, S. and Margolis, R.L., 2005. A small C-terminal sequence of Aurora B is responsible for localization and function. *Molecular biology of the cell*, **16**(1), pp.292-305.
107. Quartuccio, S.M. and Schindler, K., 2015. Functions of Aurora kinase C in meiosis and cancer. *Frontiers in cell and developmental biology*, **3**, p.50.
108. Li, S., Deng, Z., Fu, J., Xu, C., Xin, G., Wu, Z., Luo, J., Wang, G., Zhang, S., Zhang, B. and Zou, F., 2015. Spatial compartmentalization specializes the function of Aurora A and Aurora B. *Journal of Biological Chemistry*, **290**(28), pp.17546-17558.
109. Kimmins, S., Crosio, C., Kotaja, N., Hirayama, J., Monaco, L., Hoog, C., van Duin, M., Gossen, J.A. and Sassone-Corsi, P., 2007. Differential functions of the Aurora-B and Aurora-C kinases in mammalian spermatogenesis. *Molecular endocrinology*, **21**(3), pp.726-739.
110. Willems, E., Dedobbeleer, M., Digregorio, M., Lombard, A., Lumapat, P.N. and Rogister, B., 2018. The functional diversity of Aurora kinases: a comprehensive review. *Cell division*, **13**(1), pp.1-17.
111. Joukov, V. and De Nicolo, A., 2018. Aurora-PLK1 cascades as key signaling modules in the regulation of mitosis. *Science signalling*, **11**(543), p.eaar4195.

112. Reboutier, D., Benaud, C. and Prigent, C., 2015. Aurora A's functions during mitotic exit: the guess Who game. *Frontiers in Oncology*, **5**, p.290.
113. Büchel, G., Carstensen, A., Mak, K.Y., Roeschert, I., Leen, E., Sumara, O., Hofstetter, J., Herold, S., Kalb, J., Baluapuri, A. and Poon, E., 2017. Association with Aurora-A controls N-MYC-dependent promoter escape and pause release of RNA polymerase II during the cell cycle. *Cell reports*, **21**(12), pp.3483-3497.
114. Magnaghi-Jaulin, L., Eot-Houllier, G., Gallaud, E. and Giet, R., 2019. Aurora A protein kinase: to the centrosome and beyond. *Biomolecules*, **9**(1), p.28.
115. Adhikari, B., Bozilovic, J., Diebold, M., Schwarz, J.D., Hofstetter, J., Schröder, M., Wanior, M., Narain, A., Vogt, M., Dudvarski Stankovic, N. and Baluapuri, A., 2020. PROTAC-mediated degradation reveals a non-catalytic function of AURORA-A kinase. *Nature chemical biology*, **16**(11), pp.1179-1188.
116. Otto, T., Horn, S., Brockmann, M., Eilers, U., Schüttrumpf, L., Popov, N., Kenney, A.M., Schulte, J.H., Beijersbergen, R., Christiansen, H. and Berwanger, B., 2009. Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer cell*, **15**(1), pp.67-78.
117. Gheghiani, L., Loew, D., Lombard, B., Mansfeld, J. and Gavet, O., 2017. PLK1 activation in late G2 sets up commitment to mitosis. *Cell reports*, **19**(10), pp.2060-2073.
118. Hirota, T., Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Nitta, M., Hatakeyama, K. and Saya, H., 2003. Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells. *Cell*, **114**(5), pp.585-598.
119. Giesecke, A. and Stewart, M., 2010. Novel binding of the mitotic regulator TPX2 (target protein for Xenopus kinesin-like protein 2) to importin-α. *Journal of Biological Chemistry*, **285**(23), pp.17628-17635.
120. Cavazza, T. and Vernos, I., 2016. The RanGTP pathway: from nucleo-cytoplasmic transport to spindle assembly and beyond. *Frontiers in cell and developmental biology*, **3**, p.82.
121. Kufer, T.A., Silljé, H.H., Körner, R., Gruss, O.J., Meraldi, P. and Nigg, E.A., 2002. Human TPX2 is required for targeting Aurora-A kinase to the spindle. *The Journal of cell biology*, **158**(4), pp.617-623.
122. Polverino, F., Naso, F.D., Asteriti, I.A., Palmerini, V., Singh, D., Valente, D., Bird, A.W., Rosa, A., Mapelli, M. and Guarguaglini, G., 2021. The aurora-A/TPX2 Axis directs spindle orientation in adherent human cells by regulating NuMA and microtubule stability. *Current Biology*, **31**(3), pp.658-667.
123. Burgess, S.G., Mukherjee, M., Sabir, S., Joseph, N., Gutiérrez-Caballero, C., Richards, M.W., Huguenin-Dezot, N., Chin, J.W., Kennedy, E.J., Pfuhl, M. and Royle, S.J., 2018. Mitotic spindle association of TACC3 requires Aurora-A-dependent stabilization of a cryptic α-helix. *The EMBO journal*, **37**(8), p.e97902.
124. Lindon, C., Grant, R. and Min, M., 2016. Ubiquitin-mediated degradation of aurora kinases. *Frontiers in Oncology*, **5**, p.307.
125. Fukuda, T., Mishina, Y., Walker, M.P. and DiAugustine, R.P., 2005. Conditional transgenic system for mouse aurora a kinase: degradation by the ubiquitin proteasome pathway controls the level of the transgenic protein. *Molecular and cellular biology*, **25**(12), pp.5270-5281.
126. Guarino Almeida, E., Renaudin, X. and Venkitaraman, A.R., 2020. A kinase-independent function for AURORA-A in replisome assembly during DNA replication initiation. *Nucleic Acids Research*, **48**(14), pp.7844-7855.
127. Pugacheva, E.N., Jablonski, S.A., Hartman, T.R., Henske, E.P. and Golemis, E.A., 2007. HEF1-dependent Aurora A activation induces disassembly of the primary cilium. *Cell*, **129**(7), pp.1351-1363.
128. Lens, S.M., Voest, E.E. and Medema, R.H., 2010. Shared and separate functions of polo-like kinases and aurora kinases in cancer. *Nature Reviews Cancer*, **10**(12), pp.825-841.
129. Anand, S., Penrhyn-Lowe, S. and Venkitaraman, A.R., 2003. AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer cell*, **3**(1), pp.51-62.
130. D'assoro, A.B., Quatraro, C., Amato, A., Opyrchal, M., Leontovich, A., Ikeda, Y., Ohmine, S., Lingle, W., Suman, V., Ecsedy, J. and Iankov, I., 2014. The mitotic kinase Aurora-A promotes distant metastases by inducing epithelial-to-mesenchymal transition in ERα+ breast cancer cells. *Oncogene*, **33**(5), pp.599-610.
