Structural characterisation of human protein complexes in nutrient signalling by X-ray crystallography, cryo-EM and cross-linking mass spectrometry

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Yana Dem‘yanenko

2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Timm Maier, Michel Steinmetz, Oliver Hantschel

Basel, 30 March 2021

_____

Prof. Dr. Marcel Mayor

i. Abstract

Nutrient availability regulates cellular growth and metabolism through modulation of the mTORC1 activity. Aberrant activation of this serine/threonine kinase complex has been implicated in a host of metabolic disorders such as cancer and type 2 diabetes. Much work has been done to understand the processes involved in metabolic disease pathogenesis and to develop treatment strategies. However, limited effectiveness of available therapies targeting mTORC1 highlights the need for better understanding of the mechanistic aspects of regulation of this pathway, as well as complex crosstalk and feedback mechanisms towards other systems. In this work I have focused on high-resolution structural characterisation of proteins and protein-protein interactions involved in glucose metabolism and amino acid sensing pathway, as well as development of a cross-linking mass spectrometry method for studying transient and dynamic protein interactions. I have solved a structure of a naturally occurring R42H mutation in the hexokinase 2 that was previously shown to decrease enzyme activity and subsequently lead to development of glucose intolerance in Mexican cavefish. With help of differential scanning fluorimetry I have shown that this mutation destabilises the protein without a measurable effect on the overall substrate and product binding. However, it is possible that the mutation influences the overall flexibility and regulation of the enzyme's activity, which requires further investigation. Characterisation of RagAC-Raptor biding through generation of several nucleotide-binding mimicking mutation combinations allowed in vitro assembly of the full lysosomal mTORC1 complex for single particle cryo-EM structure determination which is in agreement with currently proposed models of the mega complex. Using the established RagAC system I have developed a protocol for *in vitro* zero-length cross-linking, enrichment and identification of cross-linked peptides with 1,1-carbonyldiimidazole (CDI) and in vivo incorporation of photo-leucine and photo-methionine incorporation for cross-linking studies of similar dynamic systems. The wide range of cross-link distances (7- 140 Å) obtained in the CDI cross-linking experiments have revealed a likely conformational heterogeneity and flexibility of the complex in solution, highlighting the weaknesses of this method for studying heterogeneous mixtures and dynamic protein complexes. Although on a peptide level there was a substantial overlap, the identification of the cross-link site differed considerably between different software, therefore the results of similar studies should be interpreted with caution. The analysis of photo-leucine and photo-methionine incorporation rates revealed overall values on par with previous studies (of up to 10% and 30% respectively). More than 85% efficiency of UV activation, with some variation depending on the residue location, is promising for further studies. However, reduced protein production, large variations between different residue positions, and poor identification of modifications on less abundant peptides should be considered when designing cross-linking experiments with photo-amino acids.

ii. Acknowledgements

My path to completing this thesis has certainly not been easy. I am grateful to have been awarded the Biozentrum Fellowship for Excellence to enable me to come to Switzerland at all. I am also grateful for the community University of Basel and Biozentrum have created and all the people that I could meet along the way: other PhD students, PhD representatives, retreat organisation committee, and brilliant faculty members. Particularly, I would like to thank my thesis advisor Prof. Timm Maier for supporting my projects, even if they did not completely align with his vision, allowing me to learn in a relaxed environment. The work produced in this thesis would not have been possible without him, just as it would not be possible without my teammates at the Maier lab. I would like to give special thanks to Roman Jakob, who was always a pleasure to be around and incredibly helpful, providing valuable day-to-day advice; former members Edward Stuttfeld and Stefan Imseng who have made getting started in the lab a breeze; Shubham Singh for his expertise and thoughtfulness when providing advice, training and his general honest support throughout my time at the Biozentrum. The cross-linking mass-spectrometry part of my project could not have happened without the help and support of Bastian Kohl (S. Hiller lab) and Danilo Ritz (PCF), so I would like to give special thanks to them for invaluable discussions and support.

The writing of this manuscript was a great challenge for me, especially during the pandemic, amidst lockdowns and restrictions, away from home and family. Throughout this unique time, my friends and neighbours Sebastian Till, Livia Marras, Jean Noel Gos, Natalie Carducci, Tin Cugelj, Perrine Devillers and the rest of our lovely Oetlinger community. They have helped me stay sane and positive by a) allowing almost exclusive use of our sofa and living room, b) providing the best possible entertainment in the shape of “Monday night Curd dinners”, late-night and evening jazz concerts in the living room and our backyard over the summer and some of the most interesting conversations, c) just being lovely people it was a pleasure to be around.

Being the first person in the family with a Master’s degree, much less a PhD, it is certainly a lot of pressure. I would link to thank my family in Ukraine who have always supported me in whatever I set out to do, setting high expectations and pushing me to do better. It’s been a tough year for all of us and I’m glad to have been able to help organise and join the New year’s feast this year and hope to be able to visit more often in the coming years.

Last, but not least, I would like to thank my partner Connor Sharp, who has been there for me from start to finish. Although not in flesh, but certainly in spirit. Thank you for being a wonderful constant in my life among all the chaos and uncertainty. Spending nearly the whole year apart due to restrictions has been difficult for both of us. Nonetheless, I’ve always felt your support even across the English Channel and a couple of hundred miles. Thank you for our late-night “work-chats” with Only Connect breaks, thanks you for being there for me all this time.

iii. Table of Contents

1 Introduction

1.1 Overview of metabolic regulation in mammalian cells

During their lifetime, all cells need to balance their activity and growth with availability of resources and environmental factors. For example, in the absence of nutrients each cell needs to ensure that the resources are used most efficiently and in a way that would be most beneficial to either the cell itself or the community. Under these circumstances, all most energy-expensive anabolic processes such as protein synthesis, and DNA replication for cell division would be typically down-regulated, while resource-management processes such as gluconeogenesis and autophagy would be upregulated instead. To ensure optimal outcome in the perpetually changing environment, the response to internal and external stimuli must be rapid, precise, and flexible.

Protein phosphorylation is a fundamentally simple, flexible and robust means for such regulatory responses in eukaryotic cells. The process of the phosphate group transfer from an ATP molecule to a specific residue of an effector protein is catalysed by a family of enzymes called kinases. The reaction is rapid and results in an extra negative charge added to the surface of a protein substrate which may influence its conformation, enzymatic activity, protein-DNA or protein-protein interactions. To reverse the effect of a phosphorylation, another enzyme family - the phosphatases, can quickly hydrolyse the bond and release the phosphate group. Both, kinases and phosphatases can be directly regulated in response to the environment and are often substrates of themselves or other members of the family ensuring the flexibility of such a signalling pathway. The specificity of the response is usually conferred by the amino acid sequence of substrate proteins, where each of the kinases only phosphorylates substrates with a specific recognition motif. Substrates of the same kinase are typically a structurally and functionally diverse family of proteins with little similarity to each other, except at their respective kinase recognition sites. Many of the substrates may be regulated by more than one kinase and therefore harbour several different recognition motifs. This provides the required complexity to the system and ensures the robustness of the signalling response.

At the simplest level, this kind of signalling system consists of receptor proteins which activate the transducing kinases, the messenger kinases themselves, and, of course, substrate effector proteins. Receptor proteins may either be membrane integrated (e.g., tyrosine-receptor kinases, G-protein coupled receptors), or intracellular (e.g., AMP activated kinase (AMPK), cAMP-dependent protein kinase (PKA)). They either directly or indirectly influence the activity of downstream regulator kinases such as target of rapamycin kinase (TOR), protein kinases B and C, and extracellular signal-regulated kinases 1/2 (ERK1/2) etc. Most of the signal integration happens at this level through the crosstalk between different pathways and multiple feedback loops. These master regulatory kinases in turn phosphorylate or influence the phosphorylation of ultimate effector proteins: transcription factors, metabolic enzymes to produce the desired

outcome for the cell (e.g., in starvation conditions protein synthesis and glycolysis would be downregulated while nutrient transport, and autophagy (recycling mechanisms) in turn upregulated).

Energy (ATP), food and building material (glucose, amino acids and lipids), as well as stress level (both environmental and metabolic) are the main inputs to the kinase signalling pathways which influence cell growth and proliferation. The energy availability can be sensed indirectly through a change in ratios of ATP to its metabolites ADP and AMP by the AMP-dependent kinase (AMPK)[1]. Binding of AMP or ADP (in absence of the former) has been shown to promote activating phosphorylation of Thr172 by upstream kinases such as LBK1 as well as block dephosphorylation at this residue[2-7]. Additionally, in many species AMP binding also promotes allosteric activation of already phosphorylated AMPK[8, 9]. When activated, it stimulates energy-producing (catabolic) processes and downregulates energy consumption (anabolism) by phosphorylating its downstream targets[10, 11]. The response to nutrient availability (e.g. glucose, amino acids, insulin, etc.) is mainly regulated by the target of rapamycin kinase (TOR)[12]. When activated, this pathway promotes anabolic activities such as protein and glycogen synthesis while inhibiting autophagy and gluconeogenesis[13]. At the same time, the response to environmental and metabolic stress is regulated through mitogen-activated protein kinase (MAPK) pathways which can interfere with normal nutrient and energy signalling and collectively produce an Integrated Stress Response condition that results either in induced expression of repair genes or death (apoptosis)[14].

Dysregulation of metabolic pathways is frequently observed in cancers alongside the phenomenon associated with increased glucose uptake which is known as the “Warburg effect”. It represents a shift in the mechanism of ATP production from mainly respiration (oxidative phosphorylation) to up to equal parts respiration and fermentation (final step of glycolysis: pyruvate to lactic acid conversion) even in aerobic conditions[15-17]. Despite extensive research done in the last century, the ontology of this metabolic rewiring remains under debate. There is, however, no doubt that such dual energy production mode together with upregulated nutrient uptake is utilised by many rapidly growing cells[18].

When the demand for amino acids, nucleotides and fatty acids is high, the processes geared towards increasing their intracellular concentration are upregulated concurrently with the processes that utilise them[19]. If the regulatory mechanisms fail, the cell enters an unsustainable vicious cycle: excessive growth -> high nutrient demand -> upregulated intake -> high intracellular nutrient availability -> growth. Potential intervention points here are reduction of nutrient and energy supply to quickly exhaust the resources of rapidly growing cells, or inhibition of growth to downregulate the nutrient demand. The advantage of the first strategy is that it is inherently most effective against aberrant rapidly growing cells. However, implementation of such an intervention is challenging due to multiplicity and redundancy of metabolic pathways involved[20]. While inhibiting

overall growth may seem easier through inhibition of the central kinase mTOR, due to complex nature of its signalling pathways as well as its ubiquitous expression in all cell types, currently available therapies have only shown mild efficacy and are sometimes associated with severe toxic effects[21, 22]. In many cases this is due to either drug-induced metabolic rewiring, or cytostatic nature of such treatments as mTOR also regulates the nutrient uptake and recycling pathways. One possible strategy to generate the metabolic imbalance required for cytotoxic effect could be selective targeting of a subset of mTOR functions together with independent inhibitors of nutrient uptake, synthesis or autophagy[22]. Understanding the complex relationship between the regulatory mechanisms involved in these signalling branches is, therefore, paramount.

The work described in this thesis was done in relation to mammalian signalling pathways mediated by amino-acid and glucose availability. At their core lies the target of rapamycin kinase (TOR, or mTOR in mammalian cells). It is a conserved phosphatidylinositol 3-kinase-related kinase discovered in yeast in 1991 as the protein responsible for inhibition of cell proliferation by an immunosuppressant rapamycin[23]. Three years later its analogue was found in human cells by four other independent groups[24-27]. The action of TOR is directly opposed by that of AMPK in a yin-yang fashion[12]. While AMPK predecessors can be found in all domains of life from bacteria and archaea to humans, TOR mediated branch only appears in eukaryotes together with mitochondrial acquisition [28].

mTOR is involved in regulation of cell growth and proliferation, motility, metabolism, proteogenesis, and autophagy as the core component of two structurally and functionally distinct complexes: TORC1 and TORC2 (Fig 1.1). The latter is involved in regulation of lipogenesis, glycolysis and cytoskeleton homeostasis, and is stimulated mainly via growth

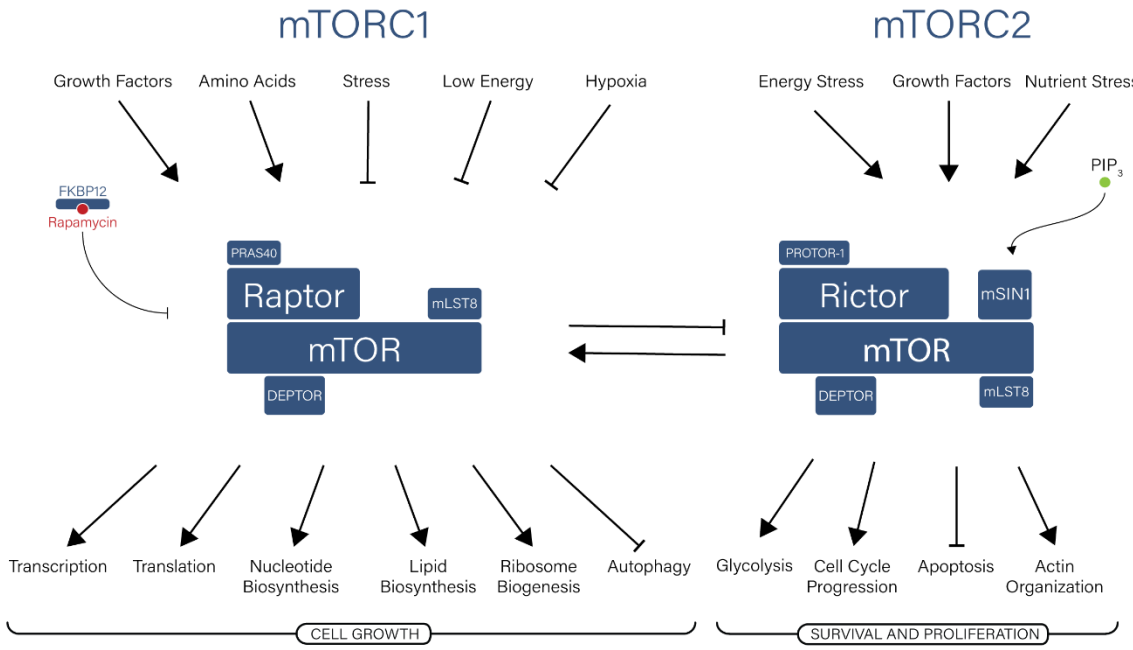

Figure 1.1.2.1.1 Organisation of mTORC1 and mTORC2 complexes and their signalling input and output. Adapted from a figure provided by M. Wälchli and S. Imseng.

factors and insulin[13]. TORC1 complex, on the other hand, is regulated by the presence of amino acids and glucose and is known to impact protein synthesis and DNA replication. Mutations in any of the components of this pathway have been known to cause growth defects, tumour development and contribute to metabolic disorders[13, 29].

TORC1 comprises TOR kinase, Raptor (regulatory associated protein of TOR), and LST8. Raptor is known to act as a substrate recognition protein while LST8 acts to enhance mTOR activity and stabilise mTOR and Raptor interaction[30]. In TORC2, Raptor is replaced by a rapamycin-insensitive substrate binding protein Rictor[31, 32].

1.2 Amino acid signalling

Amino acids play a vital role in multiple metabolic processes inside the cell. Not only are they needed for protein synthesis, but they also serve as intermediate metabolites in several essential biosynthetic pathways (e.g., sphingolipids, nucleotides, neurotransmitters, nicotinamide compounds, hormones etc.).

Although mTORC1 is the essential regulator of cell proliferation in response to intracellular amino acid level, none of the components of the complex can act as their direct sensors. Instead, in presence of amino acids, mTORC1 activity is upregulated through both enzyme activation and suppression of inhibition mediated by a host of effector proteins.

Activation of mTORC1 kinase takes place at the lysosome through an interaction with a membrane bound small GTPase Rheb[33-37]. This binding causes a conformational change in mTOR that results in more efficient enzymatic activity as well as promotes access of its substrates to the binding site on Raptor[33, 36-38]. While Rheb-mTORC1 interaction is dependent on presence of amino acids, it is not influenced by them directly, rather through a cascade of protein-protein interactions that leads to relocation of mTORC1 to the lysosomal membrane where Rheb resides.

1.2.1 Amino acid sensors

Free amino acids can become available in the cytoplasm through import, synthesis, and protein recycling systems. The signalling pathways employed by the cells are specific for each amino acid. Conversely, their influences on regulation of mTORC1 activity are also distinct. An overview of this system is shown in figure 1.2.1.

There are two main groups of amino acids: essential and nonessential. Essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) cannot be synthesised by the cells and therefore must come from the outside. Nonessential amino acids can be taken up by the cells, but they can also be synthesised from other metabolites or other amino acids inside the cell. These include alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine, and tyrosine. While amino acid dependent branch of mTOR signalling network is extensively studied, due to its complexity it is not yet entirely understood.

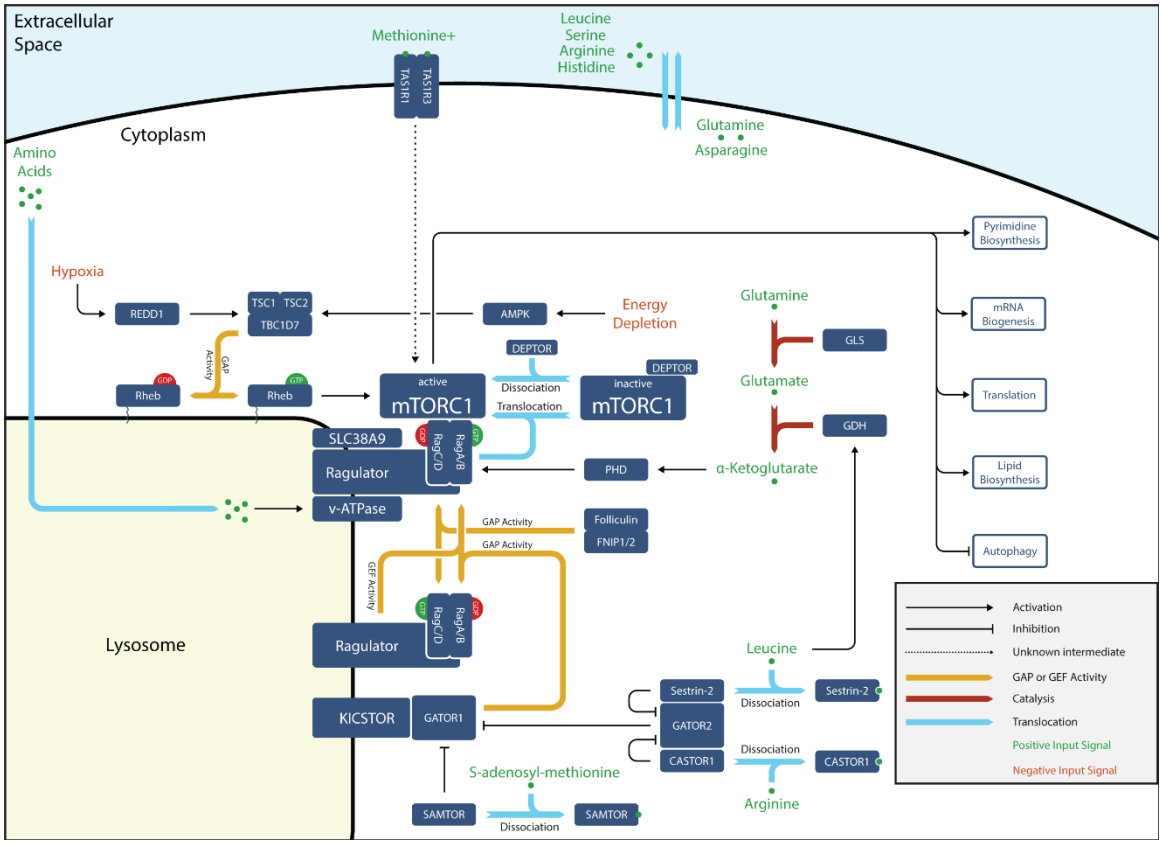

Figure 11.2.1.1 Overview of amino acid signalling to mTORC1. Translocation and activation of the mTORC1 is regulated by a cascade of interactions. Amino acids shown in green bind to sensor proteins and transduce the signal via inhibition of mTORC1 negative regulators (blunt end arrow). Enzyme activity of active translocation are shown in bold. Feedback loops, activation signalling interactions and downstream processes are marked with slim arrows. Adapted and from a figure by S. Imseng and F. Mangia

It is, however, known that mTORC1 signalling is influenced by branched chain amino acids (leucine, isoleucine and valine) as well as arginine and methionine[39-45]. In addition, it is indirectly regulated by glutamine and asparagine metabolism as they serve as amino acid exchange factors in import of leucine and other essential amino acids as well as arginine, serine and histidine[39]. Extracellular amino-acid levels, particularly methionine have also been shown to play a role in mTORC1 activation and consequently upregulation of protein and nucleotide synthesis[46].

The most studied and well-understood mechanism of amino acid signal transduction to mTORC1 is through mediation of the nucleotide binding state of RagAC GTPases through guanidine exchange factors and GTPase activating proteins (GEFs and GAPs)[47-52]. The activity of these proteins is mediated by intracellular amino acid sensors.

In addition to cytoplasmic sensors, amino acid transporters in cytoplasmic and lysosomal membranes also play a role in activating mTORC1[53]. For example, SLC38A9 upon binding arginine promotes efflux of glutamine from the lysosome and interacts with RagAC to promote FLCN:FNIP mediated GTP hydrolysis by RagC[54-57]. Extracellular levels

of methionine are thought to be sensed through the TAS1R1-TAS1R3 complex which is an extracellular amino-acid sensor for the majority of the 20 amino acid[46]. It has been shown to be required for mTORC1 translocation to the lysosome and subsequent activation[58].

Rag-independent pathways for mTORC1 activation are also known: Arf1 and vATPase - mediated glutamine-dependent mTORC1 activation, as well as Rab1A-mediated activation[42, 59-63].

1.2.2 Rag GTPases – conserved shuttle that recruits mTORC1 to the lysosome and connects it to amino acid signalling

The recruitment of mTORC1 to the lysosomal surface is mediated by a constitutive heterodimer of small GTPases RagA or RagB with RagC or RagD through its ability to bind to both: Raptor on the mTORC1, and Ragulator complex at the lysosome[64-66].

Rag proteins are members of the Ras-related GTPase family and contain a nucleotide binding domain at their N-terminus and a dimerization domain at their C-terminus.

RagA and RagB as well as RagC and RagD have high sequence identity to each other (>97% between RagA and RagB, and >79% between RagC and RagD), but <25% identity to their partner (i.e., RagA to RagC or RagB to RagD). Although sequence identity between the partner proteins is low, as is the identity between the human proteins and their yeast counterparts (~55 and 45% to Gtr1 and Gtr2), both adopt a similar tertiary structure which is conserved from yeast to human[67-72].

The dimerization domains have “roadblock” architecture like that of the Lamtor2-3 and 4-5. This organisation allows RagA to dimerise with RagC in a way that results in an anti-parallel orientation of the N-terminal GTPase(G)-domains. Although it has been shown that RagA can interact with both RagC and RagD, and the same is true for RagB, there seems to be a preference of RagA for RagC and RagB for RagD. As the most studied combination of the GTPases is RagAC[40, 66, 73], for simplicity reasons both dimers would be henceforth referred to as RagAC assuming what is true for this combination also translates to RagBD.

The tertiary structure of the G-domains or RagA-D resembles that of other Ras-related GTPases with six β-strands and 4 α-helices providing the scaffold and two semi-ordered switch regions (each contains a short α-helix) directly involved in nucleotide binding. Upon GTP hydrolysis, both switch I and II undergo large conformational changes and move away from the nucleotide binding site releasing the Mg^{2+} ion and lowering the affinity for product nucleotide (GDP)[67, 69]. This can be seen in the crystal structure of $RagA^{GTP}$-$RagC^{GDP}$ where GTP-bound RagA(Q66L) shows ordered switch I and switch II regions, while the same regions of the GDP-bound RagC T90N were disordered and not visible in the structure (Fig 1.2.2). Although the switch I region was ordered in the structure of the RagC(S75N) mutant, its conformation was distinct from that in the GTP bound RagA.

Presence of amino acids influences the nucleotide binding state of the RagAC complexes which acts as a switch for mTORC1 translocation to the lysosome through guanine nucleotide exchange factors and GTPase activating proteins (GEFs and GAPs) [64, 74, 75]. Only GTP-bound RagA/B in complex with GDP-bound RagC/D can bind Raptor and recruiting mTORC1 to the lysosomal surface[40].

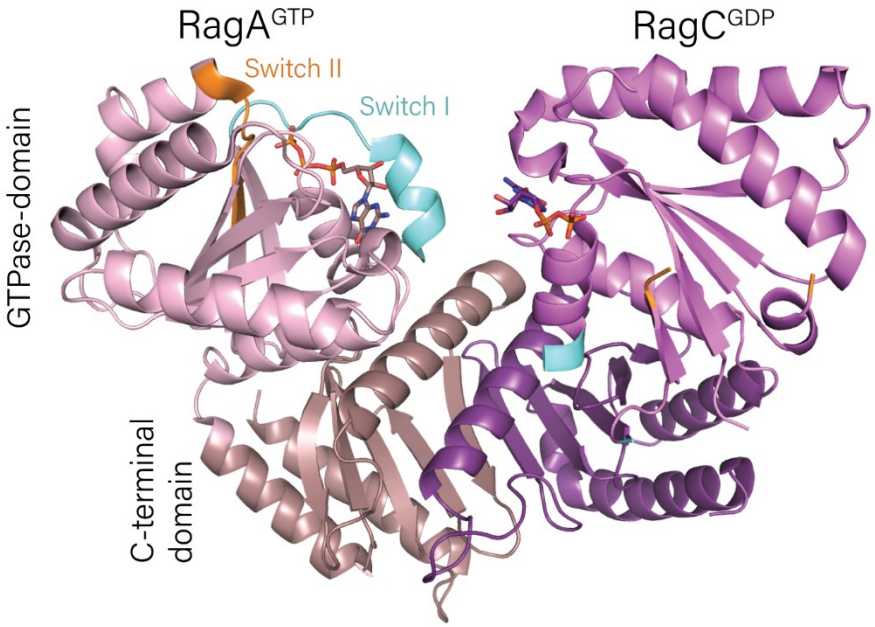

Figure 1.2.2.1 Crystal structure of RagA GTP(Q66L)-RagC GDP(T90N) (PDBid: 6S6A). Where switch I and II regions are disordered, the first flanking residues are coloured cyan and orange, respectively.