131. Littlepage, L.E., Wu, H., Andresson, T., Deanehan, J.K., Amundadottir, L.T. and Ruderman, J.V., 2002. Identification of phosphorylated residues that affect the activity of the mitotic kinase Aurora-A. *Proceedings of the National Academy of Sciences*, **99**(24), pp.15440-15445.
132. Fadaka, A.O., Sibuyi, N.R.S., Madiehe, A.M. and Meyer, M., 2020. MicroRNA-based regulation of Aurora A kinase in breast cancer. *Oncotarget*, **11**(46), p.4306.
133. Sasai, K., Treekitkarnmongkol, W., Kai, K., Katayama, H. and Sen, S., 2016. Functional significance of aurora kinases–p53 protein family interactions in cancer. *Frontiers in Oncology*, **6**, p.247.
134. Du, R., Huang, C., Liu, K., Li, X. and Dong, Z., 2021. Targeting AURKA in Cancer: molecular mechanisms and opportunities for Cancer therapy. *Molecular cancer*, **20**, pp.1-27.
135. Yan, V.C., Butterfield, H.E., Poral, A.H., Yan, M.J., Yang, K.L., Pham, C.D. and Muller, F.L., 2020. Why great mitotic inhibitors make poor cancer drugs. *Trends in Cancer*, **6**(11), pp.924-941.
136. Mou, P.K., Yang, E.J., Shi, C., Ren, G., Tao, S. and Shim, J.S., 2021. Aurora kinase A, a synthetic lethal target for precision cancer medicine. *Experimental & Molecular Medicine*, **53**(5), pp.835-847.
137. van Gijn, S.E., Wierenga, E., van den Tempel, N., Kok, Y.P., Heijink, A.M., Spierings, D.C., Foijer, F., van Vugt, M.A. and Fehrmann, R.S., 2019. TPX2/Aurora kinase A signaling as a potential therapeutic target in genomically unstable cancer cells. *Oncogene*, **38**(6), pp.852-867.
138. Bibby, R.A., Tang, C., Faisal, A., Drosopoulos, K., Lubbe, S., Houlston, R., Bayliss, R. and Linardopoulos, S., 2009. A cancer-associated aurora A mutant is mislocalized and misregulated due to loss of interaction with TPX2. *Journal of Biological Chemistry*, **284**(48), pp.33177-33184.
139. Levinson, N.M., 2018. The multifaceted allosteric regulation of Aurora kinase A. *Biochemical Journal*, **475**(12), pp.2025-2042.

140. Leroux, A.E., Schulze, J.O. and Biondi, R.M., 2018. AGC kinases, mechanisms of regulation. *Semin Cancer Biol* **48**, pp. 1–17.
141. Zorba, A., Nguyen, V., Koide, A., Hoemberger, M., Zheng, Y., Kutter, S., Kim, C., Koide, S. and Kern, D., 2019. Allosteric modulation of a human protein kinase with monobodies. *Proceedings of the National Academy of Sciences*, **116**(28), pp.13937-13942.
142. Burgess, S.G., Oleksy, A., Cavazza, T., Richards, M.W., Vernos, I., Matthews, D. and Bayliss, R., 2016. Allosteric inhibition of Aurora-A kinase by a synthetic vNAR domain. *Open Biology*, **6**(7), p.160089.
143. Brunet, S., Sardon, T., Zimmerman, T., Wittmann, T., Pepperkok, R., Karsenti, E. and Vernos, I., 2004. Characterization of the TPX2 domains involved in microtubule nucleation and spindle assembly in Xenopus egg extracts. *Molecular biology of the cell*, **15**(12), pp.5318-5328.
144. Bayliss, R., Sardon, T., Vernos, I. and Conti, E., 2003. Structural basis of Aurora-A activation by TPX2 at the mitotic spindle. *Molecular cell*, **12**(4), pp.851-862.
145. Zorba, A., Buosi, V., Kutter, S., Kern, N., Pontiggia, F., Cho, Y., and Kern, D. 2014. Molecular Mechanism of Aurora a Kinase Autophosphorylation and its Allosteric Activation by TPX2. *eLife* **3**, article number: 02667 [no pagination].
146. Asteriti, I.A., Polverino, F., Stagni, V., Sterbini, V., Ascanelli, C., Naso, F.D., Mastrangelo, A., Rosa, A., Paiardini, A., Lindon, C. and Guarguaglini, G., 2023. AurkA nuclear localization is promoted by TPX2 and counteracted by protein degradation. *Life Science Alliance*, **6**(5).
147. Leroux, A. E., and Biondi, R. M. 2020. Renaissance of allostery to disrupt protein kinase interactions. *Trends in Biochemical Sciences* **45**(1), pp. 27–41.
148. Lake, Eric W., et al. 2018. Quantitative conformational profiling of kinase inhibitors reveals origins of selectivity for Aurora kinase activation states. *Proceedings of the National Academy of Sciences* **115**(51), 11894-11903.
149. Fu, J., Bian, M., Liu, J., Jiang, Q. and Zhang, C., 2009. A single amino acid change converts Aurora-A into Aurora-B-like kinase in terms of partner specificity and cellular function. *Proceedings of the National Academy of Sciences*, **106**(17), pp.6939-6944.
150. Burgess, S.G., Peset, I., Joseph, N., Cavazza, T., Vernos, I., Pfuhl, M., Gergely, F. and Bayliss, R., 2015. Aurora-A-dependent control of TACC3 influences the rate of mitotic spindle assembly. *PLoS genetics*, **11**(7), p.e1005345.
151. Peset, I. and Vernos, I., 2008. The TACC proteins: TACC-ling microtubule dynamics and centrosome function. *Trends in cell biology*, **18**(8), pp.379-388.
152. Hood, F.E., Williams, S.J., Burgess, S.G., Richards, M.W., Roth, D., Straube, A., Pfuhl, M., Bayliss, R. and Royle, S.J., 2013. Coordination of adjacent domains mediates TACC3–ch-TOG–clathrin assembly and mitotic spindle binding. *Journal of Cell Biology*, **202**(3), pp.463-478.
153. Cascales, H.S., Burdova, K., Middleton, A., Kuzin, V., Müllers, E., Stoy, H., Baranello, L., Macurek, L. and Lindqvist, A., 2021. Cyclin A2 localises in the cytoplasm at the S/G2 transition to activate PLK1. *Life Science Alliance*, **4**(3).
154. Lössl, P., Brunner, A.M., Liu, F., Leney, A.C., Yamashita, M., Scheltema, R.A. and Heck, A.J., 2016. Deciphering the interplay among multisite phosphorylation, interaction dynamics, and conformational transitions in a tripartite protein system. *ACS central science*, **2**(7), pp.445-455.
155. Raab, M., Matthess, Y., Raab, C.A., Gutfreund, N., Dötsch, V., Becker, S., Sanhaji, M. and Strebhardt, K., 2022. A dimerization-dependent mechanism regulates enzymatic activation and nuclear entry of PLK1. *Oncogene*, **41**(3), pp.372-386.
156. Cyphers, S., Ruff, E.F., Behr, J.M., Chodera, J.D. and Levinson, N.M., 2017. A water-mediated allosteric network governs activation of Aurora kinase A. *Nature chemical biology*, **13**(4), pp.402-408.
157. Taylor, S.S. and Kornev, A.P., 2011. Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences*, **36**(2), pp.65-77.
158. Modi, V., and Dunbrack, R. L. 2019. A structurally-validated multiple sequence alignment of 497 human protein kinase domains. *Scientific Reports* **9**(1), pp. 1-16.
159. Tsai, C.C., Yue, Z. and Shen, J., 2019. How electrostatic coupling enables conformational plasticity in a tyrosine kinase. *Journal of the American Chemical Society*, **141**(38), pp.15092-15101.
160. Meyerowitz, J. G. Weiss, W. A., and Gustafson, W. C. 2015. A new “angle” on kinase inhibitor design: Prioritizing amphosteric activity above kinase inhibition. *Molecular & Cellular Oncology*. **2,** pp. 10–12.
161. Dodson, C.A., Yeoh, S., Haq, T. and Bayliss, R., 2013. A kinetic test characterizes kinase intramolecular and intermolecular autophosphorylation mechanisms. *Science signaling*, **6**(282), pp.ra54-ra54.
162. Burgess, S.G., Grazia Concilio, M., Bayliss, R. and Fielding, A.J., 2016. Detection of Ligand-induced Conformational Changes in the Activation Loop of Aurora-A Kinase by PELDOR Spectroscopy. *ChemistryOpen*, **5**(6), pp.531-534.
163. Dodson, C.A. and Bayliss, R., 2012. Activation of Aurora-A kinase by protein partner binding and phosphorylation are independent and synergistic. *Journal of Biological Chemistry*, **287**(2), pp.1150-1157.
164. Ruff, E.F., Muretta, J.M., Thompson, A.R., Lake, E.W., Cyphers, S., Albanese, S.K., Hanson, S.M., Behr, J.M., Thomas, D.D., Chodera, J.D. and Levinson, N.M., 2018. A dynamic mechanism for allosteric activation of Aurora kinase A by activation loop phosphorylation. *Elife*, **7**, p.e32766.
165. Gilburt, J.A., Sarkar, H., Sheldrake, P., Blagg, J., Ying, L. and Dodson, C.A., 2017. Dynamic Equilibrium of the Aurora A Kinase Activation Loop Revealed by Single-Molecule Spectroscopy. *Angewandte Chemie*, **129**(38), pp.11567-11572.
166. Lim, D.C., Joukov, V., Rettenmaier, T.J., Kumagai, A., Dunphy, W.G., Wells, J.A. and Yaffe, M.B., 2020. Redox priming promotes Aurora A activation during mitosis. *Science signaling*, **13**(641), p.eabb6707.
167. Tomlinson, L., 2022. Mass Spectrometry-Guided Structural Analysis Of Protein Kinase Inhibitor Complexes (Doctoral dissertation, University of Liverpool).
168. Akamine, P., Wu, J., Xuong, N.H., Ten Eyck, L.F. and Taylor, S.S., 2003. Dynamic features of cAMP-dependent protein kinase revealed by apoenzyme crystal structure. *Journal of molecular biology*, **327**(1), pp.159-171.

169. Musavizadeh, Z., Grottesi, A., Guarguaglini, G. and Paiardini, A., 2021. Phosphorylation, Mg-ADP, and Inhibitors Differentially Shape the Conformational Dynamics of the A-Loop of Aurora-A. *Biomolecules*, **11**(4), p.567.
170. Zorba, A., 2013. Mechanism of Aurora A Kinase Activation and Catalysis (Doctoral dissertation, Brandeis University).
171. Richards, M.W., Burgess, S.G., Poon, E., Carstensen, A., Eilers, M., Chesler, L. and Bayliss, R., 2016. Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proceedings of the National Academy of Sciences*, **113**(48), pp.13726-13731.
172. Amati, B. 2004. Myc degradation: dancing with ubiquitin ligases. Proceedings of the *National Academy of Sciences of the United States of America* **101**(24), pp. 8843-8844.
173. Chen, Y., Zhou, C., Ji, W., Mei, Z., Hu, B., Zhang, W., Zhang, D., Wang, J., Liu, X., Ouyang, G., Zhou, J., and Xiao, W. 2016. ELL targets c-Myc for proteasomal degradation and suppresses tumour growth. *Nature Communications* **7**, article number: 11057 [no pagination].
174. Morton, J. P., and Sansom, O. J. 2013. MYC-y mice: from tumour initiation to therapeutic targeting of endogenous MYC."Molecular *Oncology* **7**(2), pp. 248–258.
175. Lin, C., Liu, C., Lee, C. N., Chan, T. 2010. Targeting c-Myc as a novel approach for hepatocellular carcinoma. *World Journal of Hepatology* **2**(1), pp. 16–20.
176. Kipshidze, N. N., Iversen, P., Kim, H., Yiazdi, H., Dangas, G., Seaborn, R., New, G., Tio, F., Waksman, R., Mehran, R., Tsapenko, M., Stone, G. W., Roubin, G. S., Iyer, S., Leon, M. B., and Moses, J. W. 2004. Advanced C-Myc antisense (AVI-4126)-eluting phosphorylcholine-coated stent implantation is associated with complete vascular healing and reduced neointimal formation in the porcine coronary restenosis model. *Catheterization and Cardiovascular Interventions* **61**(4), pp. 518–527.
177. Dang, C. V., Reddy, E. P, Shokat, K. M., and Soucek, L. 2017. Drugging the 'undruggable' cancer targets. *Nature Reviews Cancer* **17**(8), 502–508.
178. Boi, D., Souvalidoi, F., Capelli, D., Polverino, F., Marini, G., Montanari, R., Pochetti, G., Tramonti, A., et al. 2021. PHA-680626 Is an Effective Inhibitor of the Interaction between Aurora-A and N-Myc. *International Journal of Molecular Sciences* **22**(23), article number: 13122 [no pagination].