1.2.3 Regulation of RagAC nucleotide binding state by amino acids

RagA is an inactive GTPase on its own and is constitutively loaded with GTP under nutrient rich conditions[76]. However, as a consequence of amino acid depletion, GATOR1 complex (Nprl2, Nprl3 and DEPDC5) can bind to RagA preventing mTORC1 access and acts as a GAP towards RagA promoting GTP hydrolysis and subsequent inhibition of mTORC1 activity through loss of Raptor binding[47, 77]. A lysosomal complex KICSTOR (KPTN, ITFG2, C12orf66 and SZT2) is required for this binding[78, 79]. Absence of SZT2 leads to constitutive activation of mTORC1 in nutrient depletion conditions resulting in neonatal lethality in mice[80].

GATOR1 activity is in turn mediated by its interaction with the GATOR2, and SAMTOR[43, 47]. While the latter is a direct S-adenosyl-methionine sensor and its binding to GATOR1 complex is regulated directly by the ligand, GATOR2 complex binds to GATOR1 only in absence of interactions with direct amino acid sensors CASTOR1 (Arginine) and Sestrin2 (Leucine)[43, 44, 81]. Amino acid binding to these proteins directly inhibits complex formation with GATOR2 which can then bind to GATOR1 relieving the mTORC1 inhibition caused by GATOR1-RagA binding.

RagC nucleotide binding state is regulated by folliculin (FLCN)-FNIP1/2 complex which acts as a GAP and promotes GTP hydrolysis in response to amino acids[82]. Folliculin is recruited to the lysosomes by RagA in a GDP-bound state during amino acid starvation[51, 83-85]. Binding of FLCN-FNIP complex obscures the binding site of Raptor on the RagAC dimer, prevents nucleotide exchange in RagA and GAP activity of Folliculin towards RagC[84, 85]. While the exact mechanism is yet unknown, it is evident that in nutrient rich conditions FLCN is capable of initiating GTP hydrolysis in RagC[50, 82]. GDP-bound RagC is then no longer capable of binding FLCN-FNIP and it dissociates[84]. This also allows nucleotide exchange to happen in RagA resulting in a locked active dimer conformation which can recruit Raptor and with it the entire mTORC1 to the lysosome[84]. What initiates this process and whether the dissociation on FLCN-FNIP or GTP hydrolysis happens first is unclear.

In addition to the regulation mechanisms described above, RagAC dimer regulates its own nucleotide loading state such that GTP can only be stably bound to one of the proteins[76]. Binding of the second GTP molecule triggers its hydrolysis while the pre-bound GTP stays intact[76]. This ensures that the dimer stays in the active or inactive conformation.

Structurally, it has been shown that depending on whether RagA or RagC are loaded with GTP, the relative position of the G-domains changes in relation to the roadblock domains. In the active $RagA^{GTP}$-$RagC^{GDP}$ conformation, the two G-domains come close together leaving ~19 Å gap between the bound nucleotides, whereas in the inactive $RagA^{GDP}$-$RagC^{GTP}$ conformation the distance increases to 30Å[69, 84, 86]. This facilitates selective access to the protein binding interface: Raptor peptide in the closed conformation and FLCN-FNIP in the open conformation[77, 84, 87].

1.2.4 $RagA^{GTP}$-$RagC^{GDP}$ binding to Raptor

From recent structural studies, it is known that helices α24, α26, and α29 within Raptor α solenoid interacts extensively with RagA switch I and inter-switch region through a network of hydrogen bonds and salt bridges[86]. Additionally, two short regions of amino acids 916-936, and 795-806 within intrinsically disordered regions of Raptor interact with the C-terminal domain of RagC and nucleotide made accessible by GTP hydrolysis which results in disordered switch I and switch II regions and a slight rotation of the G-domain away from the centre[69, 86].

1.2.5 Recruitment of RagAC to the Lysosome

Ragulator is a pentameric complex of LAMTOR1 and four “roadblock” domain proteins LAMTOR2-5 that is constitutively anchored to the lysosomal membrane via lipidation of LAMTOR1. In yeast this complex is known to be comprised only of 3 proteins (Ego1, Ego2, and Ego3) and is localised to the vacuolar membrane[88-90]. Though different by number of subunits, yeast EGO complex is structurally similar to the mammalian Ragulator complex. The N-myristoylated component (LAMTOR1 or Ego1) is wrapped around a

heterodimer (tetramer in mammalian cells) of “roadblock” domain proteins Ego2-3 or LAMTOR2-5 [71, 86, 89, 91, 92].

“Roadblock” domain architecture (αββαββββα) is not unique to Ragulator. It is present in many proteins across all kingdoms and is often associated with NTPase regulatory functions. Five β strands make up a beta-sheet that is flanked by one α-helix on one side and two helices on the other. This configuration usually requires a partner with similar organisation to complete the β-sheet to 10 strands and result in 2 interacting helices on one side and four on the other.

Recent structural studies have shown that to localise to the lysosome, RagAC binds to Ragulator with its “roadblock” dimerization domains[70-72, 91]. The interface of this interaction is analogous to LAMTOR2-3-LAMTOR4-5 interaction within the Ragulator complex itself. Similarly, LAMTOR1 binding is required to stabilise this interaction. Upon binding, residues 61-70 in the N-terminal intrinsically disordered region (IDR) of LAMTOR1 become ordered. This is 26 amino acids away from the structured core of the Ragulator complex. Remaining 60 residues at the N-terminus remain disordered.

1.2.6 At the lysosome: Ragulator, Rheb, SLC39A9, v-ATPase and TSC1/TSC2

At the lysosome, mTORC1 interacts with several proteins permanently localised to the lysosome: Ragulator complex, Rheb, SLC38A9, v-ATPase and TSC1/TSC2 complex.

Previous studies have shown that Ragulator and an amino acid transporter SLC38A9 may together act as a GEF to RagA. However, RagAC dimer does not undergo significant conformational changes upon Ragulator binding, and nucleotide binding sites of both RagA and RagC face away from the membrane and do not contact LAMTOR proteins. This makes conventional GEF activity unlikely and suggests that either this process is unique to this system and is not yet understood or there is an additional unknown player involved.[93].

SLC38A9 is known as a sodium-coupled amino-acid transporter that is required for efflux of essential amino acids produced in the lysosome through proteolysis into the cytoplasm[55]. Leucine produced and transported this way has been shown to be able to re-activate mTORC1 once in the cytoplasm[54]. In presence of arginine, the transporter was shown to interact with Ragulator-Rag complex and is therefore thought to be an arginine sensor for the system[56, 94]. The cytoplasmic tail of the protein is known to destabilise the inactive Ragulator-RagAC-FLCN/FNIP complex that prevents GTP binging to RagA and GTP hydrolysis by RagC. Once this interaction is destabilised, it triggers GAP activity of the FLCN/FNIP towards RagC, allowing for recruitment of mTORC1[95]

The process of mTORC1 recruitment to the lysosome is also influenced by lysosomal v-ATPase through its amino-acid sensitive interactions with the Ragulator complex[96]. In

addition, v-ATP participates in Ragulator-AXIN/LKB1/AMPK complex in response to starvation[97].

Once recruited to the lysosome, mTORC1 activation is initiated though complex formation with a small GTPase Rheb (Ras homologue enriched in brain). It can bind to mTORC1 in a GTP-sensitive manner and induce a conformational change that greatly improves the mTOR kinase activity and substrate recruitment [33, 37]. The nucleotide loading state of Rheb is at least in part mediated by TSC2, which acts as its GAP in absence of arginine[98-100]. Presence of arginine induces TSC2/TSC1 complex dissociation from the lysosomal membrane, making it unable to activate Rheb GTPase and promoting mTORC1 activation by GTP-bound Rheb.

Additionally, TSC1/TSC2 is an important point of integration of multiple signalling pathways regulated by mTOR and AMPK[101]. It connects mTORC1 and mTORC2 pathway as it undergoes inhibitory phosphorylation by mTORC2 substrate Akt (PKB) upon insulin stimulation. This prevents GTP hydrolysis by Rheb and promotes mTORC1 activation. It was also shown that TSC1/TSC2 complex enhance mTORC2 activation through an unknown mechanism independent of Rheb nucleotide state[102, 103]. To promote cell survival in glucose starvation conditions TSC1/TSC2 was shown to prevent excessive mTORC1 activation and therefore ATP depletion and cell death that follows, although the mechanism of this sensing and its relationship to AMPK pathway is not yet understood[104-106].

1.3 Glucose signalling

Extracellular glucose availability is sensed mainly through insulin-mediated activation of PI3K/AKT pathway that is activated in an mTORC2 dependent manner. While, as mentioned earlier, it is yet unclear how intracellular availability of glucose is communicated to AMPK and mTORC1 signalling pathways. Several proteins and systems are thought to be involved: lysosomal v-ATPase, together with RagGTPases and Ragulator, ULK1 and leucyl-t-RNA-synthetase, TBC1D7 and TSC1/TSC2, glycolysis enzymes Aldolase A, PFKFB3 and hexokinase 2. In glucose limiting conditions lysosomal v-ATPase was shown to interact with Ragulator, Axin and AMPK in fructose-1,6-bisphosphate dependent manner and disrupt mTORC1 translocation to the lysosome and subsequent activation[107]. ULK1 was shown to phosphorylate leucyl-tRNA synthetase and block its binding to leucine to inhibit mTORC1 activation[108]. Additionally, TSC1/TSC2 complex protein TBC1D7 was shown to be required for mTORC1 downregulation through an unknown mechanism [109]. Hexokinase 2 was in turn shown to bind to mTORC1 and downregulate its activity in absence of glucose, although the mechanism of this inhibition is also unknown[110]. In contrast, glycolysis enzymes PFKFB3 and PFK has been shown to promote mTORC1 translocation to the lysosome and subsequent activation in glucose rich conditions through interactions with Ragulator-RagB complex [111]. Recent research has also identified aldolase A product, dihydroxyacetone, as a possible glycolysis intermediate

that may signal glucose sufficiency to mTORC1 through GATOR1-GATOR2-RagGTPase axis[112]. However, the sensor protein for this molecule has not yet been identified. Therefore, despite the advances and improved understanding of the processes involved in glucose signalling to mTORC, the exact mechanism of glucose sensing itself remains elusive. This is due to the missing unambiguous links between the known nutrient signalling events directly preceding mTORC1 activation and the impact of glucose or other glycolysis intermediate binding to "sensor" proteins in terms of their conformation, localisation and relevant downstream protein-protein interactions.

1.4 Metabolic pathway crosstalk

While all nutrient signalling cascades are activated through distinct pathways and metabolites and affect different, and sometimes opposing cellular processes, they need to produce a coordinated and balanced response to the sum of all environmental factors. For example: anabolic processes such as protein and DNA production upregulated via mTORC1 branch do not only require sufficient substrate like amino acids and nucleotides but also happen to be the most energy intensive processes in the cell, therefore energy availability regulated by AMPK pathway must also be considered. This is ensured through several feedback loops and crosstalk mechanisms.

1.4.1 Energy (AMPK) vs nutrients and growth factors (mTORC1&2)

To reduce the ATP demand under energy deficient conditions AMPK downregulates mTORC1 activity directly: by phosphorylating Raptor on Ser792/722, and indirectly: by phosphorylating ULK1 on Ser317 and TSC2 at multiple sites between 1330-1350[110, 113-118]. Folliculin (FLCN) is also known to interact with both AMPK and mTORC1 pathway components with and without its partners FNIP1 and FNIP2. This dual activity places it in perfect position for the cross-talk between the pathways[119]. On one hand it is a GAP for RagC/D and thus contributes to mTORC1 activation, it interacts with AMPK via its partners FNIP1/FNIP2[120]. While this interaction was shown to increase phosphorylation of folliculin by AMPK and promote formation of the FLCN-FNIP complex, it was also implicated in induction of apoptosis and modulation of AMPK activity. While several other proteins related to regulation of both AMPK and mTORC1 we also shown to interact with FLCN-FNIP, more information is needed to understand this connection [119-121].

1.4.2 Interplay between mTORC1 and mTORC2

As with mTORC1-AMPK crosstalk, the mechanisms of feedback and mutual regulation of mTORC1 and mTORC2 complexes are poorly understood. While each complex has a unique set of effectors and targets, several proteins which either interact with, or influence and are influenced by both complexes have been identified. Some of these are functional kinases like: PKB, PI3K, and IRS1; others like TSC1/TSC2 complex, have another type of enzymatic activity as well as a number of protein interactors, some are transcription factors (FoxO), and some have no known enzymatic function but are nonetheless required for mTORC activation or inhibition (SIN1, PRAS40)[122-124]. It was

also recently shown that lysosome positioning can affect the rate of activation of both mTORC1 and mTORC2 complexes through an unknown mechanism, despite substantial differences in localisation patterns between these proteins[125]. Although some mTORC1 signalling related complexes Ragulator, RagAC and FLCN/FNIP were shown to be involved in lysosome positioning and mTORC1 recruitment, this mechanism does not seem to apply to mTORC2[119, 126, 127].

AMPK, mTORC1 and mTORC2 pathways are important therapeutic targets in cancer and other metabolic disorders. The relationships between these systems are complex and difficult to study. As a result, several early treatments based on their inhibition or activation have failed due to robust rewiring and support of the feedback loops. Therefore, it is vital that we understand the feedback mechanisms involved just as well as the main signalling streams and use this knowledge to improve effectiveness and safety of future therapies.

1.5 Protein-protein interactions (PPIs) in cell signalling

Due to the complex, dynamic and responsive nature of signalling networks, a high degree of control and adaptability of protein-protein interactions is required. As multi-component super-complexes typically combine functions of several proteins into one their activation can be controlled based on complex assembly. This is achieved through multiple overlapping mechanisms. Interaction partners can be separated by cellular compartments until the conditions are right. For mTORC1, this is mediated through its lysosomal translocation where, upon nutrient stimulation, mTORC1 is activated by Rheb. As described in section 1.2.5, this translocation also requires multiple protein-protein interactions where each is regulated by distinct upstream stimuli. To consolidate these inputs, several upstream regulators can bind to one downstream effector with overlapping binding sites. For example: GATOR1, FLCN/FNIP and Raptor use overlapping binding sites on RagAC to ensure that all relevant pre-conditions are met for mTORC1 activation [77, 84-86]. To improve the control and stability of this type of interactions, some protein-protein interfaces are compound. RagAC-Raptor interaction described in section 1.2.4 is a good example of this. There are three separate interfaces where one is specific to the “activated” state, and the rest contribute to the overall stability of the complex. The same mechanism is often employed by the kinases and their scaffold proteins to recruit the substrates[128-130]. An additional complication lies in the mode of regulation of these PPIs. While many of the interactions at the sensing end of the network depend on conformational changes at the interface induced by small molecule binding (nucleotides, amino acids, glucose etc.), kinase binding sites on the downstream effectors may be located on accessible disordered regions, and not necessarily adjacent to the modification site[128]. These disordered regions also often contain phospho-sites for other kinases, the modification state of which often influences the ability of other kinases to bind and/or phosphorylate nearby residues[131].

The transient nature, coupled with multi-level regulation and generally lower affinity of the kinase-substrate protein interactions, and in some cases also scaffold and regulatory interactions, poses difficulty in their identification and characterisation. However, several high- and low-throughput methods have been developed to reliably obtain information about interacting proteins and interaction interfaces within these networks. Each of the methods has its own advantages and limitations depending on the system in question as outlined below.

1.5.1 Approaches for PPI identification

Several high-throughput methods are available for non-targeted, proteome wide protein-protein interaction identification, The classic yeast two-hybrid screening (Y2H) can be used to study pair-wise interactions[132]. In this system two halves of a transcription factor are fused to a pair of genes of interest and expressed in yeast or, in a variation of the method in mammalian cells; when interacting proteins are expressed in the nucleus, the transcription factor would be reconstituted promoting expression of a reporter gene. Similar approaches have since been developed to extend the applicability of two-hybrid screens: a split-ubiquitin method for interactions with membrane tethered complexes, a fluorescent protein based two-hybrid approach – to facilitate small molecule PPI inhibition screening and response of the PPIs to stimuli through a fluorescence colour change, and a kinase-substrate based approach for identification of PPIs within native mammalian signalling networks[133-136]. Each of these approaches can be automated and scaled to create screening libraries and allow high throughput experiments. While well-established and widely used, this method has a few limitations: target proteins need to be localized in the nucleus for the traditional method, bulky fusion parts could block interactions or small molecule binding sites resulting in false negatives, the growth conditions must be favourable for the interaction to be detected, and false positives could be identified through non-specific interactions. Although this method can be “high throughput” by automation, still only two (in some cases three) partners can be analysed at a time. Additionally, due to the proximity required for fusion partner interactions, PPIs to and within larger macromolecular assemblies may be missed.

Affinity purification and co-immuno-precipitation (co-IP) are high-throughput techniques suitable for identification of components of more complex protein assemblies. Here, a ‘bait’ protein can be either natively (or over-) expressed in cells or spiked into a cell lysate[137-140]. Then, interacting proteins would be pulled down with a ‘bait’-specific antibody or method for bait capture. As a result, interacting proteins from the samples can be identified by tandem mass spectrometry. While scalable and useful for finding members of multi-protein complexes, this method also has its limitations. Due to rigorous washing procedures and reliance on lysis and elution buffers, the sensitivity of this method is limited to relatively stable complexes with higher affinity. Using co-IP, it is also inherently difficult to assess whether interactions are direct or indirect. Usually several

different ‘bait’ proteins and use of knock-out libraries are required for deciphering complex topologies.

As neither of these methods are particularly suitable for capture of weak or transient interactions, label transfer mass spectrometry methods like BioID have been developed[141-144]. A ‘bait’ protein is fused to a biotin ligase protein, which generates a reactive ATP-Biotin intermediate, which can then be transferred to primary amine groups within ~10 nm radius. The affinity to avidin variants is then used for stable enrichment of biotinylated ‘interactor’ proteins, which are again analysed by tandem MS. Despite its ability to identify potential stable and transient, as well as direct or indirect interactions, this method also has several limitations: 1) due to the diffusive nature of proximity labelling, the proteins tagged with biotin may also be proximal proteins which are not direct interactors will be labelled; 2) as with two-hybrid systems, the ‘bait’ protein is fused to a ~30 kDa protein which may interfere with interactions and solubility or localisation of the bait protein; 3) interacting partners with low abundance within the cell may be undetectable; 4) biotinylation relies on relative abundance of surface lysine residues near the binding interface, and may influence peptide detection or interfere with secondary interactions. Similar to the results obtained with co-IP and affinity purification methods, further low-throughput targeted screening is required to confirm bona fide interactions and untangle the interaction maps. For the latter, a complementary method could be implemented, whereby the biotin transferase is split between the ‘bait’ and the ‘prey’ proteins and can only be activated when brought together by ‘bait’- ‘prey’ interaction[145]. However, this still leaves the other limitations of the BioID method, and the results obtained would need to be validated through other means. Unfortunately, here, the problem of low affinity transient binding within signalling networks resurfaces, as the recovery of such complexes for any further studies would be low.

1.5.2 Methods for PPI characterisation

Once protein-protein interactors are identified and confirmed, their binding mode and interfaces need to be examined to understand the mechanism of action and, for example, to guide the design of therapeutics interfering with or enhancing select PPIs. Stable interacting and rigid multi-protein complexes are easiest to study because they can be purified. In these cases, conventional X-ray crystallography, as well as high-resolution single particle cryo-EM structure determination are vital to build atomistic models of interactors and binding interfaces. However, these methods typically require a good level of prior knowledge about the systems, in particular for recombinant expression of protein complexes and for interpretation of output data in terms of atomistic models. For successful structure determination, we often need to know all essential components for at least a minimal stable complex, the conditions under which those interact, and how to produce the desired state for analysis.

Unfortunately, these requirements mean that the transient interactions often found in signalling networks, including those steered by post-translational modifications pose major challenges for structural interrogation. Therefore, methods for intermediate-level characterization with reduced sample requirements like chemical cross-linking combined with tandem MS analysis and subsequent computational modelling, as well as hydrogen/deuterium exchange and hydroxyl radical footprinting MS were developed[146-151]. In addition, *in silico* methods are steadily improving to allow better predictions of interaction sites via residue co-evolution and machine-learning algorithms[152].

Hydrogen/deuterium exchange method exploits the ability of solvent exposed hydrogen atoms to freely exchange with deuterium in solution. The efficiency of this exchange is measured by high resolution mass spectrometric analysis of peptide digests. Peptides and residues with lower rate of deuterium exchange under conditions of complex formation as compared to the isolated protein are likely protected from the solvent or conformationally stabilized and implicated in a protein-protein interaction[153, 154]. While hydrogen/deuterium (H/D) exchange is a useful and sensitive technique for mapping transient interaction interfaces as well as stable ones, it requires an expensive specialist instrument set-up and is non-trivial to perform. A technique similar in principle, but which uses chemical oxidation of surface residues, for example with hydrogen peroxide or synchrotron radiation, for footprinting was described as an efficient alternative to hydrogen/deuterium exchange[150, 155, 156]. While providing the advantage of the ability of following protein interactions with and without stimulus in solution without major interferences to the binding and allowing for transient interface evaluation, both of these techniques have a significant drawback. The changes in H/D exchange or footprint may also indicate conformational changes and small molecule binding in addition to any complex interfaces[153]. Therefore, interpretation of the results may not always be clear without additional information.

Chemical cross-linking experiments, on the other hand, allow a great degree of flexibility in terms of experimental design, data collection and analysis. The reagents are typically designed to form covalent links between groups of reactive residues within proteins and protein complexes. Tandem MS analysis of peptide digests form crosslinked complexes are used to determine the cross-linked positions within the complex. Chemical cross-linkers are available with a variety of functional groups and, sometimes, enrichment handles and are also used to stabilise weak interactions for preparation of diluted samples, for example for cryo-EM analysis[157].

1.5.3 Cross-linking mass spectrometry for PPI study

Glutaraldehyde, a simple amine-reactive crosslinker is commonly used for EM sample preparations due to its strong reactivity and high efficiency of cross-link formation. It is, however, not suitable for MS-based applications that aim to identify transient partners or

cross-linked pairs of residues due to its propensity to form polymers[158-160]. For structural applications of crosslinking, reagents that are less stable and non-self-reacting in aqueous buffer are preferred to suppress cross-linking of non-interacting proteins or protein regions caused by random diffusional interactions[161]. Several different types of residue-specific functional groups for crosslinking are currently available: N-hydroxysuccynimide (NHS)-esters and carbonyl-diimidazole for lysines, maleimide and haloacetyl for cysteines, and carbodiimides for carboxyl-to-amine cross-link formation[162]. Photo-activatable residue-unspecific cross-linkers that can insert into any CH and NH bonds upon UV-stimulation via creation of a reactive radical: aryl azides, phenyl azides, benzophenones and diazirines. Bifunctional crosslinkers staple complexes together near the interaction site, provided that reactive residues are within the distance

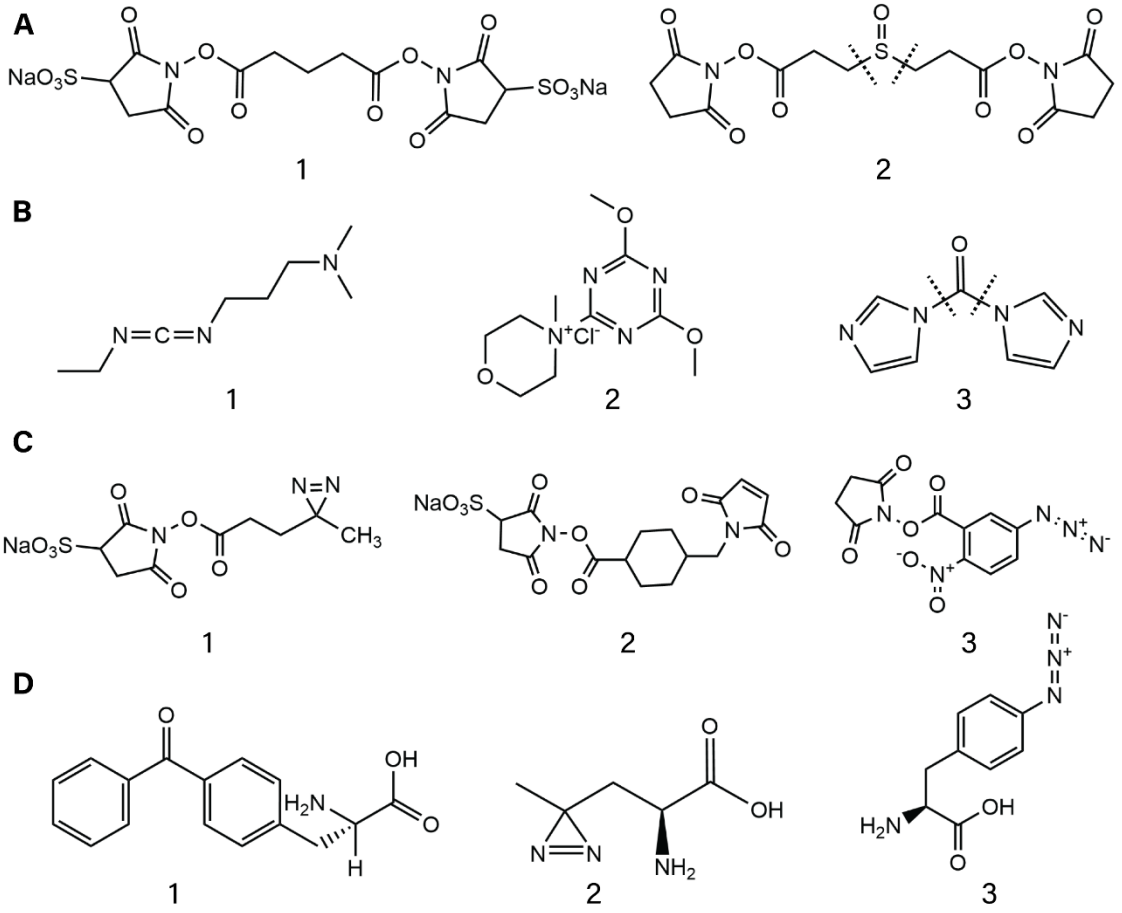

Figure 1.5.3.1 Cross-linker chemistry available for XL-MS. (A) Lysine-reactive N-hydroxysuccinimide linkers 1: Bis-Sulfosuccinimidyl-glutarate (BS2G), water soluble; 2: disuccinimidyl sulfoxide (DSSO), insoluble. (B) Zero-length cross-linkers 1: (EDC), carboxyl-amine, 2: (DMTMM), carboxyl-carboxyl, 3: (CDI), amine-amine/hydroxyl. (C) Heterobifunctional cross-linkers 1: Sulfo-SDA, Sulfo-NHS-diazirine, lysine-any active hydrogen, photo-activatable; 2: ANB-NOS, NHS-phenyl azide, lysine-any, photo-activatable; (D) Photoreactive unnatural amino acids 1: Benzoyl-phenylalanine (BPA), BPA-any active hydrogen, 350-365 nm excitation; 2: L-photo-Leucine (pLeu), diazirine, 350-365 nm excitation; 3: Azido-phenylalanine, 250nm excitation. Dashed lines represent MS-cleavable bonds.

accessible to the ends, which are typically separated by hydrocarbon or other non-reactive spacers of 4-20 Å in length. Examples shown in fig 1.5.3.1 (A-C)

With developments in mass spectrometers, cross-linkers and search algorithms, it is now possible to routinely identify pairs of specifically cross-linked residues from a variety of samples[161, 163-166]. Distance restraints can then be estimated based on the spacer length (and any allowances for flexibility) and be employed for modelling of protein assemblies or for de novo structure prediction when high-resolution data is not available[161] .