179. Jiang, J., Wang, J., Yue, M., Cai, X., Wang, T., Wu, C., Su, H., Wang, Y., Han, M., Zhang, Y., Zhu, X., Jiang, P., Li, P., Sun, Y., Xiao, W., Feng, H., Qing, G., and Liu, H. 2020. Direct phosphorylation and stabilization of MYC by Aurora B kinase promote T-cell Leukemogenesis. *Cancer Cell* **37**(2), pp. 200–215.
180. Schramm, A., Bignon, C., Brocca, S., Grandori, R., Santambrogio, C. and Longhi, S., 2019. An arsenal of methods for the experimental characterization of intrinsically disordered proteins–how to choose and combine them?. *Archives of biochemistry and biophysics*, **676**, p.108055.
181. Srivastava, A., Nagai, T., Srivastava, A., Miyashita, O. and Tama, F., 2018. Role of computational methods in going beyond X-ray crystallography to explore protein structure and dynamics. *International journal of molecular sciences*, **19**(11), p.3401.
182. DePristo, M. A., de Bakker, P. I. W., and Blundell, T. L. 2004. Heterogeneity and inaccuracy in protein structures solved by x-ray crystallography. *Structure* **12**(5), pp.831–838.
183. DeForte, S., and Uversky, V. N. 2016. Resolving the ambiguity: making sense of instrinsic disorder when PDB structures disagree. *Protein Science* **25**(3), pp. 676–688.
184. Søndergaard, C. R., Garrett, A. E., Carstensen, T., Pollastri, G., and Nielsen, J. E. 2009. Structural artifacts in protein−ligand x-ray structures: implications for the development of docking scoring functions. *Journal of Medical Chemistry* **52**(18), pp.5673–5684.
185. Kermani, A.A., 2021. A guide to membrane protein X-ray crystallography. *The FEBS journal*, **288**(20), pp.5788-5804.
186. Sinz, A. 2006. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews* **25**(4), pp.663–682.
187. Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. 2010. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Molecular & Cellular Proteomics* **9**(8), pp. 1634–1649.
188. Parfentev, I., Schilbach, S., Cramer, P., Urlaub, H. 2020. An experimentally generated peptide database increases the sensitivity of XL-MS with complex samples. *Journal of Proteomics* **220**, article number: 103754 [no pagination].
189. Kao, A., Chiu, C., Vellucci, D., Yang, Y., Patel, V. R., Guan, S., Randall, A., Baldi, P., Rychnovsky, S. D., and Huang L. 2011. Development of novel cross-linking strategy for fast accurate identification of cross-linked peptides of protein complexes. *Molecular & Cellular Proteomics* **10**(1), article number: M110.002212 [no pagination].
190. Kaake, R. M., Wang, X., Burke, A., Yu, C., Kandur, W., Yang, Y., Novtisky, E. J., Second, T., Duan, J., Kao, A., Guan, S., Vullecci, D., Rychnovsky, S. D., and Huang, L. 2014. A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Molecular & Cellular Proteomics* **13**(12), pp. 3533–3543.
191. Iacobucci, C., Götze, M., Piotrowski, C., Arlt, C., Rehkamp, A., Ihling, C., Hage, C. and Sinz, A., 2018. Carboxyl-photo-reactive MS-cleavable cross-linkers: unveiling a hidden aspect of diazirine-based reagents. *Analytical chemistry*, **90**(4), pp.2805-2809.
192. Dau, T., Bartolomucci, G., and Rappsilber, J. 2020. Proteomics using protease alternatives to trypsin benefits from sequential digestion with trypsin. *Analytical chemistry* **92**(14), pp. 9523–9527.
193. Rozbesky, D., Rosulek, M., Kukačka, Z., Chmelík, J., Man, P., and Novák, P. 2018. Impact of chemical cross-linking on protein structure and function. *Analytical Chemistry* **90**(2), pp. 1104–1113.
194. Erde, Jonathan, Rachel R. Ogorzalek Loo, and Joseph A. Loo. "Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments." Journal of proteome research 13.4 (2014): 1885-1895.
195. Ludwig, Katelyn R., Monica M. Schroll, and Amanda B. Hummon. "Comparison of in-solution, FASP, and S-trap based digestion methods for bottom-up proteomic studies." Journal of proteome research 17.7 (2018): 2480-2490.

196. Barré, A., Tintas, M., Levacher, V., Papamicaël, C., and Gembus, V. An overview of the synthesis of highly versatile N-hydroxysuccinimide esters. *Synthesis* **49**(3), pp. 474–483.
197. Iacobucci, C., Götze, Ihling, C. H., Piotrowski, C., Arlt, C., Schäfer, M., Hage, C., Schmidt, R., and Sinz, A. 2018. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. *Nature Protocols* **13**(12), pp. 2864-2889.
198. Cline, G. W., and Hanna, S. B. 1988. Kinetics and mechanisms of the aminolysis of N-hydroxysuccinimide esters in aqueous buffers. *Journal of Organic Chemistry* **53**(15), pp. 3583–3586.
199. Cline, G. W., and Hanna, S. B. 1987. The aminolysis of N-hydroxysuccinimide esters. A structure-reactivity study. *Journal of the American Chemistry* **109**(10), pp. 3087–3091.
200. Muraoka, A., Matsuura, Y., Naitow, H., Ihara, M. and Kunishima, N., 2018. Availability of NHS-biotin labeling to identify free protein lysine revealed by experiment and MD simulation. *Analytical biochemistry*, **557**, pp.46-58.
201. Mädler, S., Bich, C., Touboul, T., and Zenobi, R. 2009. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *Journal of Mass Spectrometry* **44**(5), pp. 694–706.
202. Mädler, S., Gschwind, S., and Zenobi, R. 2010. Role of arginine in chemical cross-linking with N-hydroxysuccinimide esters. *Analytical Biochemistry* **398**(1), pp. 123–125.
203. Lima, D.B., Melchior, J.T., Morris, J., Barbosa, V.C., Chamot-Rooke, J., Fioramonte, M., Souza, T.A., Fischer, J.S., Gozzo, F.C., Carvalho, P.C. and Davidson, W.S., 2018. Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins. *Nature protocols*, **13**(3), pp.431-458.
204. Hogan, Donna. 2017. Technique for normalization of cross-linked peptide ion intensity to elucidate enzymatic conformational changes. (Doctoral dissertation, University of New Hampshire).
205. Bandyopadhyay, P., and Kuntz, I. D. 2009. Computational investigation of kinetics of cross-linking reactions in proteins: importance in structure prediction. *Biopolymers* **91**(1), pp. 68–77.
206. Kalkhof, S. and Sinz, A., 2008. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Analytical and bioanalytical chemistry*, **392**, pp.305-312.
207. Rappsilber, J. 2011. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of Structural Biology* **173**(3), pp. 530–40.
208. Dupree, E. J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B. A., and Darie, C. C. 2020. A critical review of bottom-up proteomics: the good, the bad, and the future of the field. *Proteomes* **8**(3), article number: 14 [no pagination].
209. Shevchenko, A., and Shevchenko, A. 2001. Evaluation of the efficiency of in-gel digestion of proteins by peptide isotopic labeling and MALDI mass spectrometry. *Analytical Biochemistry* **296**(2), pp. 279–283.
210. Rogers, J.C., and Bomgarden, R.D. 2016. Sample preparation for mass spectrometry-based proteomics; from proteomes to peptides. In: *Advances in Experimental Medicine and Biology*. New York: Springer, pp. 43–62.
211. Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Förster, F., and Aebersold, R. 2012. Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & Cellular Proteomics* **11**(3), article number: M111.014126 [no pagination].
212. Schmidt, R., and Sinz, A. 2017. Improved single-step enrichment methods of cross-linked products for protein structure analysis and protein interaction mapping. *Analytical and Bioanalytical Chemistry* **409**(9), pp. 2393–2400.
213. Fritzsche, R., Ihling, C. H., Götze, M., and Sinz, A. 2012. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Communications in Mass Spectrometry* **26**(6), pp. 653–658.
214. Bartman, C. E., Metwally, H., and Konermann, L. 2016. Effects of multidentate metal interactions on the structure of collisionally activated proteins: insights from ion mobility spectrometry and molecular dynamics simulations. *Analytical Chemistry* **88**(13), pp. 6905–691
215. Cavanagh, J., Benson, L. M., Thompson, R., and Naylor, S. 2003. In-Line Desalting Mass Spectrometry for the Study of Noncovalent Biological Complexes. *Analytical Chemistry* **75**(14), pp. 3281–3286.
216. Speicher, K., Kolbas, O., Harper, S., and Speicher, D. 2000. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *Journal of Biomolecular Techniques* **11**(2), pp. 74–86.
217. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., and Mann, M. 2007. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**, 2856– 2860.
218. Petrotchenko, E. V., Serpa, J. J., Cabecinha, A. N., Lesperance, M., and Borchers, C. H. 2014. “Out-Gel” Tryptic Digestion Procedure for Chemical Cross-Linking Studies with Mass Spectrometric Detection. *Journal of Proteome Research* **13**(2), pp. 527–535.
219. Bullock, J. M. A., Schwab, J., Thalassinos, K., and Topf, M. 2016. The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking mass spectrometry. *Molecular & Cellular Proteomics* **15**(7), pp. 2491–2500.
220. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. 2012. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **11**(12), pp. 1709–1723.
221. Makarov, A. 2019. Orbitrap journey: taming the ion rings. *Nature Communications* **10**, article number: 3743 [no pagination].
222. Fenn, J. B., Mann, M., Meng, C.K., Wong, S. F., and Whitehouse, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246** (4926), pp. 64–71.
223. Banerjee, S., and Mazumdar, S. 2012. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International Journal of Analytical Chemistry*, article number: 282574 [no pagination].
224. Nadler, W. M., Waidelich, D., Kerner, A., Hanke, S., Berg, R., Trumpp, A., and Rösli, C. 2017. MALDI versus ESI: The impact of the ion source on peptide identification. *Journal of Proteome Research* **16**(3), pp. 1207–1215.
225. Stieger, C. E., Doppler, P., and Mechtler, K. 2019. Optimized fragmentation improves the identification of peptides cross-linked by MS-cleavable reagents. *Journal of Proteome Research* **18**(3), pp. 1363–1370.

226. Iacobucci, C. et al. 2019. First community-wide, comparative cross-linking mass spectrometry study. *Analytical Chemistry* **91**(11), pp. 6953–6961.
227. Ihling, C. H., Piersimoni, L., Kipping, M., and Sinz, A. 2021. Cross-linking/mass spectrometry combined with ion mobility on a timsTOF Pro instrument for structural proteomics. *Analytical Chemistry* **93**(33), pp. 11442–11450.
228. Senko, M. W., et al. 2013. 2013. Novel parallelized quadrupole/linear Ion Trap/Orbitrap Tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical Chemistry* **85**(24), pp. 11710–11714.
229. Ngoka, L. and Gross, M.L., 1999. Multistep tandem mass spectrometry for sequencing cyclic peptides in an ion-trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, **10**(8), pp.732-746.
230. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. 2007. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**, pp. 709–712.
231. Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., Gygi, S. P. 2011. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & Cellular Proteomics* **10**(12), article number: M111.009910 [no pagination].
232. Michalski, A., Neuhauser, N., Cox, J., and Mann, M. 2012. A systematic investigation into the nature of Tryptic HCD spectra. *Journal of Proteome Research* **11**(11), pp. 5479–5491.
233. Arlt, C., Götze, M., Ihling, C., Hage, C., Schäfer, M., and Sinz, A. 2016. Integrated workflow for structural proteomics studies based on cross-linking/mass spectrometry with an MS/MS cleavable cross-linker. *Analytical Chemistry* **88** (16), 7930–7937.
234. Hodge, K., Ten Have, S., Hutton, L., Lamond, A. I. 2013. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics* **88**, pp. 92–103.
235. Beveridge, R., Stadlmann, J., Penninger, J. M., and Mechtler, K. 2020. A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes. *Nature Communications* **11**(1), pp. 1–9.
236. Yugandhar, K., Zhao, Q., Gupta, S., Xiong D., and Yu, H. 2021. Progress in methodologies and quality-control strategies in protein cross-linking mass spectrometry. *Proteomics* **21**, pp. 23–24.
237. Lenz, S., Sinn, L. R., O'Reilly, F. J., Fischer, L., Wegner, F., and Rappsilber, J. 2021. Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nature Communications* **12**, article number: 3564 [no pagination].
238. Nesvizhskii, A. I., Vitek, O., Aebersold. 2007. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* **4**(10), pp. 787–97.
239. Fischer, Lutz, and Rappsilber. 2017. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Analytical Chemistry* **89**(7), pp. 3829–3833.