1.5.4 Considerations for crosslinking studies of PPIs in signalling

NHS-esters with varied linker chains are the most widely used reagents for cross-linking mass-spectrometry applications[167]. They form specific covalent bonds with primary amine (NH_2) groups of lysine residues. In contrast to glutaraldehyde, the NHS-based cross-linking reaction competes with concurrent hydrolysis of the active compound and once-reacted intermediate[162]. This way, cross-polymerisation is minimised, and the effective distance between cross-linked residues should not be more than the length of the crosslinker. As lysines are relatively common on protein surfaces, NHS-ester-based crosslinking is a robust method suitable for a wide range of targets. Due to ease of use and specificity of the NHS-ester compounds, similar chemistry is also used in mass-spectrometry based absolute and relative protein quantification (such as tandem-mass-tagging method) and protein immobilisation or fluorescent labelling[162, 168]. Despite dozens of successful structural studies of protein complexes that rely on data obtained with NHS-ester cross-linkers (DSS, BS3, DSSO), limitations still exist in terms of applicability to the often low-abundant, highly flexible proteins and transient interactions frequently found in signalling networks[146, 169-171].

Lysines are common at the protein surface and, in principle, have a high likelihood of being found near a protein-protein binding interfaces to generate multiple cross-links required for successful modelling. However, comparison to experimentally determined structures shows that lysine-based crosslinking is not truly specific for short distances between crosslinked side chains. First, lysines are highly flexible resulting in allowable distance restrains between Cα of up to 26-30 Å[172]. Second, for larger proteins and assemblies, like those involved in cell signalling, intra- and inter-domain dynamics also have a substantial impact on cross-linking efficiency. In consequence, cross-links obtained with DSS or BS3 often bridge distances of up to 60 Å when compared to experimental structures. Even when many cross-links are identified, such large ranges and long variable distance cut-offs can result in poor convergence during modelling and limit the use of this method for the analysis of highly dynamic PPIs.

While there are integrative modelling methods that allow to include extra restraints and explicitly account for side-chain flexibility and even large-amplitude intra-protein motions by using protein ensembles, flexibility features need to be specified in advance[173, 174].

For many complexes consisting of multiple multi-domain proteins with large regulatory unstructured regions the required information about possible conformational ensembles in solution is, however, not available. Therefore, other strategies for efficient and precise analysis of dynamic PPIs are needed. On one hand, improvements in cryo-EM technology can help to provide insights into large-scale motions for generation of smarter restraints with cross-linking. On the other hand, novel strategies of cross-linking combined with mass spectrometry hold promise for providing more meaningful inter-residue restraints. First: zero-length cross-linking with different reaction chemistries, and second: metabolic incorporation of cross-linkable amino acids into proteins.

1.5.5 Zero-length cross-linking

As the name suggests, zero-length cross-linking reagents modify target residues in a way that promotes covalent bond formation with the nearest reactive residue without leaving any part of the cross-linker behind. Additionally, cross-linkers, which have a spacer region less than 3 Å (one carbonyl group) are also often referred to as “zero-length”. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium (DMTMM) are true zero-length cross-linkers that have been developed to promote amide bond formation between the carboxylic side chains of aspartate or glutamate, and lysine at moderately low and neutral pH (pH 4.7-6.0 for EDC and 7.0-7.5 for DMTMM). DMTMM can also be used to cross-link carboxylic residues to each other when used in combination with dihydrazides[175]. Use of 1,1-carbonyldiimidazole (CDI) to cross-link lysines to lysines, or to residues with hydroxyl groups (Ser, Thr, and Tyr) with a spacer length of 2.6 Å has also been described[162, 176]. While EDC and DMTMM are true zero length cross-linkers, the advantage of the CDI’s carbonyl linker is that it can be cleaved in the mass-spectrometer for the formation of reporter ions that can be used for faster, more sensitive and more reliable cross-link identifications (fig. 1.5.5.1) [176]. The additional benefit of these zero-length cross-linkers is that the intermediate products are unstable in aqueous solution and thus rather result in turned-over reagent than in unspecific or long-range cross-links[162].

Despite these advantages, there are also substantial caveats to using zero-length cross-linkers for characterisation of protein complexes. First, due to the short cross-linker length and rapid decomposition of reagents, these are likely to produce fewer identifiable cross-links and may not provide sufficient data for modelling on their own. Second, it has been reported that for test proteins DMTMM provided only marginally shorter distance restraints than the conventional lysine-lysine cross-linker DSS, further highlighting the impact of intra-protein dynamics on cross-linking results[175]. Despite these limitations, with different reaction chemistry and residue specificity compared to DSS and BS3, these compounds have the potential to provide complementary information to guide model development for dynamic PPIs, but their use is just being established for simple model protein systems[175-180].

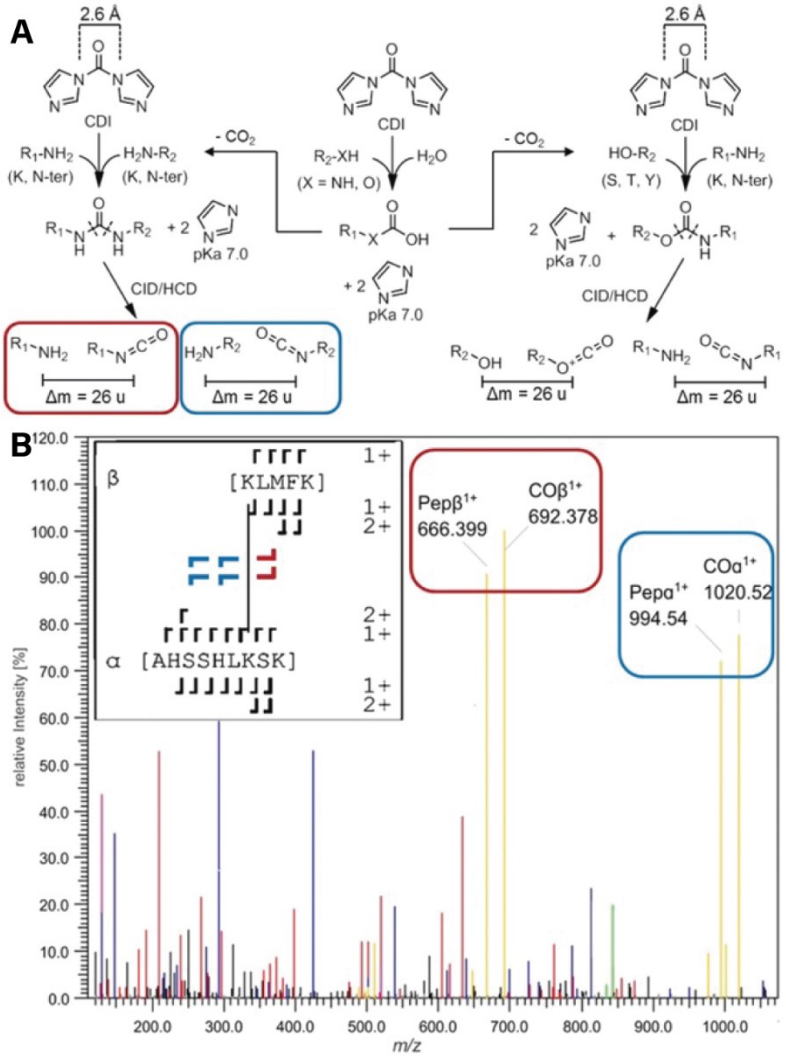

Figure 1.5.5.1 (A) Carbonyldiimidazole crosslinking reactions (B) Sample ion spectra of identified crosslink, with reporter fragments highlighted in yellow. This figure is reproduced from the original publication and is reproduced with permission of John Wiley and Sons publishing [citation in text].

1.5.6 Photoactivatable amino-acid analogues for cross-linking

Chemical cross-linkers (zero-length or otherwise) require to be added to protein solutions for both *in vivo* and *in vitro* experiments. This means that for *in vitro* studies the complexes or their components need to be extracted form cells, or the cross-linker must be membrane permeable for *in cell* studies[181]. As mentioned above, *in vitro* cross-linking provides little benefit for interface characterisation of unstable or poorly understood protein complexes as it requires purification of a stable complex or precise knowledge of complex formation requirements.

It has recently been show that *in cell* cross-linking with membrane permeable reagents like DSS or DSSO (disuccinimidyl sulfoxide) is dependent on protein concentration in the

cell, where the most abundant species act as cross-linker sponges obscuring the formation and detection of desired cross-links[171]. While under conditions tested this effect could be overcome by raising the cross-linker:protein ratio, this trend does not bode well for dynamic or transient interactions common in signalling networks. Since the complex formation in signalling networks is typically dependant on diverse external stimuli and multiple post-translational modifications, an inducible, native system to target complexes of interest may be preferred to chemical-based approaches.

Such opportunity is provided by site-specific translational incorporation of unnatural amino acids. Proteins of interest are usually engineered to include (rare) stop codons, which are reassigned to unnatural amino acids by co-expression of specialised adapted tRNAs and tRNA synthetases[182, 183]. Dozens of unnatural amino acids with bio-orthogonal chemistries have been designed to include fluorescent groups or cross-linker handles. Notably, para-benzoyl-phenylalanine (BPA) and para-azido-phenylalanine (APA) (fig. 1.5.3.1 D) incorporated this way are used to permanently cross-link small molecule-protein and protein-protein interaction partners in a position-specific and UV-activated manner[184, 185]. Since incorporation of these amino acids is position specific and requires gene mutagenesis and expression of specific tRNA and tRNA synthetase, prior knowledge about the system is required to reap the benefits of site-specific unnatural amino acid incorporation for cross-linking. Particularly, knowledge of the binding site on at least one of the interaction partners is required for the experimental design for further interface or binding partner identification. For signalling pathway research this could, for example, be useful for studying overlapping binding sites with a one-to-many relationship under different stimuli. While identification of cross-linked peptides obtained with these non-natural amino acids is complicated by their non-discriminatory reaction chemistry where the cross-links can form to any side chain through interaction with any C-H bond, several studies were successful in unambiguous cross-linked peptide identification[186-189].

For poorly characterised protein complexes, an alternative approach is provided by the non-site-specific incorporation of photoactivatable diazirine analogues of L-Leucine, L-Methionine, and L-Lysine exploiting native tRNA promiscuity[190, 191]. Such a mode of unnatural amino acid incorporation does not require modified DNA or tRNA constructs. In contrast to genetic unnatural amino acid incorporation, which typically produces one specific substitution per protein molecule, the diazirine analogues are randomly incorporated at a variable rate (up to 6% for Photo-Leucine, 20% for Photo-Methionine and 4-40% for Photo-Lysine were reported) instead of their natural variants[190, 191]. Due to random native incorporation of these residues, a library of photo-amino acid incorporation positions would be created automatically during every experiment. As L-Leucines and L-Methionines, together with other hydrophobic residues, are frequently found at protein-protein interfaces, this could allow for efficient space sampling[191]. Similar to BPA and APA, upon irradiation of the photo-amino acids with ~360 nm UV-light

the diazirine group forms a reactive carbene (while N_2 dissociates). This enables insertion into the nearest C-H or N-H bonds or addition to any unsaturated groups to form cross-links to any proximal residue. To date, both: photo-Leucine and photo-Methionine have been used to identify novel interactions through affinity-purification method coupled with tandem MS[192-196]. Preliminary MS studies on cross-linked peptide levels have shown that the cross-linked peptides remained intact during tandem MS and could, in principle, be identified for further structural studies.[197, 198]

This makes photoactivatable amino acid incorporation an attractive potential method for structural characterisation of regulated transient interactions. Because photo-amino acids would be distributed across the protein at a given rate (unlike BPA or APA), only a fraction of native amino-acids at each position would be replaced by an analogue. To compensate for this and allow for MS/MS-based cross-link identification, efficient incorporation, protein production and enrichment of both cross-linked proteins and peptide would be paramount. However, there is currently limited information on how photo-amino acid incorporation is distributed across protein sequences and how it affects protein expression, stability, and protein-protein interactions.

2 Aims of this thesis

The starting point for this thesis was to identify and structurally characterize protein crosstalk in nutrient-mediated metabolic regulation, with particular emphasis on amino acid sensing via Rag GTPases and mTORC1 as well as glucose-mediated metabolic control. At the outset of this project, I decided to follow a dual approach to study relevant interactions by established methods for high-resolution structural analysis, such as X-ray crystallography and Cryo-EM, but also to tackle the challenge of developing zero-length crosslinking as a method for analysing transient PPIs in metabolic signalling.

In chapter 3.1 I aim to characterise the impact of a naturally occurring mutation in the glycolytic enzyme hexokinase 2 linked to the development of insulin resistance on overall structure and catalytic throughput by X-ray crystallography in combination with biophysical analysis.

In chapter 3.2 I aim to reconstitute *in vitro* the interaction of RagAC with Raptor and mTORC1 for cryo-EM analysis to understand the mechanism of amino-acid mediated mTORC1 activation.

In chapter 3.3, I use the RagAC system as a model to develop and evaluate current zero-length cross-linking mass spectrometry methods for applications to transient and dynamic multiprotein complexes in cell signalling networks.

3 Materials and methods

3.1 Materials

All chemicals used in the study are analytical grade, unless otherwise mentioned.

3.1.1 Strains and media

Cell line	Description	Growth media
HEK 293	Human embryonic kidney cell line	DMEM +10% FBS
HEK 293T	HEK 293 + SV40 Large-T-antigen	DMEM +10% FBS
T-REx	HEK 293 stably expressing tetracycline repressor protein 1	DMEM +5-10% FBS (+5µg/mL blasticidin)
SF21	Spodoptera frugiperda ovarian cell line	HyClone SFX media
E. coli DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(r_{k^-}, m_{k^+}) phoA supE44 thi-1 gyrA96 relA1 λ-	LB, TB
E. coli NEB10β	F– mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara-leu)7697 galU galK λ– rpsL (Str^R) nupG	LB, TB
E. coli pir1	F- Δlac169 rpoS (Am) robA1 creC510 hsdR514 endA recA1 uidA (ΔMluI)::pir-116	LB, TB
E. coli NEB Turbo	F' $proA^+B^+$ $lacI^q$ ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)Tet^S endA1 thi-1 Δ(hsdS-mcrB)5	LB, TB
E. coli EMBacY	MultiBac (Geneva-biotech.com) F-replicon Kan^R LacZ (mini-attTn7) ΔchiA ΔV-cath Chl^R YFP	LB, TB

3.1.2 Media and additives:

Additive	Use	Supplier
E. coli		
Polyethylene glycol MW 3350 (PEG 3350)	TSS competent cells	Aldrich
$MgSO_4$	TSS competent cells	AppliChem
Ampicillin	Plasmid amplification	Carl Roth
Gentamycin	Plasmid amplification	Carl Roth
Kanamycin	Plasmid amplification	Carl Roth
Spectinomycin	Plasmid amplification	Calbiochem
Tetracycline	Plasmid amplification	Carl Roth
Luria-Bertani media (LB)	Culture	LLG Labware
Terrific Broth (TB)	Culture	LLG Labware
IPTG	Culture/ Protein expression	Carl Roth

5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	Blue-white selection for Bacmid generation	Carl Roth
GenElute Miniprep	Mini Prep DNA Purification	Sigma-Aldrich
QIAprep Spin Miniprep	Mini Prep DNA purification	Sigma-Aldrich
ZymoPURE II Maxiprep	Maxi Prep DNA Purification	Zymo Research
Wizard SV Gel and PCR	PCR-Cleanup System	Promega
S. frugiperda		
FuGENE	Transfection	Promega
HyClone SFX	Culture maintenance	GE Healthcare
DMSO Cell culture grade	Cell freezing	AppliChem
H. sapiens		
Dulbecco's modified Eagle medium (DMEM)	Cell line maintenance	Sigma-Aldrich
Dulbecco's modified Eagle medium w/o L-Leucine and L-Methionine (DMEM-LM)	Photo-leucine and Photo-methionine incorporation	Thermo Fischer Scientific
Foetal Bovine Serum (FBS)	Cell line maintenance	Gibco
Trypsin-EDTA	Cell line maintenance	Sigma-Aldrich
Soya bean trypsin inhibitor	Cell harvesting	Sigma-Aldrich
Ca2+ free Phosphate buffered saline (PBS)	Cell line maintenance	Sigma-Aldrich
Dialysed foetal bovine serum (DFBS)	Photo-leucine and Photo-methionine incorporation	Gibco
L-Leucine	Photo-leucine and Photo-methionine incorporation	Alpha Aesar
L-Methionine	Photo-leucine and Photo-methionine incorporation	Alpha Aesar
L-Photo-leucine	Photo-leucine and Photo-methionine incorporation	Iris Biotech GmbH
L-Photo-Methionine	Photo-leucine and Photo-methionine incorporation	Iris Biotech GmbH
Blasticidin	T-REx cell line maintenance	Invitrogen
FuGENE	Transfection	Promega
PEI (Polyethyleneimine)	Transfection	Sigma Aldrich/ Polysciences
Valproic acid, sodium salt	Protein expression, growth inhibitor	Sigma Aldrich

3.1.3 Protein purification and cross-linking buffers and additives

All generic materials not mentioned in the table were of analytical grade, sourced from AppliChem, or Sigma Aldrich.

Bicine and HEPES buffers	Protein Purification and storage	AppliChem

Nucleotides (GxP, ATP)	Ligand additive	Jena Biosciences/Sigma Aldrich
PEG and PEG MME	Crystallisation	Sigma Aldrich
Glucose-6-phosphate	Ligand additive	Sigma Aldrich
Phosphoramidon	Protease inhibitor	PeptaNova
Bestatin	Protease inhibitor	PeptaNova
Pepstatin A	Protease inhibitor	PeptaNova
Phenanthroline	Protease inhibitor	PeptaNova
E-64	Protease inhibitor	PeptaNova
PMSF	Protease inhibitor	Sigma Aldrich
Glutaraldehyde	Crosslinking	Electron Microscopy Sciences
BS3	Crosslinking	Thermo Fisher Scientific
DSS	Crosslinking	Thermo Fisher Scientific
CDI	Crosslinking	Sigma Aldrich
DMTMM	Crosslinking	Aldrich
ADH	Crosslinking	Alpha Aesar
EDC	Crosslinking	Sigma Aldrich
Anhydrous DMSO	Crosslinking	Sigma Aldrich
Acetonitrie, 0.1% TFA (LC/MS grade)	Peptide SEC fractionation	Fisher Scientific
Water with 0.1% TFA (LC/MS grade	Peptide SEC fractionation	Fisher Scientific
Ni-Magnetic beads	Pull-down	Cube Biotech
Superdex S200 Increase resin	SEC	GE Healthcare
Superdex S30 Increase	Peptide SEC	GE Healthcare
His-Tag Novagen® Monoclonal Ab 70796-M	Western Blot	Sigma Aldrich
Rabbit polyclonal Anti-mouse Ab – HRP conjugate	Western Blot	Abcam
Fat-free dry milk	Western Blot blocking	Coop
Clarity ECL substrate	Western Blot detection	Bio-Rad
Ni Sepharose Fast-Flow	IMAC	GE Healthcare
Ni Sepharose HP	IMAC	GE Healthcare
IGEPAL/NP-40	Cell lysis	Thermo Fischer Scientific / Fluka
TCEP	Protein buffer additive	Sigma Aldrich
Amicon 4/15 10,000 MWCO	Protein Concentration	Merck Millipore
Slide-A-Lyzer 3,500 MWCO	Small scale dialysis	Thermo Fisher Scientific
Ultrafree® filters	Small scale buffer filtration	Merck Millipore

3.1.4 Electrocompetent *E. coli* cell preparation and transformation

1-5uL of DNA was added to 50 µL of thawed competent cells, mixed and transferred to ice cold 0.1 cm gap electroporation cuvettes. If DNA volumes above 1 µL were used, the DNA was dialysed against Milli-Q water by placing a droplet onto a 0,05 µm pore size Millipore dialysis filter. Electroporation was done using the Bio-Rad MicroPulser electroporator on a setting for bacteria. 450-950 µL of fresh room temperature LB was added immediately. Cells were then transferred to a 1.5 mL Eppendorf tube and incubated in a heat block at 37 ˚C, shaking at 800 rpm for 1-4 hours prior to plating onto the agar plates depending on the antibiotic used for selection.

3.1.5 TSS competent *E. coli* cell preparation and transformation

To prepare the competent cells, an aliquot of the DH5α or NEB10β cells was used to grow an overnight culture in LB without antibiotic. Next day fresh media was inoculated with 1:300 v/v of the overnight culture. The cells were grown until OD_{600} was between 0.3 and 0.4. When the OD was reached, the cells were spun down at 3000xg at 4˚C for 10 minutes and resuspended in $1/10^{th}$ or $1/20^{th}$ of the original culture volume of ice cold TSS solution. TSS solution was made using 10% PEG3350, 5% DMSO, and 20 mM $MgSO_4$ according to the protocol described by Chung et al. [199]. The solution was sterile filtered and chilled on ice until use. Cells were either used immediately or flash frozen in liquid N_2. KCM mix contains 0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$ in Milli-Q water. Sterile filtered prior to use.

Transformation of TSS competent cells was done using the following protocol. An appropriate amount of purified DNA or a cloning reaction mix was added to 20 µL of ice cold KCM mixture and diluted with autoclaved MiliQ water to 100 µL. 100 µL of competent cells were then added to this mix and incubated on ice for 10 minutes. 400-800 µL of fresh room temperature LB was added for outgrowth. Cells were then incubated in a heat block at 37 ˚C, shaking at 800 rpm for 1-4 hours prior to plating onto the agar plates depending on the antibiotic used for selection.

3.2 Expression construct generation

The following gene constructs were used in this work:

Name	Proteins	NTerm-tag	Vector	Expression system	Generated by
	$RagA^{WT}$	10His	pAB2G	Not used for expression	E. Sauer
	$RagA^{Q66H}$	10His	pAB2G	Not used for expression	E. Sauer
	$RagC^{WT}$	-	pIDS	Not used for expression	E. Sauer
	$RagC^{S75N}$	-	pIDS	Not used for expression	E. Sauer

Cre21	$RagA^{WT}$-$RagC^{WT}$	10His-RagA	pAB2G-pIDS	SF21	E. Sauer. /E. Stuttfeld
Cre22	$RagA^{G65A}$-$RagC^{S75N}$	10His- RagA	pAB2G-pIDS	SF21	E. Sauer. /E. Stuttfeld
Cre23	$RagA^{Q66H}$-$RagC^{G119V}$	10His-RagA	pAB2G-pIDS	SF21	E. Sauer. / E. Stuttfeld
Cre24	$RagA^{Q66H}$-$RagC^{WT}$	10His-RagA	pAB2G-pIDS	SF21	This work
Cre25	$RagA^{Q66H}$-$RagC^{WT}$	10His-RagA	pAB2G-pIDS	SF21	This work
Cre26	$RagA^{Q66H}$-$RagC^{S75N}$	10His-RagA	pAB2G-pIDS	SF21	This work
A43	Raptor	10His-Myc-Flag	pAB2G	SF21	S. Imseng
pFrA	eGFP	10His	NA	HEK293/HEK293T/T-Rex cells	S. Singh
MX-01	C-term P2A-eGFP	NA	NA		S. Singh
GA-HEC-01	$RagA^{WT}$-$RagC^{S75N}$	10His-RagA	MX01	HEK 293/HEK 293T/T-REx	This work
pND007	$HK2^{WT}$	6His (thromobin)	pET28a-LIC	E. coli BL 21	Addgene 25529
pND008	$HK2^{\Delta 17\ R42H}$	6His (thromobin)	pET28a-LIC	E. coli BL 21	N. Dietz

3.2.1 Hybrid RagAC construct generation for insect cell expression

To produce expression clones of RagAC combinations, the appropriate variants of RagA were combined with each of the RagC variants in Cre recombinase reaction as described in Fitzgerald et al.[200]

Identity of the genes was verified through DNA sequencing. Additionally, the mutations in the purified products were confirmed by MSMS.

3.2.2 $RagA^{WT}$-$RagC^{S75N}$ generation for mammalian cell expression

To generate the construct for expression in HEK293 cells, a GATEWAY compatible construct as well as a mammalian Gateway acceptor vector were used. Outside of the attB1 and attB2 sites, the MX-01 vector contains an *E. coli* origin of replication, a CMV promoter preceding the attB1 site, EGFP open reading frame (orf) preceded by a P2A and T2A site C-terminal to the attB2 site as well as a Gentamycin resistance gene. This plasmid was obtained from Shubham Singh.

The Gateway compatible construct was designed to contain the $RagA^{WT}$ and a $RagC^{S75N}$ open reading frames separated by P2A sites with an N-terminal 10His-tag flanked by attB1 and attB2 sites. Additionally, Kozak consensus sequence (GCCACC) was added between the N-terminal Methionine and the attB1 site. Gibson Assembly (GA) protocol was used

to obtain this construct from already available vectors. Using the insect cell expression vector as a template, the attB1 site between the 10His-tag and RagA orf was deleted by PCR and blunt ends of the product were ligated with T4 Ligase. Similarly, the AttB1 site was removed from the original entry clone of the $RagC^{S75N}$ to allow for successful amplification of the constructs for GA. These vectors were then used for production of the GA fragments by PCR including the sequence of attB1 and P2A site on the RagA fragment and a P2A site and attB2 site on the RagC. pFrA (from Shubham Singh) was used as the backbone for the assembly of the Gateway entry clone. The primers were designed with overlaps of up to 60 nucleotides and the PCR was performed using in-house purified Phusion polymerase in GC buffer (NEB) in presence of dNTPs (Carl Roth) without additional additives. The PCR products were gel-purified using a Promega kit and treated with DpnI (NEB) prior to assembly using the NEB HiFi Assembly kit according to manufacturer's instructions. The final DNA was transformed into the electrocompetent *E. coli* NEB10β and plated on agar with 100 µg/mL Ampicillin. Several of the resulting colonies were sequenced at Microsynth. Correct clones were used for Gateway assembly protocol with MX-01 vector according to manufacturer's instructions.

Positive colonies were sequenced by Microsynth and the DNA was amplified in *E. coli* NEB10β and prepared for mammalian transfection using the Zymo Research Maxi prep kit according to manufacturer's instructions including the endotoxin removal step.