240. Yugandhar, K., Wang, T., Wierbowski, S. D., Shayhidin, E. E., and Yu, H. 2020. Structure-based validation can drastically underestimate error rate in proteome-wide cross-linking mass spectrometry studies. *Nature Methods* **17**, pp. 985–988.
241. Kolbowski, L., Lenz, S., Fischer, L., Sinn, L. R., O'Reilly, F. J., and Rappsilber, J. 2022. Improved peptide backbone fragmentation Is the primary advantage of MS-cleavable crosslinkers. *Analytical Chemistry* **94**(22), pp. 7779–7786.
242. Merkley, E. D., Rysavy, S., Kahraman, A., Hafen, R. P., Daggett, A, and Adkins, J. N. 2014. Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Society* **23**(6), pp. 747–759.
243. Schiffrin, B., Radford, S. E., Brockwell and Calabrese, A. N. 2020. PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Science* **29**(8), pp. 1851–1857.
244. Jumper, J. et al. 2020. AlphaFold2. Critical Assessment of Techniques for Protein Structure Prediction: Abstract Book, pp. 22–24.
245. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. 2022. ColabFold: making protein folding accessible to all. *Nature Methods* **19**, pp. 1–4.
246. Svetlov, V. and Nudler, E., 2018. Analysis of the impact of molecular motions on the efficiency of XL-MS and the distance restraints in hybrid structural biology. *bioRxiv.*
247. Kahraman, A., Malmström, L., and Aebersold, R. 2011. Xwalk: Computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**(15), pp. 2163–2164.
248. Kahraman, A., Herzog, F., Leitner, A., Rosenberger, G., Aebersold, R., and Malmström, L. 2013. Cross-link guided molecular modeling with ROSETTA. *PLoS ONE* **8**(9), article number: 73411 [no pagination].
249. Biehn, S. E., and Lindert, S. 2022. Protein structure prediction with mass spectrometry data. *Annual Review of Physical Chemistry* **73**, pp. 1–19.
250. Bonvin, A. M. J. J., Karaca, E., Kastritis, P. L., and Rodrigues J. P. G. L. M. 2018. Defining distance restraints in HADDOCK. *Nature Protocols* **13**(7), article number: 1503 [no pagination].
251. Miernyk, J.A. and Thelen, J.J., 2008. Biochemical approaches for discovering protein–protein interactions. *The Plant Journal*, **53**(4), pp.597-609.
252. Chen, Z. A. et al. 2010. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal* **29**(4), pp. 717–26.
253. Gong, Z., Ye, S., Nie, Z., and Tang, C. 2020. The conformational preference of chemical cross-linkers determines the cross-linking probability of reactive protein residues. *The Journal of Physical Chemistry B* **124**(22), pp. 4446–4453.
254. Gong, Z., Ye, S, and Tang, C. 2020. Tightening the crosslinking distance restraints for better resolution of protein structure and dynamics. *Structure* **28**(10), pp. 1160–1167.
255. Marklund, E. G., Degiacomi, M. T., Robinson, C. V., Baldwin, A. J., & Benesch, J. L. (2015). Collision cross sections for structural proteomics. Structure, 23(4), 791-799.

256. Calabrese, A. N., et al. 2020. Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients. *Nature Communications* **11**(1), pp. 1–16.
257. Uba, A. I., and Yelekçi. 2019. Crystallographic structure versus homology model: a case study of molecular dynamics simulation of human and zebrafish histone deacetylase 10. *Journal of Biomolecular Structure and Dynamics* **38**(15), 4397–4406.
258. Paul, F., Thomas, T., and Roux, B. 2020. Diversity of Long-Lived Intermediates along the Binding Pathway of Imatinib to Abl Kinase Revealed by MD Simulations. *Journal of Chemical Theory and Computation* **16**(12), pp. 7852–7865.
259. Blanc, F., Isabet, T., Benistry, H., and Houdusse, A. 2018. An intermediate along the recovery stroke of myosinVI revealed by X-ray crystallography and molecular dynamics. *PNAS* **115**(24), pp. 6213–6218.
260. Case, D. A., et al. 2020. *Amber 2020 Reference Manual*. University of California, San Francisco.
261. Cheng, Y., Zhang, F., Chen, Q., Gao, J., Cui, W., Ji, M., and Tung, C. 2011. Structural Basis of Specific Binding between Aurora A and TPX2 by Molecular Dynamics Simulations. *Journal of Chemical Information and Modelling* **51**(10), pp.2626–2635.
262. Hospital, A., Goñi, J. R., Orozco, M., Gelpi, J. L. 2015. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry* **8**, pp. 37–47.
263. Bryant, S. H., Lawrence, C. E. 1993. An empirical energy function for threading protein sequence through the folding motif. *Proteins* **16**(1), pp. 92–112.
264. Davidchack, R. L., Handel, R. and Tretyakov, M. V. 2009. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics* **130**(23), article number: 234101 [no pagination].
265. Tobias, D. J., Martyna, G. J., and Klein, M. L. 1993. Molecular dynamics simulations of a protein in the canonical ensemble. *The Journal of Physical Chemistry* **97**(49), pp. 12959-12966.
266. Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. 2015. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **11**(4), pp. 1864–1874.
267. Pearlman, D. A., et al. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91**(1-3), pp. 1–41.
268. Babu, M. M., van der Lee, R., Sanchez de Groot, N., and Gsponer, J. 2011. Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology* **21**(3), pp. 432-40.
269. Choi, U. B., Sanabria, H., Smirnova, T., Bowen, M. E., Weninger, K. R. 2019. Spontaneous Switching among Conformational Ensembles in Intrinsically Disordered Proteins. *Biomolecules* **9**(3), article number: 114 [no pagination].
270. Mitra, G. 2021. Emerging Role of Mass Spectrometry-Based Structural Proteomics in Elucidating Intrinsic Disorder in Proteins. *Proteomics* **21**(3-4), article number: 2000011 [no pagination].
271. Mittag, T., Kay, L. E., and Forman-Kay, J. D. 2010. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition: An Interdisciplinary Journal* **23**(2), pp. 105-116.
272. Gilburt, J. A. H., et al. 2017. Dynamic Equilibrium of the Aurora A Kinase Activation Loop Revealed by Single-Molecule Spectroscopy. *Angewandte Chemie* **129**(38), pp. 11567–11572.
273. Ward, S. 2020. Targeting the Aurora A/N-Myc Protein-Protein Interaction in Neuroblastoma. (Doctoral dissertation, University of Leeds).
274. Burgess, S. G., and Bayliss, R. 2015. The structure of C290A: C393A Aurora A provides structural insights into kinase regulation. *Acta Crystallographica Section F: Structural Biology Communications* **71**(3), pp. 315-319.
275. Ziemianowicz, D. S., Ng, D., Schryvers, A. B., Schriemer, D. C. 2019. Photo-Cross-Linking Mass Spectrometry and Integrative Modeling Enables Rapid Screening of Antigen Interactions Involving Bacterial Transferrin Receptors. *Journal of Proteome Research* **18**(3), pp. 934–946.
276. Sinz, A. 2014. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev. Proteomics* **11**, pp. 733–743.
277. Swart, C., Martinez-Jaime, S., Gorka, M., Zander, K., Graf, A. 2018. Hit-Gel: Streamlining in-gel protein digestion for high-throughput proteomics experiments. *Scientific Reports* **8,** article number: 8582 [no pagination].
278. Lauber, M. A. and Reilly, J. P. 2011. Structural analysis of a prokaryotic ribosome using a novel amidinating cross-linker and mass spectrometry. *Journal of Proteome Research* **10**(8), 3604–16.
279. Arlt, C., Ihling, C. H., and Sinz, A. 2015. "Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry." *Proteomics* **15**(16), pp. 2746-2755.
280. Herzog, F., et al. 2012. Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **337**(6100), pp. 1348–52.
281. Rehkamp, A., Tänzler, D., Iacobucci, C., Golbik, R. P., Ihling, C. H., and Sinz, A. 2018. Molecular details of Retinal Guanylyl Cyclase 1/GCAP-2 interaction. *Frontiers in Molecular Neuroscience* **11**, article number: 330 [no pagination].
282. Bastian, B. et al. 2018. Structure and mechanism of the two-component α-helical pore-forming toxin YaxAB. *Nature Communications* **9**(1), pp. 1-14
283. Wei, A, Iacobucci, C, Schultze, W., Ihling, C., Arlt, C., and Sinz, A. 2022. Different Oligomeric States of the Tumor Suppressor p53 Show Identical Binding Behavior Towards the S100β Homodimer. *ChemBioChem*, article number: e202100665 [no pagination].
284. Wang, Y., Li, X., Liu, Y.H., Richardson, D., Li, H., Shameem, M. and Yang, X., 2016, November. Simultaneous monitoring of oxidation, deamidation, isomerization, and glycosylation of monoclonal antibodies by liquid chromatography-mass spectrometry method with ultrafast tryptic digestion. *MAbs* **8**(8), pp. 1477-1486).
285. Ying, Y. and Li, H. 2020. Recent progress in the analysis of protein deamidation using mass spectrometry. *Methods* **200**, pp. 42–57.
286. Steigenberger, B., et al. 2019. PhoX: an IMAC-enrichable cross-linking reagent. *ACS central science* **5**(9), pp. 1514-1522.

287. Combe, C. W., Fischer, L., and Rappsilber, J. 2015. XiNET: cross-link network maps with residue resolution. *Molecular & Cellular Proteomics* **14**(4), pp. 1137-1147.
288. Piotrowski, C., Moretti, R., Ihlinh, C. H., Haedicke, A., Liepold, T., Lipstein, N., Meiler, J., Jahn, O., and Sinz, A. 2020. Delineating the Molecular Basis of the Calmodulin–bMunc13-2 Interaction by Cross-Linking/Mass Spectrometry-Evidence for a Novel CaM Binding Motif in bMunc13-2. *Cells* **9**(1), article number 136 [no pagination].
289. Guo, X., Bandyopadhyay, P., Schilling, B., Young, M.M., Fujii, N., Aynechi, T., Guy, R.K., Kuntz, I.D. and Gibson, B.W., 2008. Partial acetylation of lysine residues improves intraprotein cross-linking. *Analytical chemistry*, **80**(4), pp.951-960.
290. Hofmann, T., Fischer, A. W., Meiler, J., and Kalkhof, S. 2015. Protein structure prediction guided by crosslinking restraints–A systematic evaluation of the impact of the crosslinking spacer length. *Methods* **89**, pp 79-90.
291. Jaffe, J., Keshishian, H., Chang, B., Addona, T. A., Gillette, M. A., and Carr, S. A. 2008. Accurate Inclusion Mass Screening. *Molecular and Cellular Proteomics* **7**(10), pp. 1952–1962.
292. Hevler, J. F., et al. 2021. Selective cross-linking of coinciding protein assemblies by in-gel cross-linking mass spectrometry. *The EMBO Journal* **40**(4), article number: 106174 [no pagination].
293. Uhlenbrock, N., et al. 2019. Structural and chemical insights into the covalent-allosteric inhibition of the protein kinase Akt. *Chemical science* **10**(12), pp. 3573–3585.
294. Stieger, C.E., Doppler, P. and Mechtler, K., 2019. Optimized fragmentation improves the identification of peptides cross-linked by MS-cleavable reagents. *Journal of proteome research*, **18**(3), pp.1363-1370.
295. Mendes, M. L., et al. 2019. An integrated workflow for crosslinking mass spectrometry. *Molecular systems biology* **15**(9), article number: 8994 [no pagination].
296. Peterson, R. T., and Schreiber, S. L. 1999. Kinase phosphorylation: keeping it all in the family. *Current biology* **9**(14), pp. 521–524.
297. Lomant, A. J. and Fairbanks, G. 1976. Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [35S]diothiobis(succinimidylpropionate). *Journal of Molecular Biology* **104**(1), pp. 243–261.
298. Liu, R., et al. 2019. Assessing lysine and cysteine reactivities for designing targeted covalent kinase inhibitors. *Journal of the American Chemical Society* **141**(16), pp. 6553-6560.
299. Tüting, C., Iacobucci, C., Ihling, C. H., Kastritis, P. L., and Sinz, A. 2020. Structural analysis of 70S ribosomes by cross-linking/mass spectrometry reveals conformational plasticity. *Scientific Reports* **10**(1), article number: 12618 [no pagination].
300. Fürsch, J., et al. 2021. Structural probing of Hsp26 activation and client binding by quantitative cross-linking mass spectrometry. *Analytical Chemistry* **93**(39), pp. 13226-13234.
301. Bullock, J. M. A., Sen, N., Thalassinos, K., and Topf, M. 2018. Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. *Structure* **26**(7), pp. 1015–1024.
302. Sinnott, M., et al. 2020. Combining information from crosslinks and monolinks in the modeling of protein structures. *Structure* **28**(9), pp. 1061-1107.
303. Fabris D., and Yu, E. T. 2010 Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D. *Journal of Mass Spectrometry* **45**(8), pp. 841–860.