3.3 RagAC and Raptor expression and purification

3.3.1 Insect cell baculovirus generation

For Bacmid generation, constructs of RagAC obtained through Cre recombination, as well as a Raptor expression clone A43 (see plasmid table) were transformed into electrocompetent *E. coli* EMBacY cells and plated onto LB agar plates containing 10 µg/mL Gentamycin, Kanamycin, Tetracyclin, Spectinomycin, X-Gal and IPTG after an overnight recovery in LB medium at 37C. After 30h, two-three white colonies per construct were picked and grown overnight in LB containing Gentamycin and Kanamycin. Cells were centrifuged at 1000xg for 10 minutes and the pellet was lysed using QIAGEN DNA miniprep kit solutions 1-3. Cell lysate was cleared by centrifugation at >15000xg in a tabletop centrifuge for 25 minutes. Supernatant was collected and the DNA was precipitated by adding 700 µL of cold 100% isopropyl alcohol and incubating for 2 hours at -20 ˚C. The DNA was then centrifuged at >15000xg for 5 minutes and the pellet was washed twice with 70% ethanol. It was then dried in a laminar flow hood and resuspended in 100 µL of autoclaved Milli-Q water.

Transfection and subsequent virus generation in insect cells was carried out according to the protocol described in Fitzgerald et al. [200] For each of the constructs 0.5*10^6 of *S. frugiperda* Sf21 cells per well of a 6-well tissue culture dish were transfected with 10 µL DNA and 5 µL FuGENE transfection reagent according to manufacturer's instructions. For V_1 generation, 3 mL of each V0 were used to infect 50 mL (25*10^6 cells). V_1 was

harvested after 72-96 hours depending on cell viability. For V_2 generation, 5 mL of each V_1 was added to 100*10^6 cells in 200 mL. V_2 was harvested when cell viability was <80% (~72 hours post infection). YFP fluorescence was observed throughout.

If required, to make baculovirus infected cells (BIIC), 1 mL of V_2 was used to infect 100*10^6 cells at ~10^6 cells/mL. Cells were harvested after 24h post infection and frozen in 45% spent medium, 45% fresh medium and 10% DMSO.

3.3.2 Immobilised metal affinity purification (IMAC)

All RagAC variants and Raptor were expressed in insect cells by infecting 5-10L of culture at ~1*10*6 cells/mL with V2 at 1:300 dilution. Cell pellets were harvested by centrifugation at 1000xg for 15 minutes in Beckman-Coulter Lynx Sorvall centrifuge with an F9-6-1000 LEX rotor, weighed and stored at -80 ˚C until purification.

Initially, for RagAC constructs, GenScript High-affinity Ni-charged resin was stripped using 100 mM EDTA solution and loaded with $CoCl_2$ to improve the purity of the eluate. Initial batch of $RagA^{WT}$-$RagC^{WT}$ and $RagA^{G65A}$-$RagC^{S75N}$ constructs were purified this way. However, during purification of other variants it transpired that GE healthcare Ni-Sepharose FastFlow or HP columns provided better yield and similar purity of the samples. Second batch of $RagA^{WT}$-$RagC^{WT}$ and both: $RagA^{WT}$-Rag^{C75N} and $RagA^{Q66H}$-$RagC^{WT}$ were purified using these columns. Raptor was also purified using a 5mL Ni-Sepharose FastFlow column.

Cell pellets were thawed and resuspended in lysis buffer (50 mM Bicine pH 8.0 150 mM NaCl, 10 mM Imidazole, 2 mM $MgCl_2$, DNaseI (solid powder ~2-4 mg) + protease inhibitors: Bestatin, PMSF, Phenanthroline, E64). Lysis was done by sonication (at least 2x 5 minutes on 50% duty cycle until clear and less viscous). Lysate was cleared by centrifugation at 35,000xg for 45 minutes using Beckman-Coulter Optima X-90 ultracentrifuge with Ti 45 rotor.

IMAC columns were washed with 5 column volumes (CV) of water, and equilibrated with 5 CV of the Lysis buffer using a peristaltic pump at 4 ˚C. Cleared lysate was loaded onto the columns at ~1 mL/min flow rate once or, in some cases, left cycling overnight at 4˚ C. Subsequently, the columns were washed with 5 CV lysis buffer and attached onto the AKTA Purifier system or Bio-Rad NGC. The columns were then washed for up to further 5CV with 5% elution buffer (EB – 50 mM Bicine pH 8.0, 100 mM NaCl, 500 mM Imidazole) to make approximately 30 mM imidazole concentration and eluted over a 10 CV gradient to 100% (EB) while collecting 1 mL fractions and monitoring A_{280}. Elution peak fractions were subjected to SDS-PAGE analysis. Fractions containing all required proteins with minimum of contaminants were pooled and ~1mg TEV protease was added to cleave the tag. The mixture was dialysed against lysis buffer without imidazole overnight.

Orthologous Ni-purification was then performed to remove the remaining contaminants and majority of the uncleaved protein using the procedure described above. The flow-

through and first few ml of the wash which now contained the proteins of interest were collected, and a step elution with 100% EB was done to clean the column before regeneration. For purification of Raptor, the protein was eluted with a gradient 0-40% EB due to ability of cleaved Raptor to bind to the Ni^{2+} resin. Resulting protein solutions were concentrated using conical concentrators with 10,000 kDa MW cut off for RagAC and 30,000 kDa for Raptor. Final protein concentration was determined by measuring absorbance at 280 nm and calculation using the Beer-Lambert law. Small aliquots were flash frozen in liquid N^2 and stored at -80 ˚C.

3.4 RagAC Nucleotide exchange and analysis

3.4.1 Nucleotide exchange

The exchange of native nucleotides in the RagAC dimer was performed in accordance with previously published procedures. The protein solution was incubated in presence of 10 mM EDTA and the nucleotide mixture at 1 mM for 30 minutes at room temperature. The reaction was stopped by addition of $MgCl_2$ to a final concentration of 12 mM.

3.4.2 Strong Anion Exchange (SAX) chromatography

To analyse the nucleotide content of RagAC dimer, the protein solution was heated at 95 °C for 5 minutes and subsequently centrifuged at >15,000xg for 10 minutes. The supernatant was loaded onto the strong anion exchange column previously equilibrated with 20 mM bicine buffer. Elution was carried out with a gradient of 0-0.5 M NaCl over 15 column volumes. Absorbance at 260 and 280 nm was monitored throughout. Stock solutions of commercial GTP and GDP nucleotides were used as a reference.

3.5 Ragulator-RagAC-Raptor size exclusion chromatography

Treated and untreated RagAC mutants were combined with Raptor and/or Ragulator proteins LAMTOR1-5 at equimolar ratios. The samples were incubated for 30 minutes at room temperature prior to first injection and kept at 4C in-between. Following the initial incubation, 45 µL were applied to a self-packed Superdex S200 Increase 4.2/250 PEEK column pre-equilibrated with appropriate running buffer at 0.25 mL/min. For samples without additional nucleotides the following buffer was used: 20 mM Bicine pH8.0, 150 mM NaCl, 2 mM TCEP. For samples with nucleotides, column was re-equilibrated in the same running buffer with addition of 20 µM GTP and 5 µM GDP.

3.6 Hexokinase 2 R42H and WT purification

Hexokinase 2 R42H was purified from *E. coli* by N. Dietz as follows. TSS competent BL21 were transformed with pND007 (HK2 d16 WT), and ND008 (HK2 d16 R42H). Cultures were grown from single colonies at 37 ˚C and were induced with 1mM IPTG when the OD was 0.6-0.8. The cultures were further grown for 4 hours and cells were harvested by centrifugation at 1000 xg for 15 minutes using Beckman-Coulter Lynx Sorvall centrifuge with an F9-6-1000 LEX rotor.

8g of pellets were resuspended in 30 mL of lysis buffer (20 mM Hepes pH 8.0, 150 mM NaCl, 1 mM Mg Cl_2 and 5 mM β-Mercaptoethanol) and lysed using Thermo Spectronic French press at 1000 psi. The lysate was cleared by centrifugation at 35,000 xg for 30 minutes using Beckman-Coulter Optima XE-90 Ultracentrifuge with Ti70 rotor. Cleared lysate was then filtered through 0.45 µm syringe filter and applied to pre-equilibrated in lysis buffer 1 mL HisTrap HP columns (GE Healthcare). The columns were washed with 20 CV of lysis buffer, and proteins were eluted using 3-step elution of 20, 60 and 100% buffer B (20 mM Hepes pH 8.0, 150 mM NaCl, 500 mM Imidazole, 5 mM β-Mercaptoethanol) for 8 CV each. Absorbance at 280 and 260 nm was measured, and peak fractions were collected and concentrated using conical concentrators with 30,000 kDa MW cut-off. Concentrated eluate was then applied to pre-equilibrated Superose6 16/60 column at 1 mL/min and 2mL fractions were collected. Peak fractions were pooled, concentrated and flash frozen in liquid N_2 and stored for further use.

3.7 Hexokinase 2 R42H crystallization

Initial crystallization screening was performed using the commercial sparse matrix screen ProPlex™ (Molecular Dimensions). Robotic set up of sitting drop conditions was carried out using a Crystal Gryphon LCP robot (ARI – Art Robbins Instruments) under mixing protein and precipitant solutions at 1:1 and 2:1 volume/volume ratio for incubation at 20˚C and 4˚C. Optimization screens were designed with RockMaker™ software and prepared using a Formulator™ (Formulatrix, Inc.) microfluidic multi-channel dispenser. Sitting drop screens were set up as described above using 0.3 µl of protein solution (with and without 4 mM glucose and 5 mM ATP) and 0.15 µl of precipitant, trays were again incubated at 20˚C and 4˚C. Trays stored at 20˚C were handled by the RockImager™ (Formulatrix, Inc.) plate hotel. They were imaged in 12, 24, 48, 72h intervals using visible light wide-field microscopy, and after 48 hours also using UV-two photon excited fluorescence (TPEF) and second order non-linear imaging of non-chiral crystals (SONICC). Crystallization trials at 4˚C were inspected visually under a Leica M165 stereomicroscope.

3.8 Hexokinase 2 X-ray data collection and analysis

Several crystals from the optimization screen were harvested for data collection with addition of 1uL of 50% Glycerol v/v as a cryo-protectant to final concentration of 34.5 % v/v immediately prior to fishing with a 20 µm nylon CryoLoop (Hampton Research).

Data were collected at the PXI beamline at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) synchrotron using 20 x 30 µm X-ray beam at a wavelength of 1Å (12.4 keV) and an EIGER 16M X (133 Hz) detector at 30 cm from the sample. Data were indexed, integrated, scaled, and merged XDS (Ref) and CCP4(REF) via the SLS provided software pipeline. Phases were calculated by molecular replacement in PHASER using a structure of wild-type HK2 (PDB ID:5HEX) split into two (N-terminal residues 17-460 and C-terminal residues 480-911) as search models. The model was refined in PHENIX and REFMAC followed by manual examination and rebuilding in Coot. When the mainchain

model was in good agreement with the data (R-free <0.27), ligands were removed from the model and the b-factors were scrambled for another refinement in PHENIX. The resulting map was then used together with the previous coordinates file for ligand placement and subsequent refinement in REFMAC.

3.9 Hexokinase 2 substrate binding assays (nano-DSF)

For initial testing of Tm of apo HK2 variants wild-type and Δ16 R42H mutant HK2 were used in a concentration gradient from 0.1-7 mg/mL (1-70 µM) in 20 mM Bicine pH 8.0 with 250 mM NaCl, 1mM TCEP and 1 mM $MgCl_2$. The fluorescence measurements at 330 and 350 nm in course of thermal denaturation were recorded with NanoTemper Prometheus NT.48 using "high-sensitivity" quartz capillaries to hold the samples and applying 25-75˚C temperature gradient with a slope of 1.0˚C/min. Tm was then determined by calculating the $dR = \frac{d\left(\frac{F330}{F350}\right)}{dT}$ and its local maxima. To rule out any effects of aggregation, light scattering and it is first derivative was also measured.

For analysis of ATP/ADP/Glucose/Glucose-6-phosphate binding, the wild-type and R42H mutant proteins were used at the final concentration of 0.2 mg/mL (~2µM) in phosphate buffered saline (PBS) buffer. Tm was measured as described above, but with a 25-65˚C temperature gradient as it was sufficient to fully denature HK2.

To remove any aggregates, the protein solutions were centrifuged in a table-top centrifuge at 16000 x g for 10 minutes to remove any precipitate prior to dilution.

3.10 Chemical cross-linking

Unless otherwise stated, for the chemical cross-linking purified Raptor and RagAC, as well as Raptor, RagAC and Ragulator complexes were used at final concentration of ~5.4 µM (1.6 mg/ml) in 20 mM Bicine at pH 8.0 with 150 mM NaCl and 1mM TCEP buffer. Cross-linking reactions were performed at room temperature for 1 hour using the following stock reagents made immediately prior to use: EDC (Sigma Aldrich) - 100 mM, DMTMM (Signa Aldrich) - 100 mM, NHS-ester (Thermo Fisher Scientific) – 100 mM, ADH (Sigma Aldrich) – 100 mM, CDI (Sigma Aldrich) – 100 mM in anhydrous DMSO and BS3 (Thermo Fisher Scientific) in 20 mM Bicine at pH 8.0, 150 mM NaCl. Reactions involving NHS-ester and BS3 were quenched with Tris pH 8.0 or ammonium bicarbonate to final concentration of 100 mM.

3.11 Cross-linked protein digestion and peptide purification

100 ug of protein was diluted in 6M Urea or 1% sodium deoxycholate (SDC) in 0.1M ammonium bicarbonate. The sample with urea were reduced with 5 mM TCEP for 60 min at 37 °C and alkylated with 15 mM chloroacetamide for 30 min at 37 °C. Sample with SDC was additionally heated at 98°C for 10 minutes prior to reduction and alkylation. Due to precipitate formation in SDC sample, it was centrifuged at maximum speed for 10 minutes, and the pellet was resuspended in 6M Urea, while supernatant was kept for

further analysis as is. After diluting urea containing samples with 100 mM ammonium bicarbonate buffer to a final urea concentration of 1.6M, proteins were digested by incubation with sequencing-grade modified trypsin (1/50, w/w; Promega, Madison, Wisconsin) for 12 h at 37°C. After acidification using 5% TFA, peptides were desalted using C18 reverse-phase spin columns (Microspin, Harvard Apparatus) according to the manufacturer's instructions, dried under vacuum and stored at -20°C until further use. For complete removal of SDC prior to C18 purification acidified samples containing SDC were centrifuged for 10 minutes at maximum speed and the supernatant was carefully aspirated and applied to the column.

3.12 Cross-linked peptide enrichment

Dried peptides were resuspended in 20 µL of size exclusion chromatography (SEC) running buffer (10% Acetonitrile, 0.1% TFA) using a sonicator and centrifuged in a table-top Eppendorf centrifuge at maximum speed for 10 seconds. Resuspended peptides were applied to a Superdex S30 Increase 3.2/300 column (GE Healthcare) using the biocompatible UltiMate 3000 UHPLC purification system (Thermo Fisher Scientific) with 16 µL injection volume at 0.07 mL/min. Absorbance at 205 nm was monitored throughout the run and 100 µL fractions were collected and dried under vacuum for MSMS analysis.

3.13 Tandem mass spectrometry of cross-linked peptides

After SEC enrichment, dried peptides were resuspended in 0.1% aqueous formic acid and subjected to LC–MS/MS analysis using a Orbitrap Fusion Lumos Mass Spectrometer fitted with an EASY-nLC 1200 (both Thermo Fisher Scientific) and a custom-made column heater set to 60°C. Peptides were resolved using a RP-HPLC column (75µm × 36cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH) at a flow rate of 0.2 µLmin-1. The following gradient was used for peptide separation: from 5% B to 12% B over 5 min to 35% B over 40 min to 50% B over 15 min to 95% B over 2 min followed by 18 min at 95% B. Buffer A was 0.1% formic acid in water and buffer B was 80% acetonitrile, 0.1% formic acid in water.

The mass spectrometer was operated in DDA mode with a cycle time of 3 seconds between master scans. Each master scan was acquired in the Orbitrap at a resolution of 120,000 FWHM (at 200 m/z) and a scan range from 250 to 1500 m/z followed by MS2 scans of the most intense precursors in the Orbitrap at a resolution of 30,000 with isolation width of the quadrupole set to 1.4 m/z. Maximum ion injection time was set to 50ms (MS1) and 54 ms (MS2) with an AGC target set to 1E6 and 1E5, respectively. Only peptides with charge state 3 – 6 were included in the analysis. Monoisotopic precursor selection (MIPS) was set to Peptide, and the Intensity Threshold was set to 5e3. Peptides were fragmented by CID (Collision-induced dissociation) with collision energy set to 29% and one microscan was acquired for each spectrum. The dynamic exclusion duration was set to 30s.

3.14 Cross-link identification from MSMS data

The data was then searched with Merox 1.6.6.6, Merox 2.0.1.4 and XlinkX 2.0 for Proteome Discoverer against a database of top-20 (excluding contaminants and trypsin) sequences of proteins identified in pooled cross-linked fractions using a standard mass spectrometry protocol for protein identification. Where possible same settings were used for Merox 1/2: links to K, S, T, Y were allowed, minimum peptide length was set at 5, and where applicable maximum peptide length was set at 30; Methionine oxidation, and Asn and Gln deamidation were considered as variable modifications, dead end reaction with water and intrapeptide links were also considered; MS1 and MS2 tolerance were set at 10 ppm, RISE mode with maximum 1 missing ion was selected for both Merox 1 and Merox 2 versions (in Merox 2, max two isotope shifts were also allowed); minimum intensity cut-off was set at 10%, and minimum peptide scores was not applied; slow and precise scoring algorithms were selected, however in Merox 2 internal linear ions were excluded from scoring as this option was recommended; recommended signal-to-noise ratio of 2.0 was used for both. For XlinikX search, standard parameters were used with the same database. The following parameters differed: recommended mass tolerance of 20 ppm of MS2 fragments; signal-to-noise ratio was left at the recommended 1.5, Asn and Gln deamidation was not included as a variable modification due to increased time of the calculation, and FDR threshold was set at 1%; it was not possible to select the number of missing ions allowed with XlinkX.

3.15 Recombinant protein expression in HEK cells

Depending on the scale of expression, appropriate number of cells was seeded on cell-culture treated dishes and multiwell plates (Greiner). Smaller surfaces (all multi-well plates) were additionally treated with 0.01 mg/mL Poly-L-lysine to allow multi-step washing and media exchange if required. When the cells reached 70-80% confluence they were transfected with 1 µg of DNA/10000 cells (calculated based on plate surface area and expected cell number at 100% confluence) at a final concentration of 0.05 ug/mL. Transfection mixtures with FuGENE (Promega) were prepared according to manufacturer's protocol using 3:1 transfection reagent to DNA ratio. For large scale protein expression polyethyleneimine (PEI) was used as the transfection reagent at the same reagent:DNA ratio and final DNA concentration as FuGENE, but both DNA and PEI were diluted to 2x final concentration prior to mixing. Transfection mix was incubated for 5-10 minutes prior to addition to cell culture media. For experiments with PEI, the medium was exchanged 4-6 hours post transfection.

3.16 Metal affinity pull-down from HEK cell lysate

Cells were harvested either by scraping, or by incubating with 0.25% buffered solution of trypsin and 0.02% EDTA (Sigma-Aldrich) for 5 min at 37˚ C and centrifuged at 300 x g for 7 minutes. If trypsin was used for cell detachment, 10 ml of soya-bean trypsin inhibitor was added per 15 cm tissue culture plate prior to centrifugation. The cells were then washed

three times with Ca-free phosphate-buffered saline (PBS, Sigma Aldrich). Washed cells were lysed by resuspending them in HEK cell lysis buffer (20 mM Tris pH7.4 250 mM NaCl, 1mM TCEP, 10 mM imidazole, DNase A, 0.01% NP-40 and protease inhibitors: Phosphoramidon, Pepstatin A, Bestatin, Phenanthroline) and incubating on ice for 15-30 minutes. If the lysate was very turbid, the sample was passed through a G.27 syringe needle 10 times prior to centrifugation at 16000 x g for 10 minutes to produce cleared lysate. 5-10 µl Ni-NTA PureCube magnetic beads (Cube Biotech) were added to the supernatant and the mixture was incubated at 4 ˚C for 2 hours on a rolling mixer. The beads were then washed once with lysis buffer and twice with detergent-free PBS supplemented with imidazole to final concentration of 30 mM. The beads were then incubated twice with 50 µl of elution buffer (PBS, 300mM imidazole). For on-bead trypsin digestion, the beads were incubated with 0.5 mg of trypsin for 30 minutes in ABC, then sample was reduced and alkylated and digested with 1 mg trypsin (Promega)

3.17 Photo-leucine and photo-methionine incorporation

For all experiments with photo-leucine and photo-methionine, HEK cells were grown on tissue culture plates (96 well plates were poly-K treated) in Dulbecco's Modified Eagle Medium (DMEM) (Sigma Aldrich) with 5% foetal bovine serum (Gibco) to 70% confluency as for regular transfection. Then they were washed three times with Ca-free PBS (Sigma Aldrich) and the media was exchanged to DMEM-LM (L-Leucine and L-Methionine deficient DMEM, Thermo Fisher Scientific) with 5% dialyzed foetal bovine serum (DFBS, Gibco) and either L-leucine or L-photo-leucine (Iris Biotech) at 4 mM and either L-methionine or L-photo-methionine (Iris Biotech) at 2 mM prior to transfection. If a starvation step was included in the experiment, the cells were incubated in DMEM-LM with 5% DFBS overnight and supplemented with a version of L-leucine or L-methionine by 1:50 dilution of 100 mM and 50 mM stocks in DMEM-LM 5% DFBS pH adjusted with 100 mM KOH. For partial L-leucine and L-methionine supplementation, the stocks were diluted 100-fold prior to addition to medium.

3.18 UV cross-linking

The cells grown in 15 cm tissue culture dishes in Photo-leucine or Photo-methionine-containing medium were washed three times with Ca-free PBS (Sigma Aldrich) and covered with 10 ml PBS for cross-linking. Cross-linking was performed in a Stratagene Stratalinker (Agilent Genomics) with 365 nm light bulbs for three cycles of 5 minutes with 5-minute breaks in-between.

3.19 MSMS analysis of Photo-Leucine and Photo-Methionine incorporation rates in GFP and RagAC

Dried peptides were resuspended in 0.1% formic acid by sonication and subjected to LC-MS/MS analysis using a Q Exactive Plus Mass Spectrometer fitted with an EASY-nLC 1000 (both Thermo Fisher Scientific) and a custom-made column heater set to 60°C. Peptides

were resolved using a RP-HPLC column (75µm × 30cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH) at a flow rate of 0.2 µLmin-1. The following gradient was used for peptide separation: from 5% B to 10% B over 5 min to 35% B over 45 min to 50 % B over 10 min to 95% B over 2 min followed by 18 min at 95% B. Buffer A was 0.1% formic acid in water and buffer B was 80% acetonitrile, 0.1% formic acid in water.

The mass spectrometer was operated in DDA mode with a total cycle time of approximately 1 s. Each MS1 scan was followed by high-collision-dissociation (HCD) of the 20 most abundant precursor ions with dynamic exclusion set to 45 seconds. For MS1, 3e6 ions were accumulated in the Orbitrap over a maximum time of 250 ms and scanned at a resolution of 140,000 FWHM (at 200 m/z). MS2 scans were acquired at a target setting of 1e5 ions, maximum accumulation time of 50 ms and a resolution of 17,500 FWHM (at 200 m/z). Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events. The normalized collision energy was set to 27%, the mass isolation window was set to 1.4 m/z and one microscan was acquired for each spectrum.

3.20 MSMS for protein identification and mutant confirmation

30 ug of protein solution were resuspended in 1% Sodium deoxycholate, 10mM TCEP, 100mM Tris, pH=8.5 by sonication (Bioruptor, 10 cycles, 30 seconds on/off, Diagenode, Belgium). Samples were incubated for 10 min at 95°C, let cool down to RT followed by the addition of chloroacetamide at a final concentration of 15 mM. After an incubation of 30 min at 37°C, sequencing-grade modified trypsin (1/50, w/w; Promega, Madison, Wisconsin) was added and proteins were digested for 12 h at 37°C shaking at 300 rpm. Digests were acidified (pH<3) using TFA and desalted using iST cartridges (PreOmics, Martinsried, Germany) according to the manufacturer's instructions. Peptides were dried under vacuum and stored at -20°C.

Dried peptides were resuspended in 0.1% formic acid by sonication and subjected to LC-MS/MS analysis using a dual pressure LTQ-Orbitrap Elite mass spectrometer fitted with an EASY-nLC 1000 (both Thermo Fisher Scientific). Peptides were resolved using a RP-HPLC column (75 µm × 15 cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH) at a flow rate of 200 nl/min. The following gradient was used for peptide separation: from 5% B to 10% B over 5 min to 35% B over 45 min to 50 % B over 10 min to 95% B over 2 min followed by 18 min at 95% B. Buffer A was 0.1% formic acid in water and buffer B was 80% acetonitrile, 0.1% formic acid in water.

The mass spectrometer was operated in DDA mode set to obtain one high resolution MS scan in the orbitrap at a resolution of 240,000 full width at half maximum (at 400 m/z, MS1) followed by MS/MS (MS2) scans in the linear ion trap of the 20 most intense MS signals with scan rate set to "Rapid". The charged state screening modus was enabled to exclude unassigned and singly charged ions and the dynamic exclusion duration was set

to 30s. The ion accumulation time was set to 300ms (MS1) and 25ms (MS2). MS1 and MS2 scans were acquired at a target setting of 1E6 ions and 1E4 ions, respectively. Peptides were fragmented by CID (Collision-induced dissociation) with collision energy set to 35% and one microscan was acquired for each spectrum.

The data was analysed with MaxQuant 1.6.7.0, using a Uniprot Human proteome database (or *S. frugiperda* proteome spiked with sequences of overexpressed proteins) at 5%FDR using standard settings, except for the photo-amino acid variable modifications and match between runs options where required[201].

Pull-down data after UV-crosslinking was analysed for enrichment by D. Ritz using an in-house software SafeQuant 2.3.4[202]

4 Results

In this chapter I present a concise overview of work aimed at understanding protein-protein interactions. Of particular interest to me were highly dynamic and complex nutrient signalling networks. All work presented here is my own, except where otherwise stated. For example, for the Hexokinase 2 project, N. Dietz has started the collaboration with M. Shimobayashi and has purified the wild-type and mutant proteins, as well as set up the initial crystallization trials. M. Shimobayashi has performed the enzyme activity assay (published work referenced here). For the Ragulator-RagAC-mTORC1 complex, E. Sauer and E. Stuttfeld have obtained the RagAC plasmids and have cloned the initial mutant combinations (WT-WT, G65A-S75N); S. Werten has obtained the clones for the Ragulator complex and has produced the proteins on his visit to the Biozentrum; N. Dietz have performed cryo-EM studies and analysed the data to obtain the atomic resolution model. For the cross-linking mass spectrometry project, I have devised the study and completed the experiments, however Danilo Ritz from the PCF facility has analysed the RagAC pull-down data using SafeQuant, an in-house developed program for analysis of protein enrichment.