304. Mädler, S., Seitz, M., Robinson, J., and Zenobi, R. 2010 Does Chemical Cross-Linking with NHS Ester Reflect the Chemical Equilibrium of Protein–Protein Noncovalent Interactions in Solution? *Journal of the American Society of Mass Spectrometry* **21**(10), pp. 1775–1783.
305. Dashnaw, C, M., Koone, J. A., Abdolvahabi, A., Shaw, B. F. 2021. Measuring how two proteins affect each other's net charge in a crowded environment. *Protein Science* **30**(8), pp. 1594–1605.
306. Chu, M.L., Lang, Z., Chavas, L.M., Neres, J., Fedorova, O.S., Tabernero, L., Cherry, M., Williams, D.H., Douglas, K.T. and Eyers, P.A., 2010. Biophysical and X-ray crystallographic analysis of Mps1 kinase inhibitor complexes. *Biochemistry*, **49**(8), pp.1689-1701.
307. Pegram, et al. 2019. Activation loop dynamics are controlled by conformation-selective inhibitors of ERK2. *PNAS* **116**(31), pp. 15463–15468.
308. Tomlinson, L. J., et al. 2022. Exploring the Conformational Landscape and Stability of Aurora A Using Ion-Mobility Mass Spectrometry and Molecular Modeling. *Journal of the American Society for Mass Spectrometry* **33**(3), pp. 420-435.
309. Lössl, P., et al. 2016. Deciphering the interplay among multisite phosphorylation, interaction dynamics, and conformational transitions in a tripartite protein system. *ACS Central Science* **2**(7), pp. 445-455.
310. de Leon-Boenig, G., Bowman, K.K., Feng, J.A., Crawford, T., Everett, C., Franke, Y., Oh, A., Stanley, M., Staben, S.T., Starovasnik, M.A. and Wallweber, H.J., 2012. The crystal structure of the catalytic domain of the NF-κB inducing kinase reveals a narrow but flexible active site. *Structure*, **20**(10), pp.1704-1714.
311. Tan, K. P., Varadarajan, R., and Madhusudhan, M. S. 2011. DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Research* **39**, pp. 242–248.
312. Boll, L.B. and Raines, R.T., 2022. Context-Dependence of the Reactivity of Cysteine and Lysine Residues. *ChemBioChem*, **23**(14), p.e202200258.
313. Lepsik, M. et al. 2019. Induction of rare conformation of oligosaccharide by binding to calcium-dependent bacterial lectin: x-ray crystallography and modelling study. *European Journal of Medicinal Chemistry* **177**, pp. 212–220.
314. Moffett, A. S., Bender, K. W., Huber, S. C., and Shukla, D. 2017. A Molecular dynamics simulations reveal the conformational dynamics of Arabidopsis thaliana BRI1 and BAK1 receptor-like kinases. *Journal of Biological Chemistry* **292**(3), pp. 12643–12652.
315. Ledoux, J., Trouvé, A., and Tchertanov, L. 2021 Folding and Intrinsic Disorder of the Receptor Tyrosine Kinase KIT Insert Domain Seen by Conventional Molecular Dynamics Simulations. *International Journal of Molecular Science* **22**(14), article number: 7375 [no pagination].

316. Ravichandran, A., Araque, J., and Lawson, J. 2022. Predicting the functional state of protein kinases using interpretable graph neural networks. *Biophysical Journal* **121**(3), article number: 1148, [no pagination].
317. Graziadei, A., and Rappsilber, J. 2021. Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure* **30**(1), pp. 37–54.
318. Musavizadeh, Z., Grottesi, A., Guarduaglini, and Paiardini, A. 2021. Phosphorylation, Mg-ADP, and Inhibitors Differentially Shape the Conformational Dynamics of the A-Loop of Aurora-A. *Biomolecules* **11**(4), article number: 567 [no pagination].
319. Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research*, **31**(13), pp.3381-3385.
320. Ruff, K.M. and Pappu, R.V., 2021. AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology*, **433**(20), p.167208.
321. Modi, V. and Dunbrack Jr, R.L., 2022. Kincore: a web resource for structural classification of protein kinases and their inhibitors. *Nucleic Acids Research*, **50**(D1), pp.D654-D664.
322. Ruotolo B. T., Hyung S. J., Robinson P. M., Giles K., Bateman R. H, and Robinson, C. V. 2007. Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angewandte Chemie (International Edition in English)* **46**(42), pp. 8001-8004.
323. Meagher, K. L., Redman, L. T., and Carlson, H. A. 2003. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry* **24**(9), 1016-1025.
324. http://amber.manchester.ac.uk/
325. Homeyerr, N., Horn, A. H. C., Lanig, H., and Sticht, H. 2006. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of Molecular Modeling* **12**, pp. 281–289.
326. Lake, E.W., 2021. Biochemical Context Drives Inhibitor Selectivity in Aurora A Kinase (Doctoral dissertation, University of Minnesota).
327. Ocasio, C.A., Warkentin A.A., McIntyre, P.J., Barkovich, K.J., Vesely, C., Spencer, J., Shokat, K.M., BaylissR. 2018. Type II Kinase Inhibitors Targeting Cys-Gaktekeeping Kinases Display Orthogonality with Wild Type and Ala/Gly-Gatekeeper Kinases. *ACS Chem Biol* **13**, pp. 2956–2965.
328. Dodson, C. A., Kosmopoulou, M., Richards, M. W., Atrash, B., Bavetsias, J. B., Bayliss, R. 2010. Crystal Structure of an Aurora-A Mutant that Mimics Aurora-B Bound to Mln8054: Insights into Selectivity and Drug Design. *The Biochemical Journal* **427**(1), pp. 19–28
329. Burgess, S. G., Oleksy, A., Cavazza, T., Richards, M. W., Vernos, I., Matthews, D., and Bayliss, R. 2016. Allosteric inhibition of Aurora-A kinase by a synthetic vNAR domain. *Open Biology* **6,** article number: 160089 [no pagination].
330. de Groot, et al. 2015. A Cell Biologist's Field Guide to Aurora Kinase Inhibitors. *Frontiers in Oncology* **5**, article number: 285 [no pagination].
331. Gong, X. et al. 2019. Aurora A Kinase Inhibition Is Synthetic Lethal with Loss of the RB1Tumor Suppressor Gene. *Cancer Discovery* **9**(2), pp. 248–263.
332. McCabe, J. W., et al. 2020. First-Principles Collision Cross Section Measurements of Large Proteins and Protein Complexes. *Analytical Chemistry* **92**(16), pp. 11155–11163.