4.1 Characterization of Hexokinase 2 R42 H mutant

A recent study found that some blind Mexican cavefish (Astyanax mexicanus) have developed reduced insulin signalling as a genetic adaptation to long periods of starvation[203]. The study identified that three out of four types of fish tested had independently developed a mutation in their insulin receptor. In contrast, fish from Molino cave, while also being hyperglycaemic, had a wild-type insulin receptor and displayed normal insulin signalling, akin to diet-induced type 2 diabetes found in humans. Further investigation has shown that Molino fish harbour a genetic partial loss-of-function mutation in the hexokinase 2 gene that is likely responsible for the hyperglycaemic phenotype[204]. This Arg42 -> His42 mutation, when introduced into the human hexokinase 2 (HK2) was found to lower the activity of the enzyme compared to the wild-type by 50% when overexpressed in HEK293T cells[204].

In previously published structure of the wild-type HK2 R42 is located on helix α_3 and forms a salt bridge with aspartate 272 (fig 3.1.2.1 A)[205]. Both, the structured loop 263-293 and the helix α_3, are well-conserved across the species and within the hexokinase family. Most of the sequence of the loop and the helix is also conserved in the C-terminal domain of the hexokinase. Mutations in either of these regions have not been previously characterised in hexokinase 2. However, a triple mutation of E280A, R283A, and G284Y is known to disrupt dimer formation in HK1[206]. In both, hexokinase 1 and hexokinase 2, E280 makes a contact with K558 to stabilise the dimer.

The loop containing residues 263-293 is rich in charged and hydrophobic amino acids (10/29 each, and of the remaining 9, 4 are polar, 3 are glycines and 1 - proline). The fold

of this region is conserved in all hexokinase structures published to date, it is largely stabilised by internal interactions within the backbone and some of the charged sidechains (fig. 3.1.3.1 A). This region in hexokinases is kept in place through hydrophobic interactions, as well as two salt bridges and some polar contacts, namely: D278-R381; D272-R42; S285, N287, D282 and the backbone of K290 with K295 (fig. 3.1.3.1 A). These interactions are also conserved in the C-terminal half as D726-R829; D720-R490; D730, S733, N735 and K738 – K743. In the structure of the hexokinase 1 and wild-type hexokinase 2 (in one of the protomers) D276 also interacts with K312 further stabilising the position of the loop.

Thus, we hypothesised that R42H mutation would abolish the salt bridge to D272 and might somewhat destabilise the position of the loop. This could in turn impact either dimer formation or ligand binding by interfering with arrangement of Q291 and adjacent active site residue E294 (fig. 3.1.3.2). To address this hypothesis and determine the aetiology of the mutant's reduced activity *in vitro,* the wild-type and the R42H mutant were over-expressed in and purified from *E. coli* to evaluate any differences in substrate and product affinity as well as protein and dimer stability. The R42H mutant was also crystallized together with its substrates: ATP and glucose, and its structure in complex with glucose and glucose-6-phosphate was determined.

4.1.1 Crystallization of the HK2 R42H

The initial robotic crystallization screening yielded crystals in three conditions: (1) 0.1 M HEPES pH 7.0, 10% w/v PEG 4000, 10% v/v 2-propanol, (3) 0.1 M MES pH 6.5, 10% PEG 5000 MME 12% v/v 1-propanol, and (3) 0.1 M sodium phosphate pH 6.5 12% w/v PEG 8000. Best crystals from condition (2) diffracted to ~7Å. Based on the results of initial screening, systematic grid screening of conditions was carried out covering the condition range from 0-12% 1-propanol and 2-propanol over 4 wells, 8-20% PEG MME 5000 over 6 wells, and the buffer system 0.1 M MES at pH 6.2, 6.5, 6.8 and 7.0 for hexokinase with and without ligands (ATP and glucose) at 2:1 protein to precipitant ratio.

Already after 12-hour incubation at room temperature, crystals appeared in approximately 50% of the conditions containing ligands. For every combination of 1- or 2-propanol concentration and pH values, at least two out of six PEG concentrations resulted in crystal formation. Upon further incubation for 72 hours, some of the crystals increased in size, particularly in those conditions that had only a few crystals per drop, and some additional conditions had developed crystals at pH7 (8-12 % isopropanol and 15-20% PEG MME 5000).

Generally, higher isopropanol conditions resulted in a higher number of smaller crystals with irregular shapes. Crystals grown in higher pH also required higher PEG concentrations. Conditions required for crystal growth in 1-propanol vs. 2-propanol were nearly identical, with 2-propanol conditions yielding on average larger crystals.

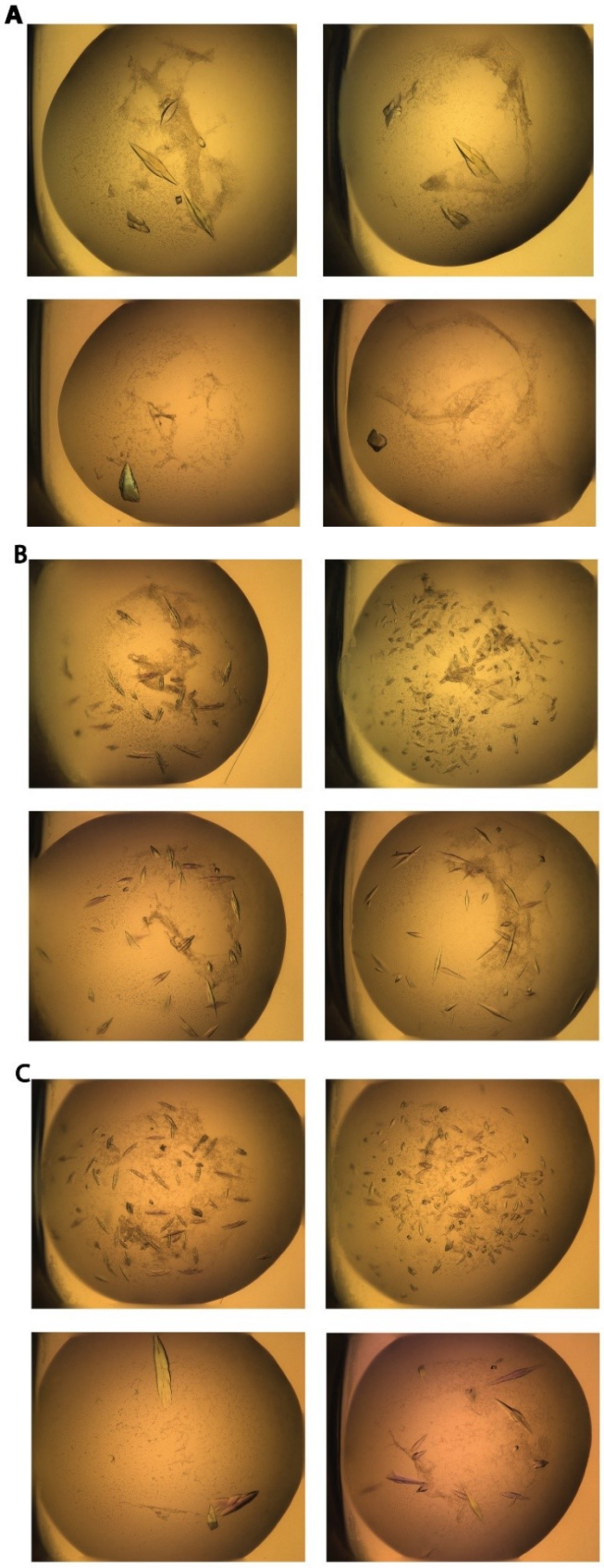

Figure 4.1.1.1 (A) Crystals of HK2 (R42H) obtained without isopropanol. Top-left: pH 6.2 PEG MME 5000 12.8%. Top-right: pH 6.5, PEG MME 5000 15.2%. Bottom-left: pH 6.8, PEG MME 5000 12.8%. Bottom-right: pH 7.0, PEG MME 5000 15.2%. (B, C) Crystals of HK2 (R42H) obtained with 4 % 1-isopropanol and 2-isopropanol, respectively. All other variables are the same as described for panel A.

Most crystals had similar morphology (elongated tetragonal bipyramids) in a range of sizes (50-300 µm at the largest dimension). Many of them had irregular edges and were visually intergrown (Figure 3.1.2.1). No crystals were obtained with ligand-free hexokinase 2 R42H, as in previous studies[205-208].

A diverse range of crystals was picked from the following conditions: 0% propanol in pH 6.2, 6.5, 6.8 with PEG MME 5000 concentration 12.8-17.6 % w/v, and 4% propanol in pH 6.5-6.8 with PEG concentration 12-17.6 % w/v.

Best diffracting crystal came from the mixture of 0.1 M MES pH 6.2, 12.8% PEG MME 5000, and 0% propanol and resulted in collection of a dataset at 2.71 Å resolution (Figure 3.1.1.2).

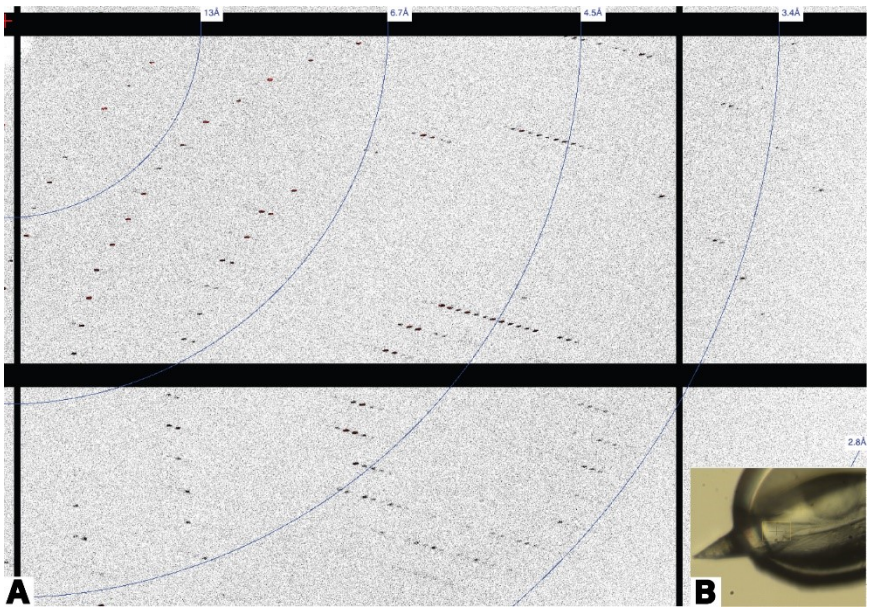

Figure 4.1.1.2 (A) Sample diffraction image. (B) Mounted crystal with beam position

4.1.2 Structure determination and model refinement

The structure was solved by molecular replacement using a previously available structure of wild-type hexokinase 2 separated at N-terminal and C-terminal domain boundary. The model was then subjected to multiple rounds of manual adjustment and refinement. Final refinement and data collection statistics can be found in table 3.1.3.1. PDB validation report can be found in Appendix I.

	Hexokinase 2 R42H
Data collection	
Space group	$P4_3\ 2_1\ 2$
Cell dimensions	
a, b, c (Å)	127.1 127.1 305.8
a, b, c (°)	90.0 90.0 90.0
Total reflections	1855477 (183612)
Resolution range (Å)	49.7-2.7
R_{sym} *or* R_{merge}	0.199(3.376)
Mean I / sigma (I)	16.71 (1.20)
Completeness (%)	99.80% (98.47)
Redundancy	27.1 (27.8)
Wilson B-factor	77.83
Refinement	
Resolution (Å)	2.715
No. reflections	68428
R_{work} / R_{free}	0.1925 (0.3247)/0.2312 (0.3360)
No. atoms	14623
Protein	14065
Ligand/ion	116
Water	442
B-factors	82.86
Protein	83.31
Ligand/ion	69.93
Water	72.04
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.390

Table 4.1.22.1 X-Ray data collection and refinement statistics. Values for the highest resolution shell in brackets

4.1.3 Crystal structure of the HK2 R42H mutant

As with other 100 kDa hexokinases, the mutant HK2 crystallized as a dimer in complex with its natural product (glucose-6-phosphate) and substrate (glucose). The previously described salt bridges at the dimer interface including K558-E280 were preserved in our structure.

The monomers are elongated and consist of two globular hexokinase domains connected via a linker helix so that the substrate binding pockets of each domain face away from each other.

When aligned separately to known hexokinase structures, the overall N- and C- terminal domains have near identical fold to the majority of known hexokinase structures, despite the diversity of crystallization conditions, space group and unit cell shape. (Table 3.1.3.1).

Table 4.1.3.1 differences of HK2 R42H to known hexokinase structures. Abbreviations: BG6 – beta-D-Glucose-6-phosphate, GLC – alpha-D-Glucose, BGC – beta-D-Glucose, G6P – alpha-D-Glucose-6-phosphate, N – N-terminal domain, C – C-terminal domain.

Protein name (PDB ID)	pH	Space group	Resolution	Ligands/ state	Oligomeric state	RMSD (Å)
Hexokinase 2 R42H ()	6.2	P43 21 21 2	2.7 Å	BG6, GLC, Na	Dimer	NA
Hexokinase 2 WT (2NZT)	8.5	P21 21 21	2.45 Å	BG6, GLC	Dimer	N=0.28 C=0.38
Hexokinase 2 WT (5HEX)	5.0	P1 21 1	2.71 Å	Inhibitor	Dimer	N=0.72 C=1.00
Hexokinase 2 WT (5HG1)	5.5	P31 2 1	2.76	Inhibitor, BG6, citrate	Monomer	N=0.53 C=0.92
Hexokinase 1 WT (1HKB)	6.0	P1 21 1	2.8 Å	BGC, G6P, Ca	Dimer	N=0.65 C=0.45
Hexokinase 1 E280A, R283A, G284Y (1CZA)	5.8-6.5	P21 21 2	1.9 Å	G6P, GLC, ADP	Monomer	N=0.52 C=0.50
Hexokinase 1 E280A, R283A, G284Y, T536A (1DGK)	5.8-6.5	P21 21 2	2.8 Å	GLC, PO4, ADP	Monomer	N=0.53 C=0.50
Hexokinase 4 WT (1V4T)	6.6	P65 2 2	3.4 Å	SO4, Na	Monomer	All to: N=0.99 C=1.51
Hexokinase 4 WT (3IDH)	7.5	P21 21 21	2.14 Å	GLC, K	Monomer	All to: N=0.80 C=0.84

In line with our initial hypothesis, the salt bridge between H42 and D272 is not supported in the mutant. However, the conformation of the 265-293 loop in the crystal structure is unaffected (RMSD 0.26 Å) when compared to that of the wild-type HK2 and other hexokinases (fig 3.1.3.1 A). The biggest observed difference in the conformation of this loop is to the apo structure of the hexokinase 4 (glucokinase) with RMSD of 0.634 Å and even then, all other conserved interactions are preserved. The position of the helix α_3 is similarly unaffected by the R42H mutation.

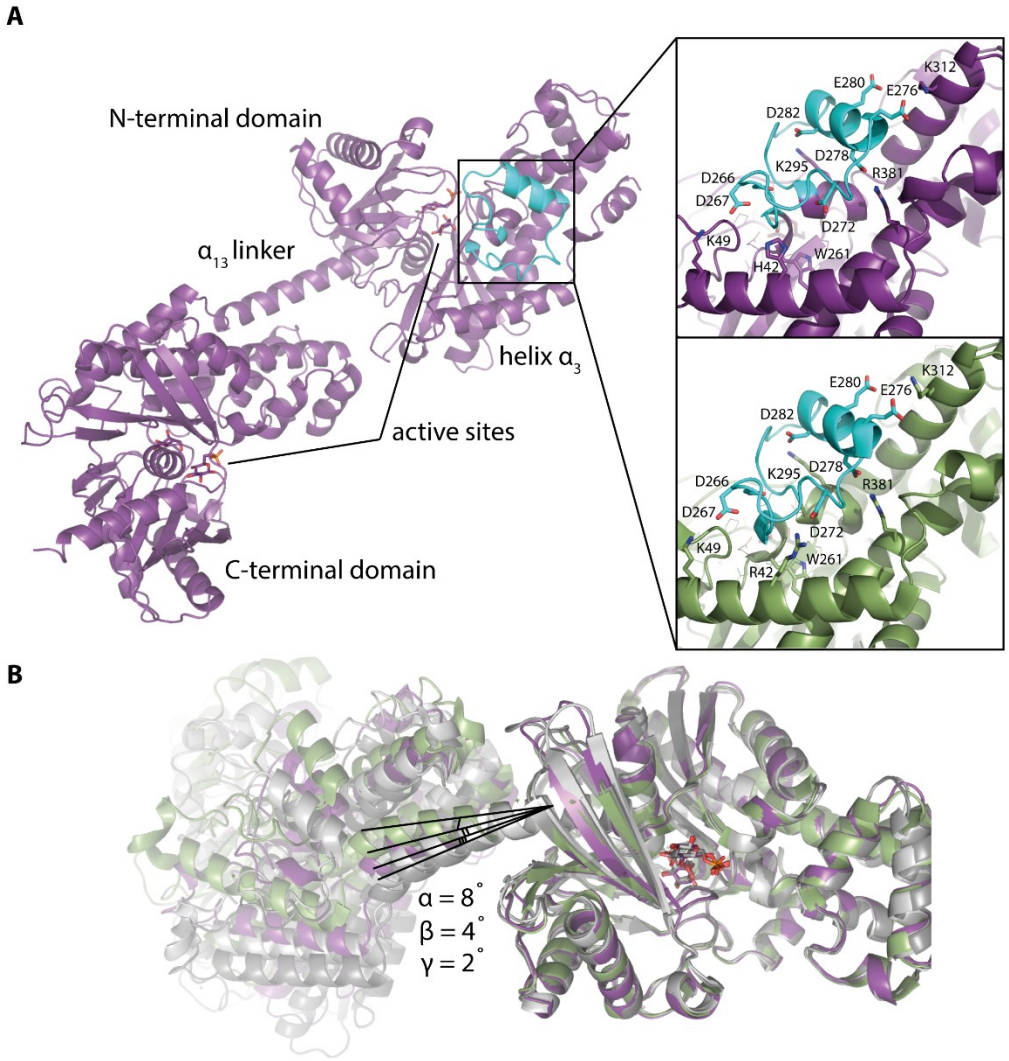

Figure 4.1.3.1 (A)Crystal structure of the hexokinase 2 R42H. For clarity only one of the chains is shown. The ordered loop 265-293 is highlighted in cyan with relevant residues shown as sticks and labelled in the insert. The same residues in WT structure are shown in green. (B) A comparison of HK2 WT (green), R42H (purple), WT HK1 (dark grey), and dimerization-incompetent mutant of HK1 (white). All structures are aligned at the N-terminus and the angles of the bend in connecting helix of the HK2 R42H mutant to WT (β) and HK1 (α) are labelled.

The major factor that prevented the calculation of a single overall alignment of all hexokinases is the flex angle of the linker helix α_{13}, which connects the N-terminal and C-terminal domains, as shown in fig. 3.1.3.1 B. Possible origins and significance of this difference from ligand binding and dimer formation were thus considered.

Of the hexokinase 1 and 2 structures compared in table 3.1.2.1, the most different conformation of this helix is observed in structures with unnatural inhibitors. With natural ligands, the active site of the N-terminal domain of the R42H mutant, occupied by β-D-glucose-6-phosphate (the product) and α-D-glucose (substrate, although not preferred) is near identical to the wild-type protein and the dimerization deficient hexokinase 1 (PDB ID: 1CZA) (fig 3.1.3.2). However, the flex angles of the linker helix were still different among these structures, which means that there are likely other factors involved in determination of this angle.

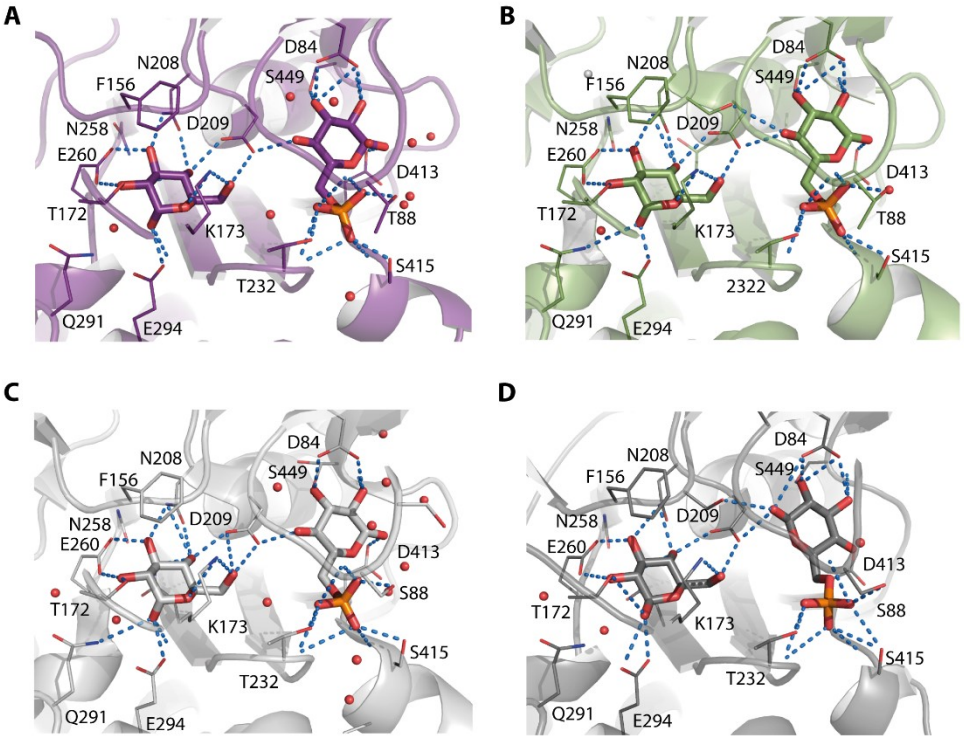

Table 4.1.3.2 Glucose and glucose-6-phosphate binding sites of the N-terminal domain in Hexokinase 2 R42H (A) and wild-type (B), and hexokinase 1 (dimerization deficient triple mutant (C)) and wild-type(D)

While the dimer interfaces in all hexokinase structures overlap, they are not identical, and the different interactions involved could provide another explanation for the difference in the bend of helix α_{13}. However, the dimerization-deficient triple mutant of hexokinase 1 (E280A, R283A, G284Y) together with the dimeric hexokinase 1 have the most similar flex angles in helix α_{13} to the R42H hexokinase 2 mutant, ruling out the contribution of the interactions at the dimer interface to the flex angle. Although it is possible that the flex angle contributed to differences in the dimer interface instead.

4.1.4 Characterisation of HK 2 R42H stability and ligand binding

As the crystal structure represents the most stable dimeric conformation inhibited by the product β-glucose-6-phosphate, it is possible that small changes in dynamics and substrate binding in either of the domains may not be reflected. To test the overall stability of the mutant, I have measured the melting temperature of both R42H mutant and wild-type protein with (and without) substrates and product using nanoDSF. In contrast to the traditional approach using a fluorescent die, this method relies on the red-shift of intrinsic tryptophan fluorescence when it becomes exposed to the solvent upon thermal denaturation[209-211]. This change is measured through the fluorescence ratio at 350/330 nm and the denaturation curve is plotted as its first derivative, as shown in fig. 3.1.4.1.

Hexokinase 2 has five tryptophane residues per monomer: three in the N-terminal part and two in the C-terminal part. Of these, only two tryptophane side chains are completely buried in both domains when the protein is folded: W171 and W261, and their counterparts W619 and W709. The remaining W94 is already almost fully exposed even in the crystal structure and likely has little effect on the fluorescence ratio change. While W619 is held in place only through interactions within the C-terminal domain, W171 is also protected from the outside by the dimer interface, if such is present in solution. Both W261 and W709 are close to the active sites of the protein and likely contribute to the arrangement of glucose binding residues E294 and E260 and their counterparts (fig. 3.1.3.1 and 3.2.3.2). They are protected from solvent by the helix α_3 and 265-295 loop (and their equivalents in the C-terminus) as seen in fig. 3.1.3.1 A.

Similar to the loop, in addition to hydrophobic interactions, the position of helix α_3 is stabilised by three salt bridges: D33-R433, E48-R396, and R42-D272. Thus, it is possible that the missing R42-D272 interaction in the R42H mutant contributes to lowering the energy required to expose W261 and therefore lowering the Tm of the protein. As protein dimerization at higher protein concentrations provides additional protection to W261 and W171 via another salt bridge to the loop region and hydrophobic interactions, the effect of the mutation may be amplified when the protein is a monomer in solution at RT.

It was previously shown that, in absence of glucose-6-P, hexokinase 1 is a monomer in solution up to a concentration of 3.6 mg/ml, but in its presence, 80% was found to be in a dimeric form already at 0.9 mg/ml. Due to overall similarity of the interfaces in crystal structures and evidence of hexokinase 2 dimerization *in vivo*, its dimerization behaviour is expected to be similar to that of HK1, with the exception of insensitivity to orthophosphate[206, 208, 212]. Unfortunately, the stability of the dimer interface cannot be monitored through this method directly (there are no tryptophane residues at the interface, the change in fluorescence of which could be measured). However, the overall T_m, would be significantly influenced by dimer formation due to several additional stabilising hydrophobic and ionic interactions.

Without ligands, and up to the concentration of 10 µM (1 mg/ml), the T_m of the R42H mutant is on average 1˚ C lower than that of the wild-type protein (fig. 3.1.4.1). At higher concentrations, the T_m of wild-type hexokinase 2 increased in a concentration-dependent manner, consistent with the known propensity for dimerization. For reference, the T_m of lysozyme only increased at concentrations above 12 mg/ml (700 µM). The melting temperature of the R42H mutant also increased at higher concentrations, but only those above 40 µM (4 mg/mL) suggesting that in absence of glucose, or glucose-6-phosphate, dimerization of the mutant may be impaired by the mutation. Additionally, unlike the wild-type protein, at concentrations of R42H mutant below 4 µM (0.4 mg/ml), the denaturation curve broadened, and the peak showed a shoulder at lower temperatures. While this had little to no effect on the calculated T_m for the entire protein, it represents a change from mostly uniform unfolding around 41-42˚ C with an additional event at around 47˚ to a stepwise unfolding, with at least part of the protein unfolding already at around 38˚ C.

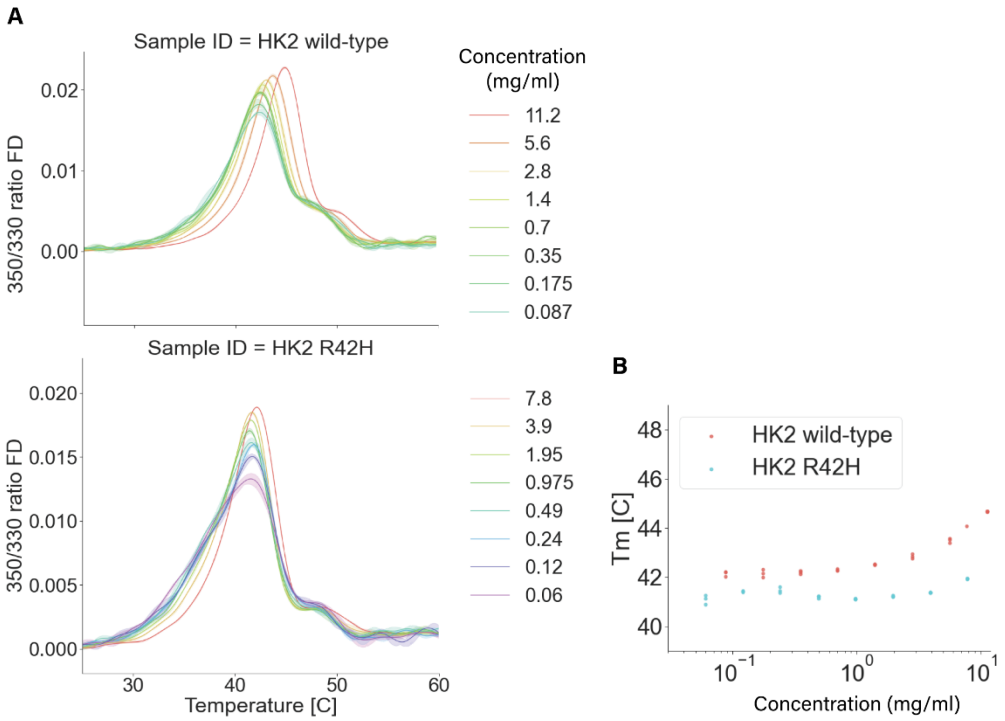

Figure 3.1.4.1 Effect of protein concentration on stability of wild-type and R42H mutant of HK2. (A) First derivative of the 350/330 nm fluorescence at varying protein concentrations. (B) Scatter plot of the melting temperatures. n=3

As the biggest difference in stability with differences in both, the shape of the peak and the overall T_m was observed at concentrations below 0.4 mg/ml, subsequent ligand binding experiments were performed at the concentration of 0.2 mg/ml to maximise the signal: noise ratio (S/N) of the measurement and still see the peak broadening effect of the mutation. The effect of substrate and product (ATP, glucose and glucose-6-phosphate) binding on T_m was tested (fig 3.1.4.2 and 3.1.4.3).

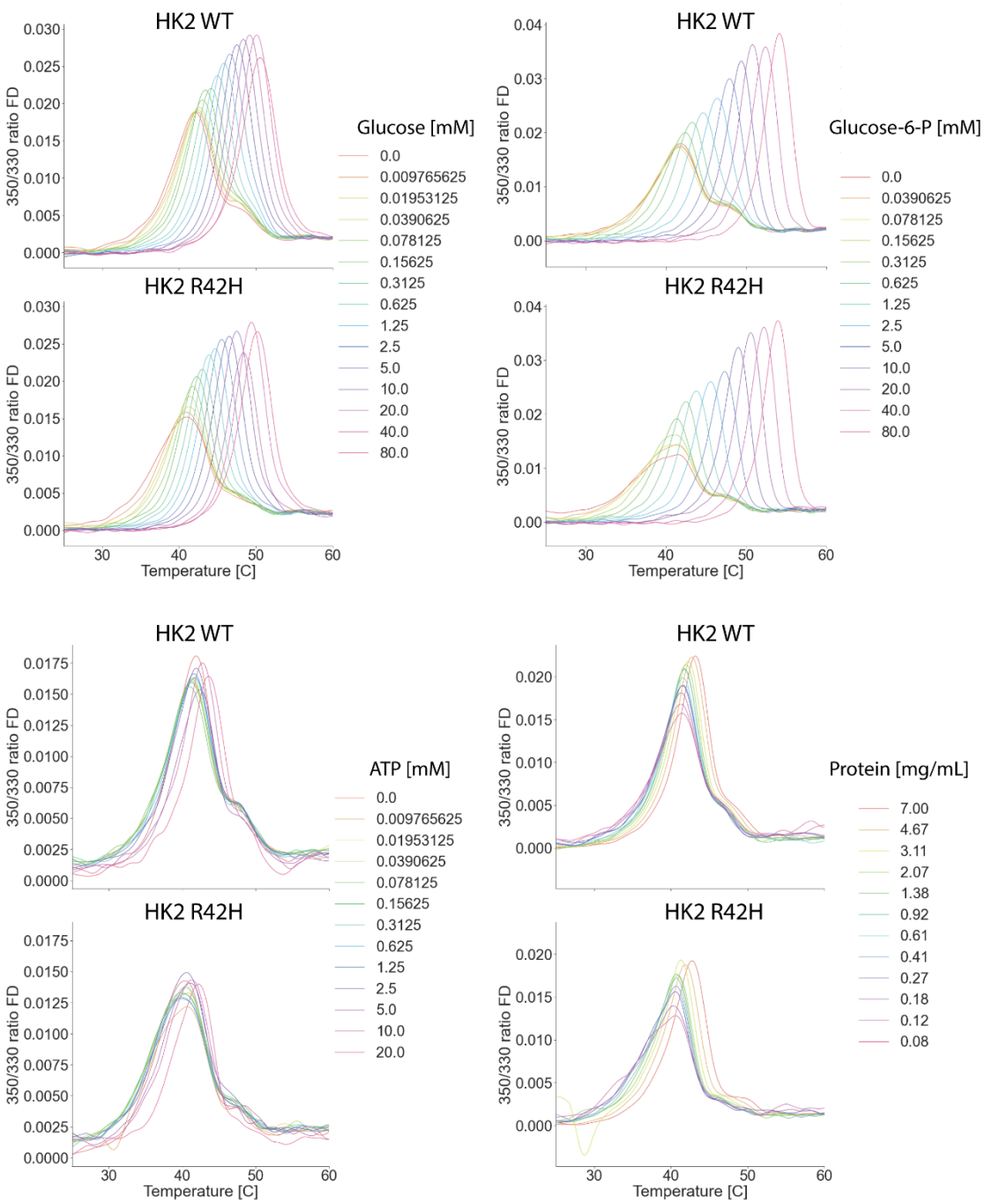

Figure 3.1.4.2 Representative thermal denaturation curves (first derivative of the fluorescence ratio at 350 and 330 nm) of the WT and R42H HK2 variants with and without ligands (n=3).

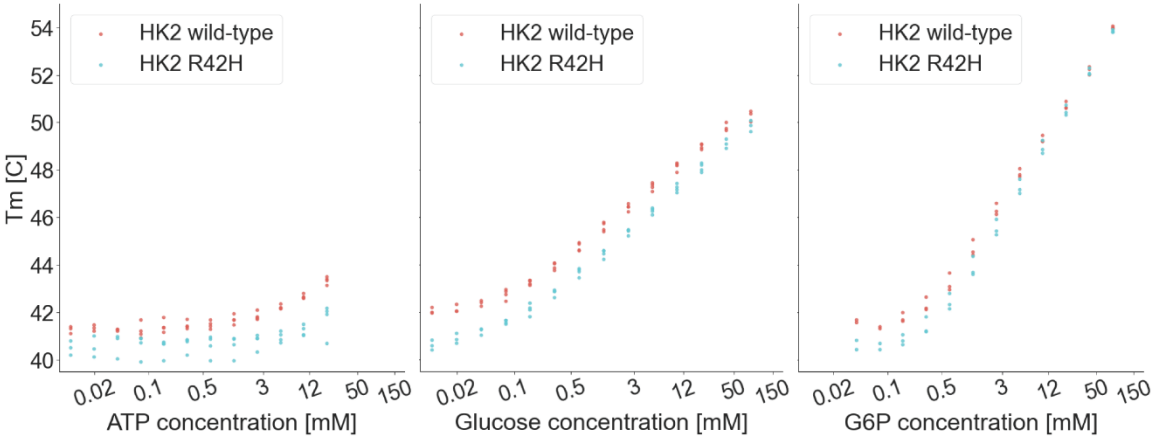

Figure 3.1.4.3 Changes in melting temperature of wild-type and R42H mutant in response to ligand binding (n=3).

ATP binding did not considerably increase the Tm of either the wild-type protein, or the mutant, until their concentration reached 5 mM, which is over 10-fold higher than the previously reported K_d for the N-terminal domain and nearly 5-fold higher than that reported K_d for the C-terminal domain[213]. However, it is in agreement with previously reported dye based DSF data[205]. Higher concentrations of ATP could not be tested due to substantial tryptophane-fluorescence reduction. Moreover, higher concentrations of ATP are not physiological, and it is more likely that ATP-binding does not induce a substantial conformational change which would increase the protein Tm. It should be noted, however, that the shoulder of the peak in the R42H mutant is only present until ATP concentration of up to 40 µM (i.e., close to the reported K_d for N-terminus of HK2), suggesting that ligand binding may slightly improve the overall stability of the 265-293 loop and helix α_3 in the mutant, leading to behaviour like that of wild-type.

As seen in figure 3.1.4.2, both glucose, and glucose-6-phosphate (G-6-P) substantially increased the T_m of both, the wild-type and the R42H mutant HK2 at concentrations above 40 and 80 µM for glucose and glucose-6-phosphate. Additionally, the shoulder of the denaturation curve present in the mutant HK2 was reduced at glucose and G-6-P concentrations of 0-80 and 0-160 µM respectively, without any changes in the overall T_m. This also indicates that, like ATP binding, glucose and G-6-P binding improved the stability of the R42H mutant without substantially improving the overall T_m of either the mutant or wild-type protein, at least initially. Further increase in concentration of both ligands at and above the previously reported K_m for glucose (0.13 mM) resulted in substantial (up to 10˚ and 15˚ C respectively) concentration-dependent increase in melting temperature (fig. 3.1.4.1 and 3.1.4.2). This is consistent with previously published findings[205].

4.2 RagAC in lysosomal mTORC1

RagAC GTPases act as a gateway between the lysosomal membrane associated complex Ragulator and mTORC1 substrate recognition component Raptor. They are also the point of integration of all amino-acid sufficiency signals that are required to initiate mTORC1 translocation to the lysosome. As differential binding of RagA and RagC to GTP and GDP is required for signalling, several mutations within their GTPase domains have been previously characterised to result in a defined nucleotide loading state *in vitro*. A mutation of Q66 to leucine in RagA makes the enzyme incapable of GTP hydrolysis, and a mutation of S75 in RagC is known to reduce the enzyme's affinity to both nucleotides, shifting the preference of the binding towards GDP. Overexpression of the double mutant has been previously shown to result in a constitutively active mTORC1 signalling in absence of amino acids. A combination of either of these mutations with a wild-type partner was also shown to have differential effects on mTORC1 signalling. However, the effect of these mutations on protein structures and protein-protein interactions was not well understood. So, the aim of the research presented in this chapter was to purify the different mutant combinations and produce a stable complex with Raptor (and mTORC1) and Ragulator for structural characterisation with cryo-EM to better understand the interactions involved in mTORC1 translocation initiation.

4.2.1 RagAC mutant combinations

Three mutant combinations of RagAC were purified and validated by MSMS (table 3.2.1.1, appendix II)

Table 4.2.1.1 RagAC mutants purified for this study.

RagA	*RagC*	*Combined effect*
WT	WT	Wild-type conformation, dependent on the native nucleotide loading state
WT	S75N	$RagC^{GDP}$. RagA state would depend on GATOR1 complex, or be altered through nucleotide loading without impact on RagC state
Q66H	S75N	$RagA^{GTP}$-$RagC^{GDP}$, fully active conformation irrespective of nucleotides.

4.2.2 Generation of RagAC complex with Raptor and Ragulator *in vitro*

There are two main strategies to produce a protein complex: co-expression and purification, or separate purification of individual components, which can then be mixed to obtain a complex. Both strategies have distinct advantages and disadvantages. Co-expression of the constructs allows for circumventing any ambiguities in requirements for complex formation which can be provided inside the cells. It is also beneficial in cases where the components of the complex are required for each other's folding and stabilisation. The disadvantages of this method are that the complex formation is more difficult to control, and the complex may be difficult to maintain during purifications. The

advantage of separate purifications, on the other hand, is that it is possible to purify mg quantities of each component at relatively high concentrations, manipulate each of the components and optimise protein concentrations and buffer to enhance complex formation. The main disadvantage of this approach is for protein complexes which are obligatory for protein folding, or if the assembly of the complex is also dependant external chaperones and co-factors. In these cases, purifying individual components is often impossible as they are unstable and are either degraded or aggregated before purification. Then, co-expression is a great alternative. For example, RagAC complexes used in this study were always co-expressed in insect cells as the proteins are obligate dimers. For RagAC-Raptor complex generation both approaches were tested.

For co-expression, insect cells were simultaneously infected with Baculovirus particles containing the Bacmid of His-tagged RagAC variants and His-Myc-Flag-tagged Raptor. A small-scale pull-down experiment was performed using Anti-FLAG antibody beads. While a sub-stoichiometric amount of the $RagA^{WT}$-$RagC^{GDP}$ and double $RagA^{GTP}$-$RagC^{GDP}$ mutant could be seen co-eluting with Raptor, the wild-type combination was only seen in the wash fractions (fig. 3.2.2.1).

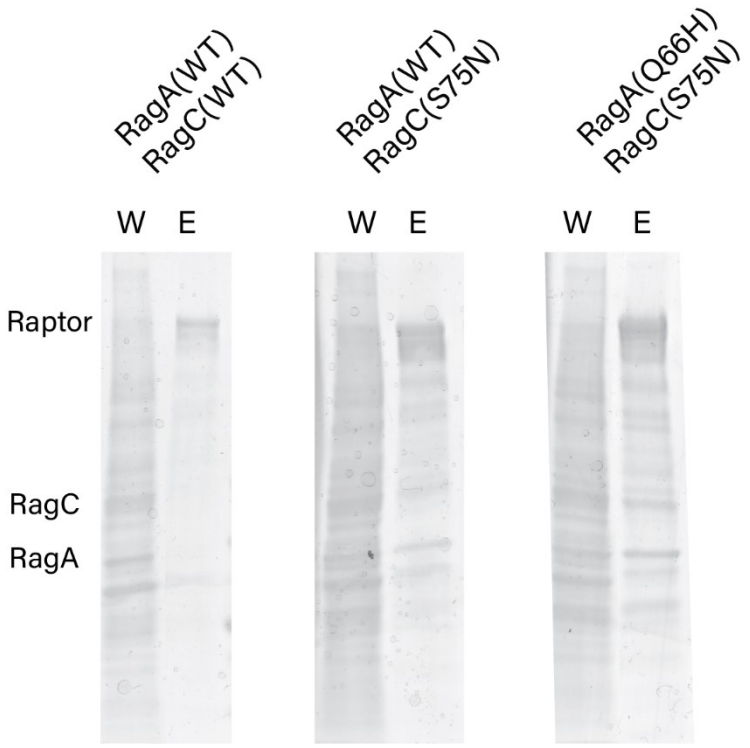

Figure 4.2.2.1 Small-scale co-expression of RagAC mutants and Raptor. Wash 1(W) and Eluate (E) fractions are shown.

To evaluate complex formation from individually purified components, analytical size exclusion chromatography was performed with and without nucleotide exchange (fig. 3.2.2.2 A, B).

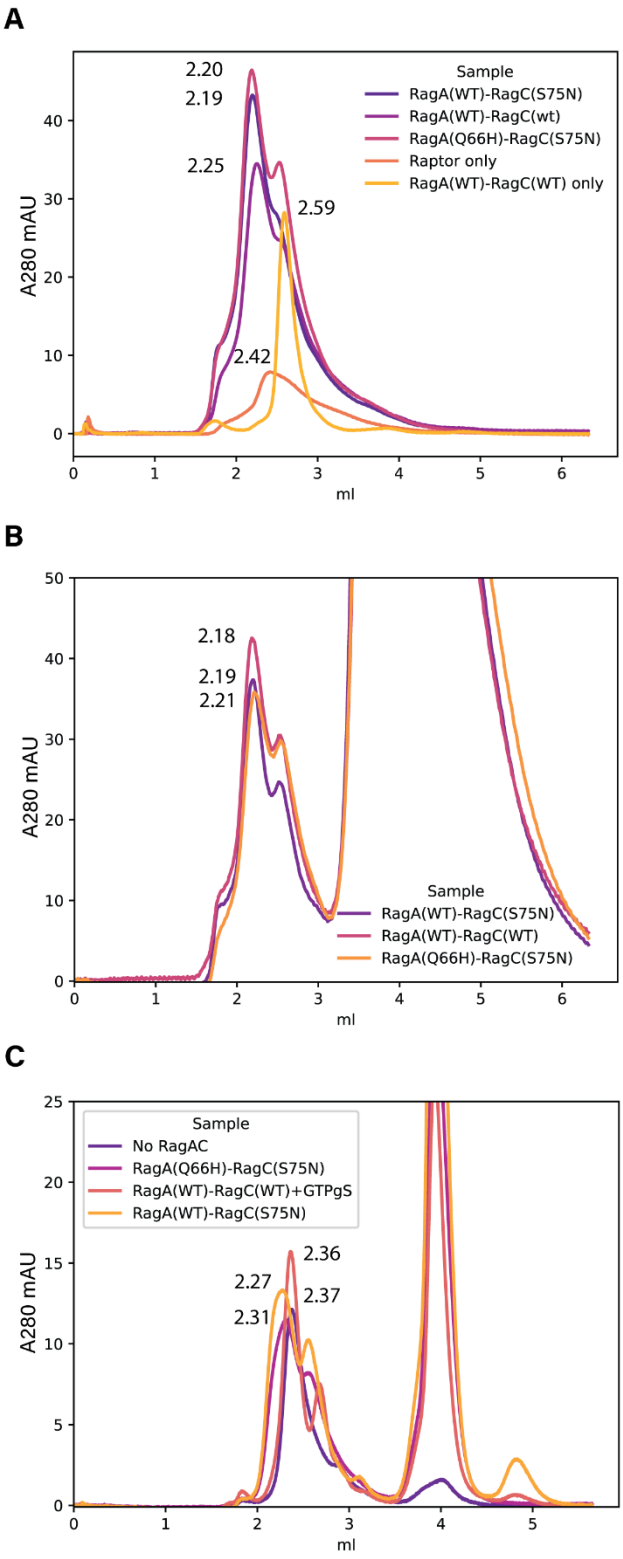

Figure 4.2.2.2 Size-exclusion chromatography. (A) Raptor binding of mutant combinations as purified from insect cells. (B) Raptor binding of RagAC variants after nucleotide exchange with 4:1 GTP:GDP ratio. (C) Ragulator-RagAC-Raptor complex. WT-WT combination nucleotide exchanged with 20 mM GTPγS, rest with GTPγS: GDP at 10:1 ratio. Peak elution volumes are labelled at the height of the peak.

All combinations of RagAC in a native state led to a slight shift in elution time of Raptor, with $RagA^{WT}$-$RagC^{S75N}$ and the double mutant producing a more substantial shift than the wild-type protein. SDS-PAGE analysis of the size-exclusion fractions showed evidence of RagAC presence in earlier fractions. Nucleotide exchange improved the shift for the $RagA^{WT}$-$RagC^{WT}$ combination, while having little effect on the mutants. Although, the peak shift was observed, further analysis of the double mutant with SEC-MALS and AUC failed to show any substantial increase in molecular weight of the largest species, indicating that the interaction was likely unstable when subjected to SEC analysis or sedimentation. To improve complex-formation, in collaboration with S. Werten from K. Scheffzek lab, which published the crystal-structure of the Ragulator-RagAC(ΔN), we included Ragulator complex for analysis as well (fig. 3.2.2.2 C). Here we used GTPγS for nucleotide loading of RagAC in a 10:1 ratio to GTP (20:2 mM) and used the WT dimer loaded with GTPγS as a negative control. As with Ragulator-RagAC-binding, both $RagA^{WT}$-$RagC^{S75N}$ and $RagA^{Q66H}$-$RagC^{S75N}$ induced a small shift in elution volume of Raptor, while GTPγS loaded wild-type dimer did not. This indicated at least some degree of binding of both mutants, but 50-100 µL change in elution volume was not consistent with the stable formation of a complex of 300 kDa. As both mutant combinations produced a larger change in elution volume than wild-type combination, and $RagA^{WT}$-$RagC^{S75N}$ resulted in a larger shift compared to the double mutant when together with Ragulator, this combination was used for EM-grid preparations of the Ragulator-Raptor-mTORC1 supercomplex.

4.2.3 Ragulator-RagAC-Raptor cryo-EM structure

As size-exclusion experiment failed to yield a stable complex, the grids were prepared with freshly assembled complex without further purification. Grid preparation, data collection and analysis to obtain a reconstruction at 3.5Å overall resolution of the mTORC1 complex and RagAC and 4.5Å at the Ragulator was performed by N. Dietz (manuscript in preparation). The structure of the core of the mTORC1 (mTOR, mLST8 and the conserved N-terminal region (CNR) of Raptor) was near identical to the previously published inactive structure of the mTORC1 without Rheb (PDB ID: 6BCX)[37]. Raptor HEAT repeats and WD-40 domain however, moved closer to the Horn domain of mTOR through a slight flex in the HEAT repeats region of Raptor (fig 4.2.3.1 A). When compared with the published RagAC-Raptor cryo-EM structure, the interfaces of RagA and Raptor, as well as the Raptor "claw" were virtually identical (PDB ID: 6U62)when superposed on RagA RMSD = 0.64 Å) (fig 4.2.3.1 B) [86], but the position of the CNR of Raptor is somewhat changed relative to the HEAT repeats and the WD-40 domain (fig. 4.2.3.1 C). In this case, there were also some differences in the arrangement of Ragulator proteins away from the Rag interface, which, together with worsened resolution of the reconstruction in this area, indicated a degree of flexibility in the Ragulator complex arrangement. In the published structures of RagAC-mTORC1 complex, and Ragulator-RagAC-Raptor proteins, there were no such conformational changes in Raptor, indicating that the assembly of the full complex is required for this change to take place. Overall, it results in improved hydrophobic

interactions between Raptor and the Horn domain of mTOR (fig. 4.2.3.1 D). And overall flattening of the supercomplex.

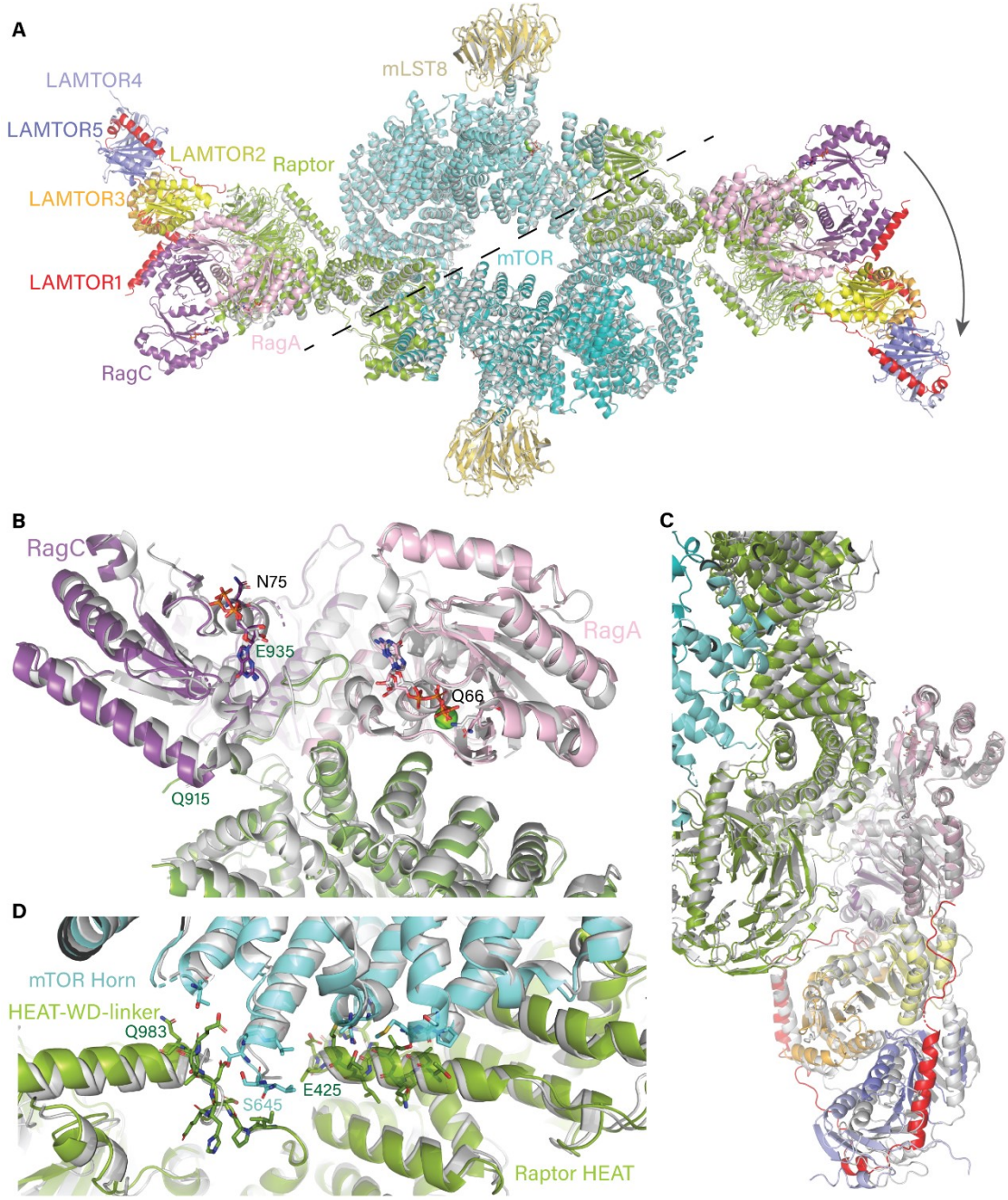

Figure 4.2.3.1 3D model reconstruction of the lysosomal mTORC1. (A) Complete mTORC1, overlayed on previously published inactive lysosomal mTORC1 model (6BCX). Arrow indicates small fluctuations in conformation and the overall direction of the RagAC-Ragulator movement relative to that structure. (B) Raptor 'claw'-RagAC binding interface superposed onto the crystal structure (PDBid:6S62). (C) Lysosomal mTORC1 model superposed on Raptor WD-domain of the crystal structure (6S62) shown in grey. (D) mTOR Horn region and Raptor HEAT repeat interactions relative to the inhibited mTORC1 structure (6BCX)

4.3 Ultra-short cross-linking mass spectrometry

To evaluate the potential of use of zero-length cross-linkers in structural studies, I tested two types of cross-linking approaches: specific chemical cross-linking of purified proteins *in vitro*, and UV-induced cross-linking *in vivo* through native incorporation of photo-reactive analogues of leucine and methionine. For chemical cross-linking, I have chosen to work with 1,2-carbonyldiimidazole (CDI) because of its reactivity with four amino acids (K, S, T, Y), which would increase the chances of cross-link formation, its reaction time and unstable intermediate for more specificity, and its MS-cleavability for ease of peptide identification.

4.3.1 CDI cross-linking on purified proteins

There are three things that are important to balance in any cross-linking reaction: protein concentration versus cross-linker concentration (i.e., cross-linker to protein ratio), reaction buffer and reaction time. Protein concentration should be high enough to support complex formation, but low enough to prevent unspecific cross-linking and aggregation. Concentrations of 1-2 mg/ml are typically recommended. Similarly, the cross-linker concentration should be sufficient to cross-link true available amino-acid pairs, but low enough to avoid aggregation. Conventional cross-linkers such as BS3 and DSS are usually used at concentrations below 1 mM. The original publication that described the use of CDI for cross-linking mass-spectrometry had tested it in concentrations of up to 0.5 mM[176] for a 66 kDa protein at 10 µM (0.66 mg/ml). As larger proteins have more reactive residues per molecule, it is logical to use higher cross-linker to protein molar ratio to compensate for this. RagAC and Raptor have a combined molecular weight of approximately 220 kDa, so higher concentrations were likely to be required. As CDI readily reacts with buffer, to test cross-linker-to-protein ration, I cross-linked 20, 5, or 1.25 µM RagAC-Raptor mixture with 5 mM CDI concentration to have the highest chances of successful cross-linking (fig 4.3.1.1(A)).

In line with expectations and previously published research, there was an inverse relationship between protein concentration and cross-linking efficiency. Although some evidence of cross-linking could be observed also for protein concentration of ~4.4 mg/ml (20 µM), 5 mM CDI was not sufficient to produce the full effect seen for the samples at 5 and 1.25 µM. These working molar ratios are 20-80 times higher than those used for BSA in the original publication. Initial experiments showed high variability between replicates, often resulting in no visible cross-linked product on the gel. Therefore, I investigated the stability of CDI in aqueous buffer by dissolving CDI stock in cross-linking buffer immediately prior to its addition to the protein solution, and comparing the cross-linking efficiency to control samples, which were instead quickly added to the CDI stock solution (in certified anhydrous DMSO) ensuring immediate mixing.

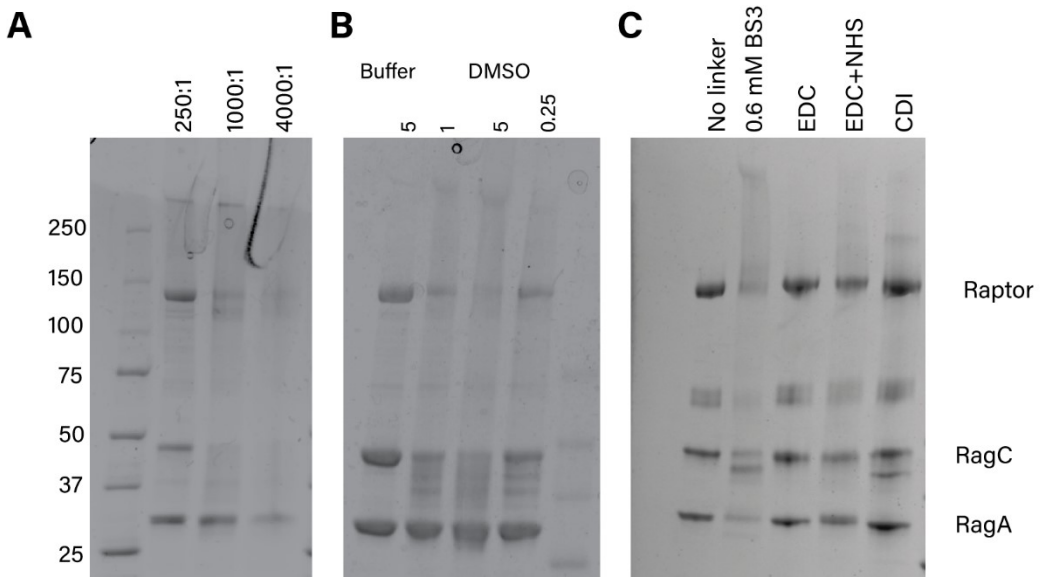

Figure 4.3.1.1 (A) Cross-linker-to-protein ratio optimisation for 5 mM CDI. Wells are labelled with molar ratio of CDI to RagAC-Raptor complex (250:1 - 20 µM, 1000:1 - 5 µM, 4000:1 1.25 µM). (B) Hydrolysis of CDI in aqueous buffer vs DMSO. Wells are labelled with CDI concentration (mM). (C) Ragulator-RagAC-Raptor complex cross-linking, comparison of cross-linker reactivity. EDC, EDC+NHS, CDI concentration – 2 mM.

As shown in figure 4.3.1.1 B, cross-linking could not be observed even when using 5mM CDI dissolved in buffer, while even 0.25 mM CDI in anhydrous DMSO showed signs of cross-linking. Thus, it is more than 20-times less effective at cross-linking after just a few seconds in aqueous buffer. As a result of this experiment, all future cross-linking with CDI was done by adding the protein solution directly to a small volume of CDI in anhydrous DMSO stored in a Sure-Seal^{TM} bottle. Additionally, CDI powder was frequently examined for signs of caking and moisture absorption and replaced regularly.

As seen from figure 4.3.1.1, RagAC tends to migrate further down the gel after cross-linking than before. This behaviour is rather unusual, so the cross-linking product was compared to that obtained with BS3 and EDC. BS3 is also lysine-lysine reactive, and EDC cross-links lysines to aspartates, or lysines to glutamates. EDC + NHS-ester pair was also tested as EDC on its own ha poor efficiency and NHS-ester is known to enhance EDC's reactivity by improving the lifetime of the intermediate cross-link. CDI and BS3, likely due to shared reactivity towards lysine residues, had very similar effects on migration pattern of the cross-linked products. EDC cross-linking efficiency was very low, and although it was improved by addition of NHS, likely higher concentrations would be needed for the cross-linking to work well using this chemistry. It should be noted that, although signs of cross-linking can be observed on the Coomassie-stained gel as a high-molecular weight smear and smearing of the individual protein bands, most of the sample likely remains un-cross-linked as evidenced by sharp individual bands of RagA and Raptor.

The choice of a denaturing reagent is known to have an impact on digestion efficiency and sequence coverage. Therefore, to maximise the chances of a successful digestion and identification Ragulator-RagAC-Raptor complex cross-linked with CDI was denatured using sodium deoxycholate or 6M urea. Due to precipitation of the samples in 1% SDC upon heating, the precipitate was resuspended in 6M urea to complete the digestion. The supernatant was kept and analysed separately.

Cross-linked peptides tend to have higher molecular weight than their uncross-linked counterparts, so size exclusion chromatography was performed to enrich cross-linked peptides relative to the more abundant background of common peptides and boost identifications. On size exclusion profiles, sample digested with urea only produced the largest peptides, and the sample digested with SDC alone resulted in mid-range to short peptides fig 4.3.1.2. To maximise the chance of sequencing events for cross-linked peptides during mass spectrometry analysis, only precursors with charge state above, or equal to 3 were selected for sequencing.

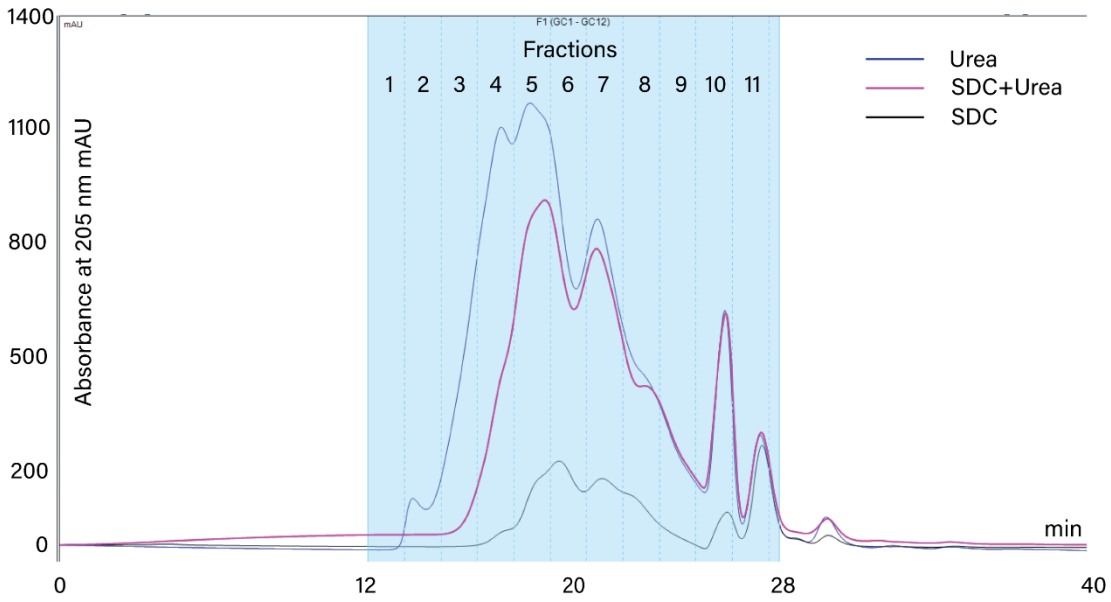

Figure 4.3.1.2 Size exclusion chromatography enrichment of cross-linked tryptic peptides.

The tandem MS data were analysed with both currently available software packages for MS-cleavable crosslinkers: Merox and XlinkX. For Merox, two major releases were available, so the data were analysed with both, using the same settings where possible. As shown in figure 4.3.1.3 A, all software versions and packages largely agreed on identification of the peptides and regions within the proteins that were involved in the cross-linking reaction (data shown here includes all unique cross-links regardless of the digestion method).

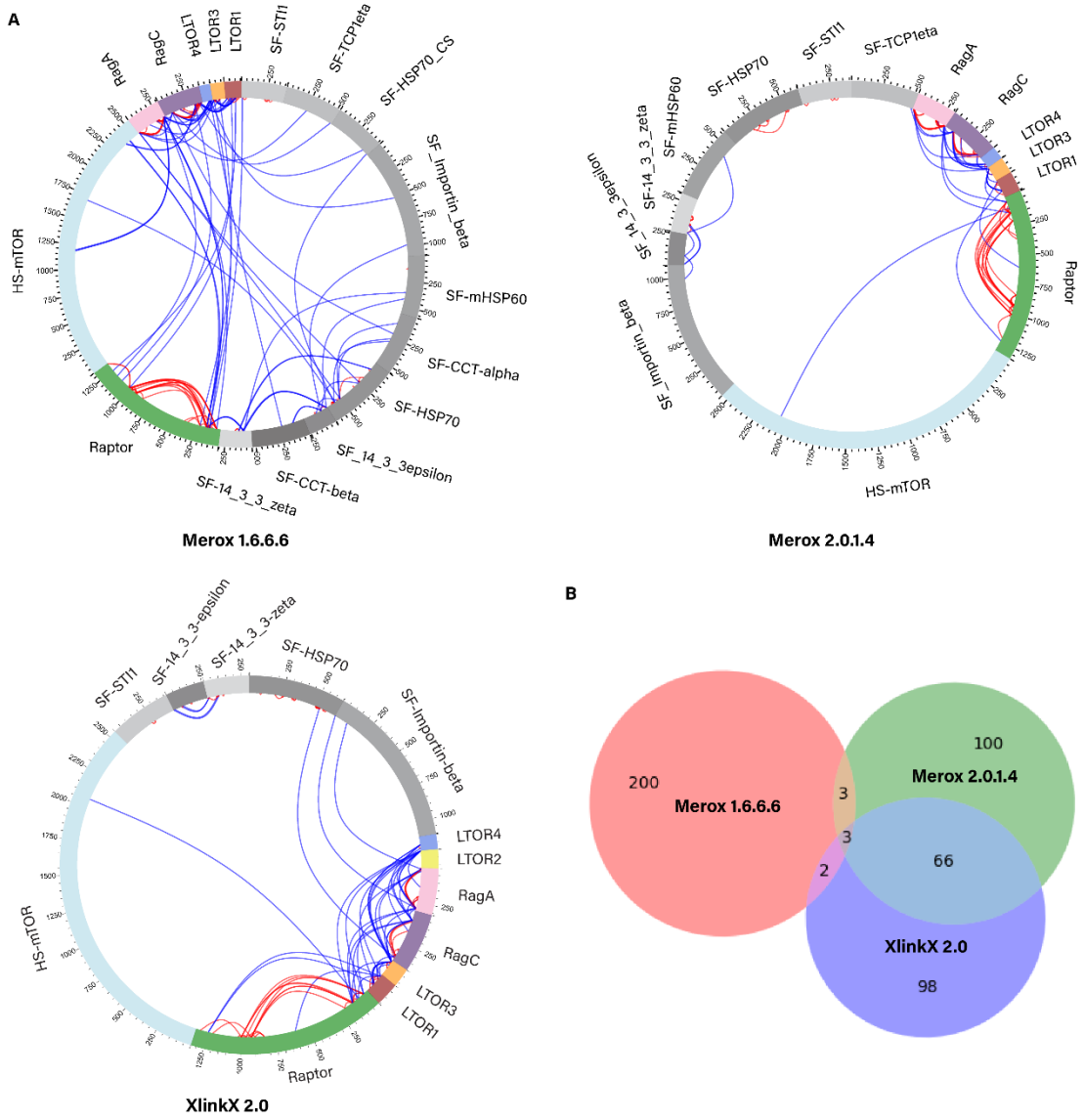

Figure 4.3.1.3 (A) Linear sequence xVis maps of cross-links identified with each software packages. (B) Venn diagram of unique cross-link position identifications per software package.

Merox 1.6.6.6 identified the highest number of cross-link identifications above the FDR cut-off (208, versus 172 for Merox 2.0.1.4, and 179 for XlinkX 2.0), all unique crosslinks identified can be found in appendix III. This increase in the overall cross-link identifications was at least in part due to an increased number of cross-links to and within the top 20 contaminating proteins from the original purifications. The identification of the reaction

site, however, proved more variable between different software, with less than 3% exact overlap between Merox 2.0.1.4 and the older Merox 1.6.6.6, and less than 50% overlap between Merox 2.0.1.4 and XlinkX (fig. 4.3.1.3 B). The difference between Merox versions is likely due to changes in the scoring algorithm, as the scores for the 6 common identifications between the Merox versions did not correlate at all (table 4.3.1.1). In contrast, Merox 2 scores and XlinkX scores for the common unique cross-link sites correlated surprisingly well, with Pearson correlation r=0.75 (fig. 4.3.1.4).

Table 4.3.1.1 Scores of the common identified cross-link sites in different software

Cross-link	*Merox 1.6.6.6*	*Merox 2.0.1.4*	*XlinkX 2.0*
RagC 79 - 79	123	174	
Raptor 132-142	112	57	113.99
Raptor 142-142	150	45	
RagC 361-361	144	222	
RagA 309-299	102	50	92.28
RagA385-361	102	17	

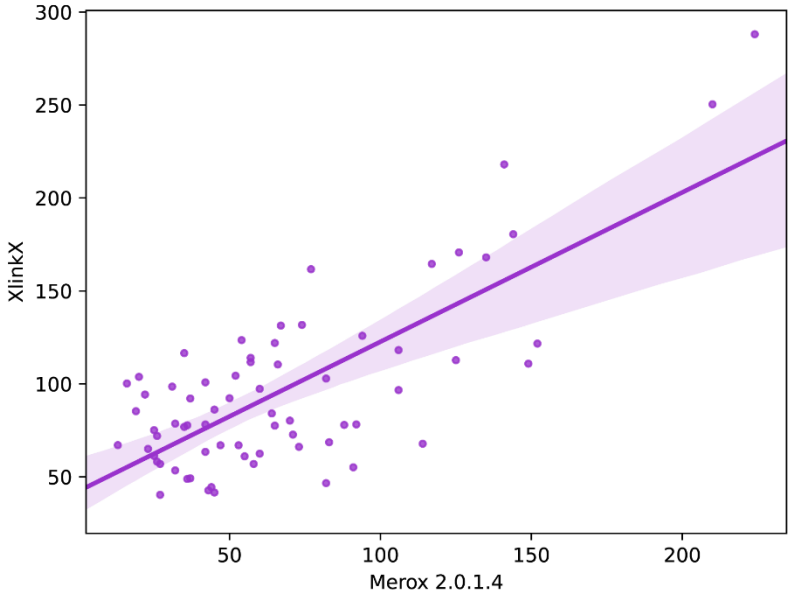

Figure 4.3.1.4 Linear correlation analysis on scores of unique cross-links identified with both XlinkX, and Merox 2.0.1.4 . Violet – line of best fit (Pearson r=0.75). Lighter shade denotes confidence of fit.

The distribution of cross-link identifications from different size exclusion fractions, and denaturation methods was similar for all the software tested (fig. 4.3.1.5 A, B). XLinkX also allowed to analyse the percentage of cross-link identifications found from spectra with

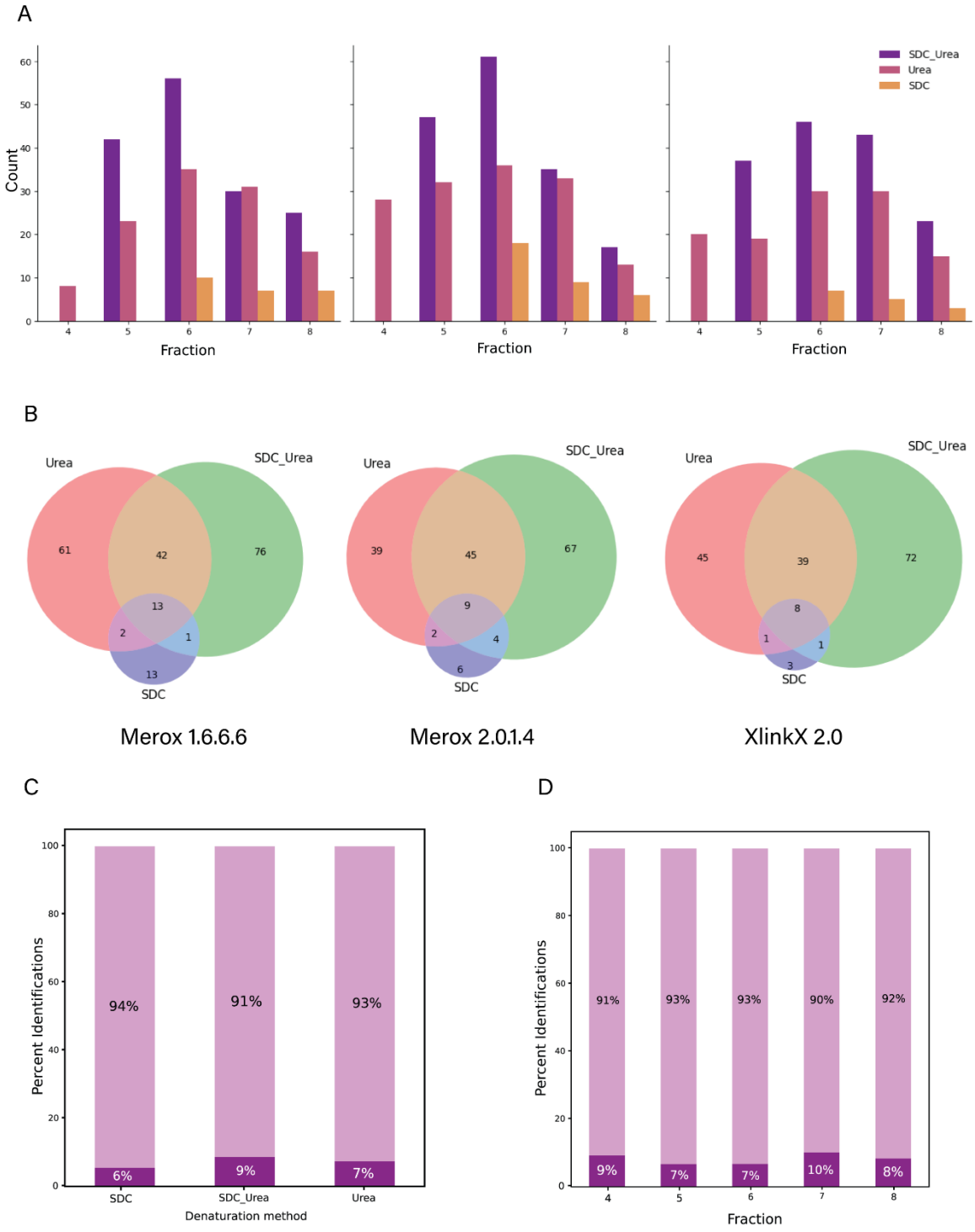

Figure 4.3.1.5 (A) Number of cross-links identified in each SEC fractions by digestion method (left-to-right: Merox 1.6.6.6, Merox 2.0.1.4, XlinkX 2.0). (B) Venn diagram of unique cross-links identified with each software by digestion method. (C, D) Per cent of spectra containing reporter ions that led to cross-link identification (by digestion method (C), by SEC fraction (D)).

identified CDI reporter ions (fig. 4.3.1.5 C, D). Only six to ten per cent of all spectra with CDI reporter ions led to a cross-link identification, without substantial differences between SEC fractions or denaturation methods.

To evaluate the quality of the cross-links identified through each of the software packages, the distances between Cα atoms of cross-linked residues in a published cryo-EM structure of the complex (PDB ID: 6U62) were calculated in PyMol. For each of the software packages, a histogram of distances was plotted (figure 4.3.1.6 A). Some of the distances could not be calculated, as the regions where they are located are not resolved in the structure, these were assigned a distance value of 0 by default. Up to 23% of total unique cross-links were in this category. With distances in the range of 6-145 Å, the distributions were similar for all software packages. However, Merox 1.6.6.6 had identified more cross-links in the distance range between 5-15 Å. Than either XlinkX or Merox 2.0.1.4 (26, 12 and 12 respectively). There was no correlation between scores and observed distances for either of the software packages (4.3.1.6 B).

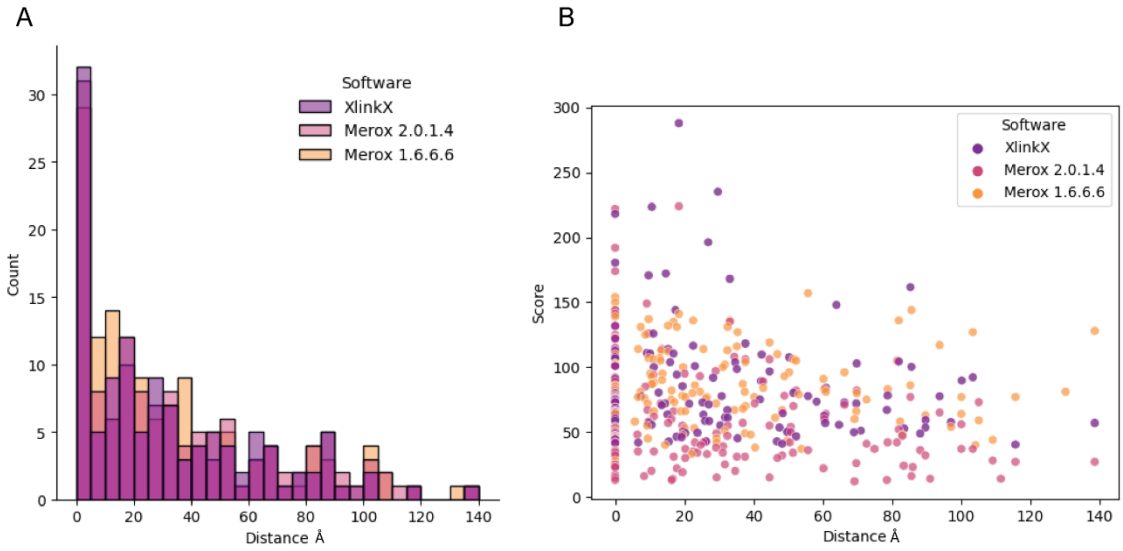

Figure 4.3.1.6 Cα-Cα distances between pairs of cross-linked residues identifies with each software. (B) Score vs. Distance for each of the cross-linked residue pairs

Of all 327 unique cross-links identified with the three programs, distance could not be calculated for 74. 14 of those were interprotein cross-links. The majority of these involved the flexible N- terminus of LAMTOR1, or C-terminus of RagA. 68 cross-links had a Cα distance below the previously reported 18 Å threshold[176]. Of those, 20 were interprotein links confirmed by the structure: RagC 361 - LAMTOR1 (55,60), and LAMTOR3 (10-12) -LAMTOR1 (103,104, 97, 96), RagA (295,299)-LAMTOR2(N-term).

While some of the longer distances could be explained by feasible, small-scare motions and rearrangements, others would be next to impossible to achieve without unfolding, based on the currently available structures. In figure 4.3.1.7, cross-links to ordered regions that were identified with both, Merox 2.0.1.4 and XlinkX are shown on the structure of the Ragulator-RagAC-Raptor complex used for distance calculations above. These

common cross-links follow the same distance distribution as all unique cross-links combined.

There are four kinds of cross-links: intraprotein, interprotein, proximal interprotein (where cross-linked residues are <20 amino acids apart in sequence and are usually located on adjacent peptides), and homeotypic cross-links (peptides and residues cross-linked to another molecule of themselves). While the score distributions between the four groups were comparable, the distance distribution of proximal cross-links was substantially smaller than that of intra- and inter-protein cross-links (fig. 4.3.1.8).

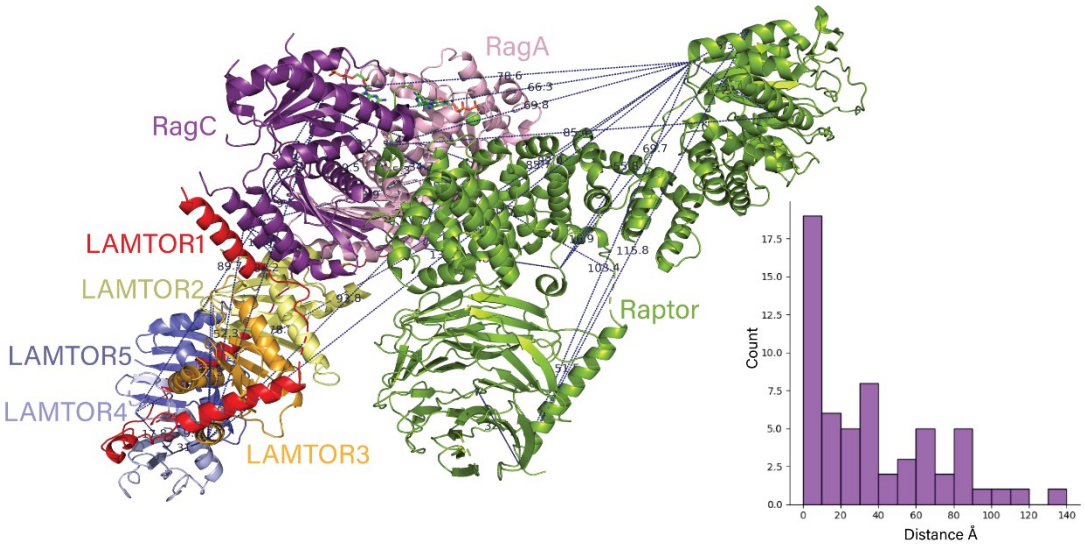

Figure 4.3.1.7 Calculated and visualised distances for cross-links identified with both Merox 2 and XlinkX 2. On the right is a distance histogram, where cross-links that cannot be plotted due to lack of structural information are assigned a value of zero and appear in the bottom 0-10 bin.

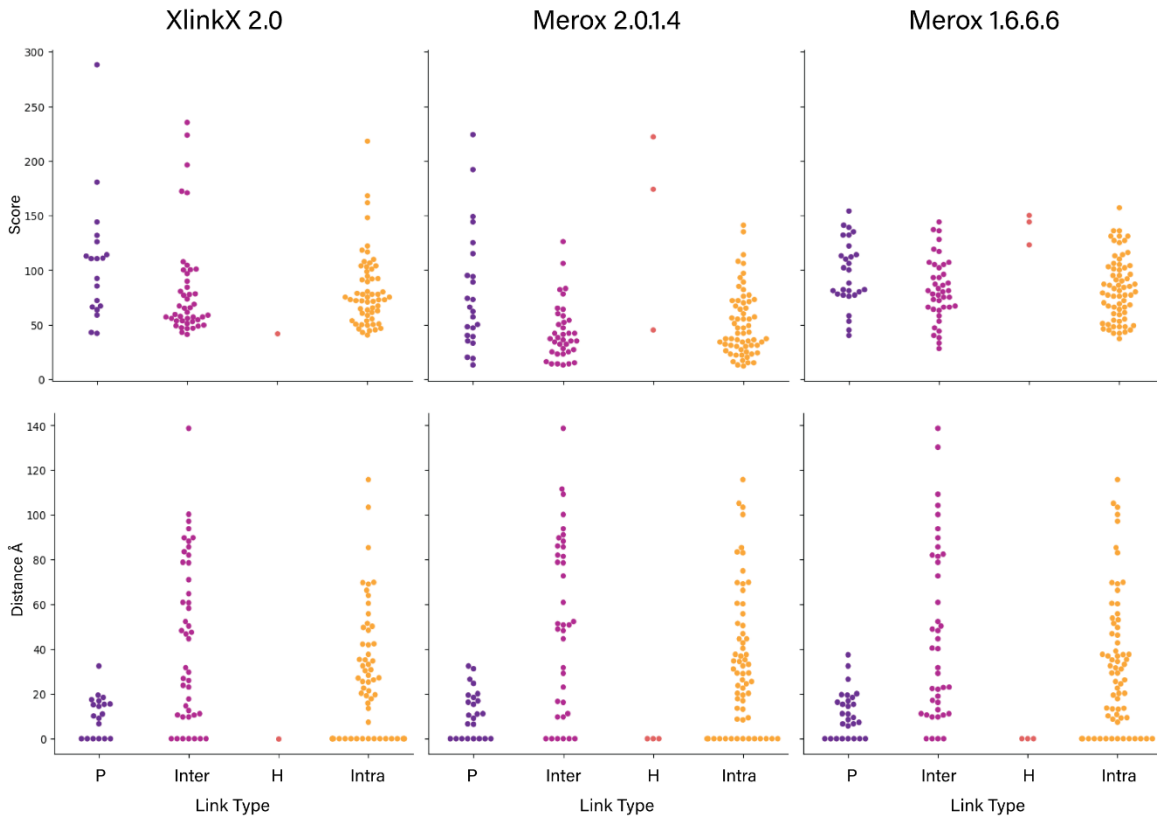

Figure 4.3.1.8 Swarm distribution plots of scores and distances calculated for each unique cross-link based on the output of each software packages. Proximal (fewer than 20 aa apart in sequence), Interprotein, Homeotypic (H) and Intraprotein cross-links were considered separately. Distances that could not be plotted were assigned a default value.

As the longest possible distance between 2 Cα in this case would be if the residues were on opposite sides of a 20-amino acid beta-sheet (70 Å). For comparison, two opposite ends of a 20-amino acid alpha-helix would result in a 30 Å Cα distance. As there are no such long beta sheets in the structure, and Cα distances in flexible loops would likely be somewhere between 30-70 Å at the longest, this kind of distribution is expected. As none of the proteins are known to form homodimers, we do not expect any homeotypic cross-links to form. However, one such link was identified with each of the software packages: RagC361-RagC361. The score of this cross-link was at the higher end of the score distribution in Merox packages, while in XlinkX it was one of the lower-scoring identifications, close to the score cut-off value. Homeotypic cross-links represent a special case for cross-link identification in cleavable peptides as only ions of one peptide contribute towards sequence information. In a rare case, cross-liking of the main peptide to a peptide with a very similar mass, but with a very poor fragmentation efficiency, could lead to a spectrum in which only the information for the first peptide is present, and this cross-link may be erroneously identified as homeotypic. Therefore, interpretation of such identifications should be done with caution.

Residues 38, 132, 142, 1008 on Raptor, 79 on RagC, 299 on RagA and 11 on LAMTOR3 were found to cross-link to multiple distinct regions within the complex. The majority of

the cross-links to these residues cannot be easily explained. All are located on very short tryptic peptides (5-8 residues) with several charged amino acids that are well fragmented in the MS.

Intriguingly, no cross-links were identified between the Raptor "claw" region and RagA or RagC, and at the main interface of RagA and Raptor. Instead, residues in that stretch of amino acids were found to be cross-linked to several residues in the main chain of Raptor, indicating that at least a proportion of Raptor was likely not bound to RagAC, and the flexible "claw" region has an alternate position somewhere between the N-terminal domain and the HEAT repeats.

4.3.2 Photo-leucine and photo-methionine impact cell growth and protein production

To evaluate cell proliferation and protein production in HEK293T cells when grown in limiting and photo-amino acid containing medium, GFP was used as a reporter gene. A vector containing 10-His-tagged $RagA^{WT}$, $RagC^{S75N}$ (GTP hydrolysis-incompetent mutant, $RagAC^{GTP}$) and GFP sequences separated by self-cleaving P2A and T2A sequences was generated by Gibson assembly and verified by sequencing. GFP-only vector was used as an expression control. Schematic view is shown in figure 4.3.2.1

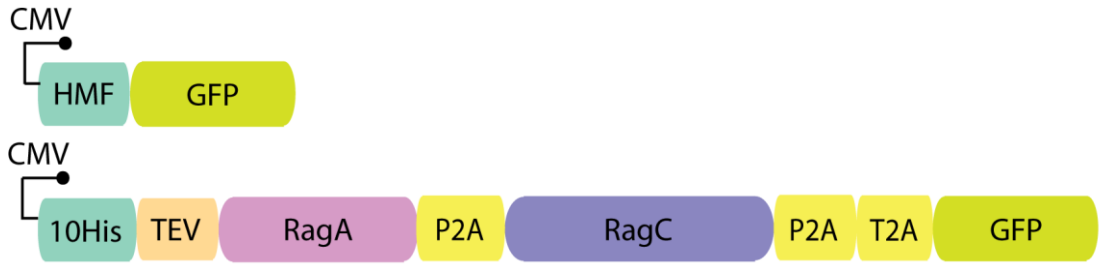

Figure 4.3.2.1 Schematic representation of the constructs used to assess photo-leucine and photo-methionine incorporation. HMF – His-Myc-FLAG tag combination, 10His – ten-histidine tag, TEV – TEV protease site, CMV – cytomegalovirus promoter and enhancer.

To minimise any serum effects and effects of leucine and methionine concentration in DMEM, the leucine- and methionine-free DMEM was supplemented with dialyzed foetal bovine serum and was supplemented with either photo- or natural L-leucine and L-Methionine. Three strategies of medium introduction were tested: Media exchange shortly before transfection, 12-hour incubation in each media prior to transfection, and a 12-hour leucine and methionine starvation period and introduction of media 2 hours prior to transfection. Cells were imaged at the centre of the well every 12 hours starting at transfection point, and cell growth was measured by confluence. Protein production was measured as a ratio of mean fluorescence per cell confluence area to account for differences in growth rates in different media, some detachment during washing, and any

seeding irregularities that may have led to non-uniform cell distribution in the well. As shown in figure 4.3.2.2, transfection and protein expression in full media have a slight inhibitory effect on cell proliferation, although it is more pronounced at later stages of expression and for the GFP-only construct. Additionally, the effect of L-photo-leucine and L-photo-methionine alone (PLM and LPM respectively), or in combination with each other (PLPM) was similar to that of leucine and methionine starvation (-LM). However, photo-leucine had a milder effect than photo-methionine overall in untransfected controls, regardless of the media introduction time. Surprisingly, when a combination of photo-methionine and photo-leucine was supplemented with 20 µM L-methionine, but not L-leucine, the growth curve was like those recorded in full media with and without transfection and independent of expression construct. As initial expression tests have shown a decrease in protein synthesis in photo-amino acid media when compared to full DMEM, a control condition was grown with 3.5 mM sodium valproate (SVA), a histone deacetylase inhibitor that is known to block G1 progression. While it had a mild effect on cells transfected with both GFP only and RagAC-GFP construct, it did not result in growth inhibition to the same extent as photo-methionine and photo-leucine. SVA had little to no effect on the growth of untransfected control cells in the timeframe of the experiment, except for the cells grown for an additional 12h in leucine- and methionine-free media, where mild inhibition was observed after 36h.

A time-course of GFP expression in these cells could also be monitored through fluorescence imaging by dividing the total fluorescence by cell area. Due to the large scale of variations in GFP expression depending on the construct and expression media, it was necessary to use three different exposure settings throughout the experiment. The data were combined in post-processing in a way that, for each point, the data with the best signal-to-noise ratio and less than 5% saturated pixels (less than 1% in most cases). They were then normalised to the values obtained with the shortest exposure time assuming linear relationship between integration time and fluorescence AU count. Compared to the growth curve data, the GFP expression per area was less variable between replicates and clear trends could be observed (fig 4.3.2.3).

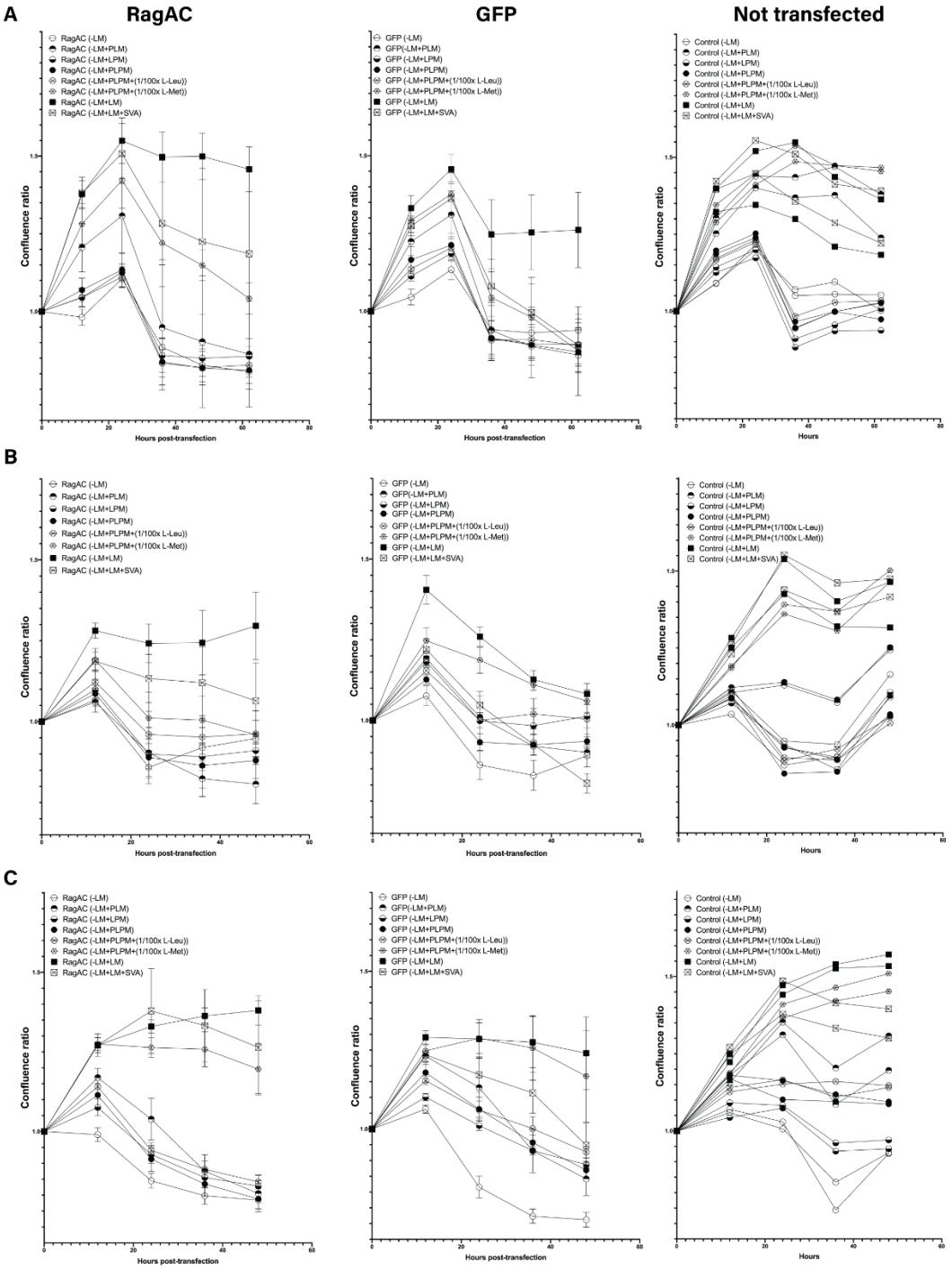

Figure 4.3.2.2 Growth of HEK293T cells after transient transfection with RagAC-GFP (left), GFP (middle) and untransfected control (right) in photo-amino acid media. (A) Transfection 2 h post-media change. (B) 12h pre-incubation with selected media prior to transfection. (C) 12h leucine and methionine starvation prior to media change and transfection. For RagAC and GFP n=5 (error bars show 95 % CI), untransfected control n=2 (here both replicates are plotted).

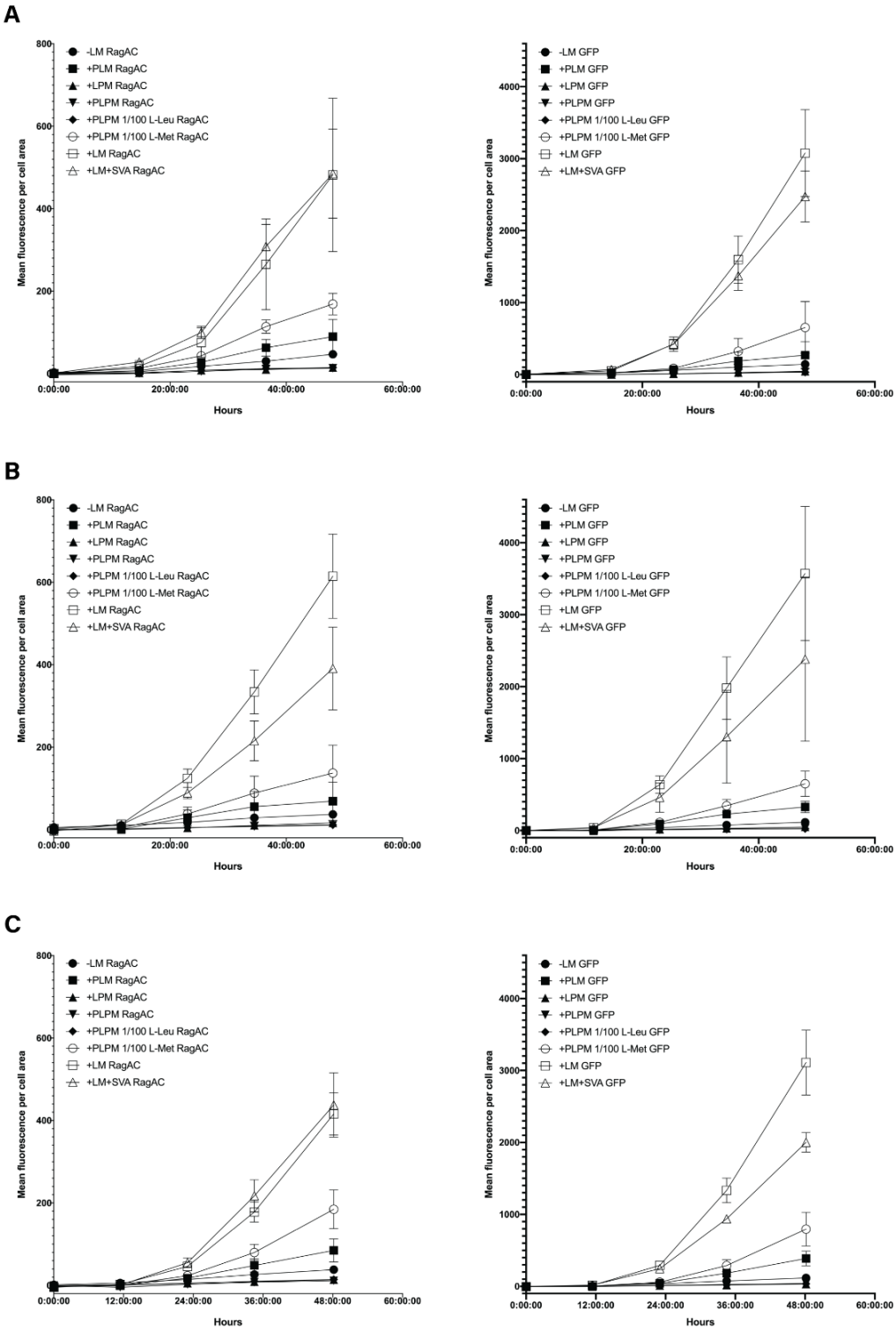

Figure 4.3.2.3 Time-course of mean GFP fluorescence per cell area in photo-amino acid containing media. (A) Transfection 2h post-media change. (B) 12h pre-incubation in expression media prior to transfection. (C) 12h leucine and methionine starvation prior to media change and transfection. n=5, error bars show 95% CI.

Independent of transfection strategy and expression time, the expression of RagAC-GFP was approximately 15% of the GFP-only expression in fully supplemented medium (+LM)

(fig. 4.3.2.4 A-C). This is almost twice as high as the expression level expected based purely on the size of the construct (3-fold increase in DNA size used for transfection and 4-fold increase in the transcript size would lead to an expectation of RagAC yield of around 8% of that of GFP).

To evaluate the effect of photo-amino acid media on each of the constructs, I calculated the proportion of GFP fluorescence per cell area in photo-amino acid media relative to that in fully supplemented media at 48-hour timepoint (fig. 4.3.2.4 D-F).

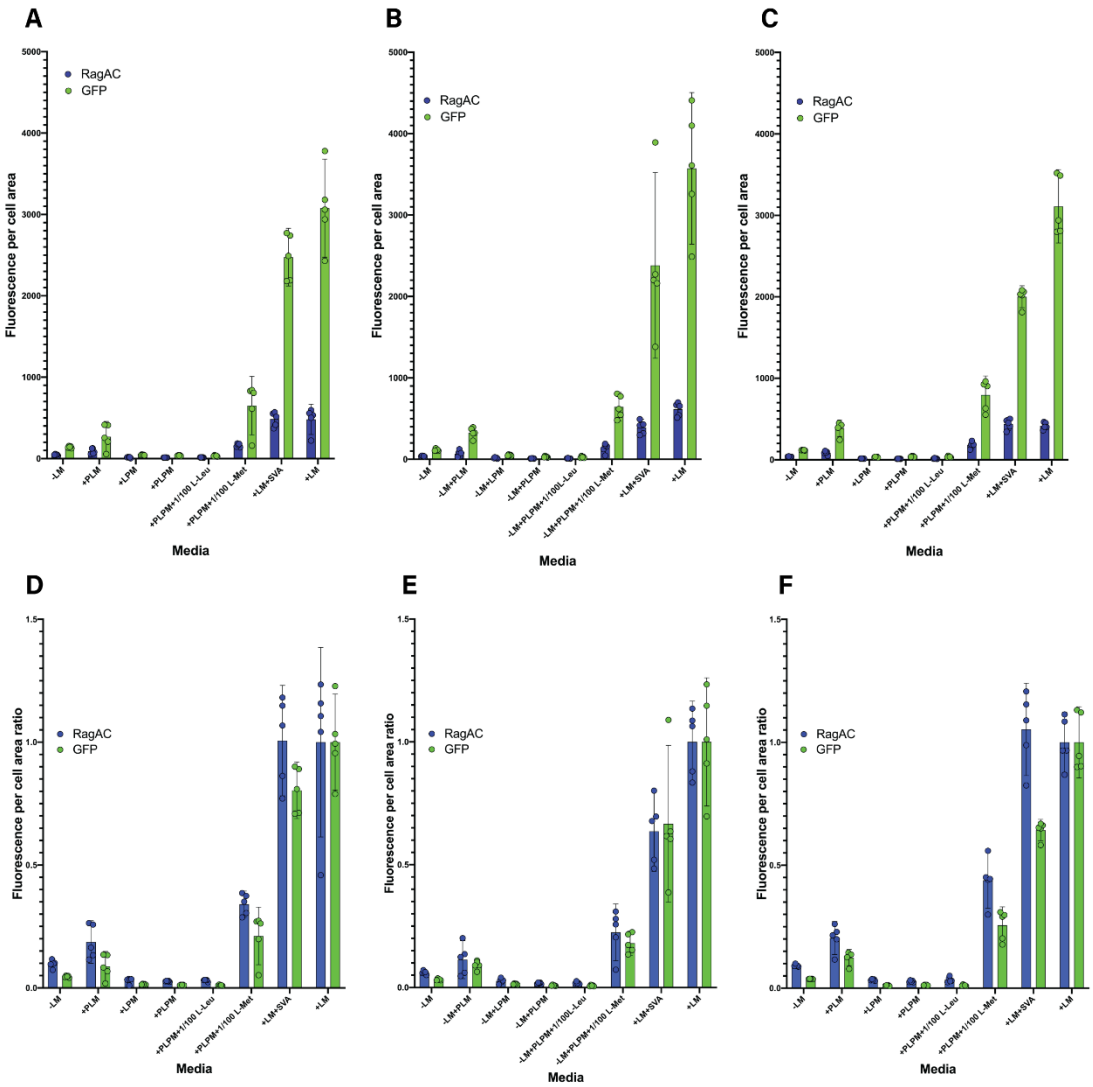

Figure 4.3.2.4 (A-C) Mean GFP expression in different media adjusted by cell area at 48h post-transfection. (D-F) Relative GFP expression in photo-amino acid containing media relative to that in fully supplemented media at 48h post-transfection. (A, D) Transfection 2h post-media change. (B, E) Transfection 12h post-media change. (C, F) Transfection 12h post leucine and methionine starvation. n=5, error bars show 95% CI

Expression in leucine- and methionine-free medium was approximately 5-10% of the yield in fully supplemented media, expression in photo-leucine only containing medium was 5-25% of that in full medium, while expression in photo-leucine and photo-methionine media supplemented with 20 µM L-methionine was 20-45% of that in full medium. Conditions without photo-methionine supplementation consistently yielded less than 5% of the expression in full media regardless of expression construct and transfection strategy.

Two-way ANOVA test showed that there was a small, but significant interaction between expression construct and media (up to 2.8% of the overall variance attributed to expression construct) in all experiments, except for those in which the cells were pre-incubated in expression media for 12 hours before transfection. As this prevented interpretation of the overall effect of different media, post-hoc analysis for GFP-only and RagAC-GFP was performed separately. The proportion of GFP expression in the photo-amino acid media relative to +LM media was compared to that in leucine- and methionine-free media (-LM), and fully supplemented media (+LM). GFP expression in all photo-amino acid containing media was significantly lower than that in +LM medium (table 4.3.2.1).

Except for the photo-leucine, photo-methionine and 20 µM L-methionine media (PLPM+L-Met), GFP expression in all other photo-amino acid media was not significantly higher than that in -LM medium. All other conditions with photo-methionine (without L-methionine supplementation) consistently showed the lowest expression, indicating that photo-methionine has a specific negative impact on protein synthesis, and that 1:100 ratio of L-methionine is enough to counteract this inhibition. While GFP expression in PLM media was not significantly different to -LM media, it was still higher than in other conditions. The difference between PLM media and PLPM+L-Met media was only significant ($p<0.0001$) when a 12h starvation step prior to transfection was included. This indicates that L-methionine at 2 mM concentration might also be inhibitory to protein synthesis under conditions tested, and this effect is remediated by addition of photo-methionine. Sodium valproate also had a significant negative effect on expression of the GFP-only construct in experiments with a 12-hour pre-incubation in expression media, and a 12-hour amino acid starvation steps. For $RagAC^{GTP}$, the effect of SVA was only significant when the cells were pre-incubated in the expression medium for 12 hours before transfection. This difference in response contributed to the interaction observed in the ANOVA test.

*Table 4.3.2.1 Post hoc T-test analysis of GFP expression at 48h post-transfection for RagAC-GFP and GFP-only in different media adjusted for multiple tests. P-values: *<0.05, **<0.01, ***<0.001, ****<0.0001*

	Transfection 2h post-media change				12h pre-incubation with the media				12h leucine and methionine starvation			
	RagAC-GFP		GFP		RagAC-GFP		GFP		RagAC-GFP		GFP	
	>-LM P-value *	<+LM P-value *	>-LM P-value *	<+LM P-value *	>-LM P-value *	<+LM P-value *	>-LM P-value *	<+LM P-value *	>-LM P-value *	<+LM P-value *	>-LM P-value *	<+LM P-value *
-LM		****		****		****		****		****		****
+PLM	ns	****	ns	****	ns	****	ns	****	ns	****	ns	****
+LPM	ns	****	ns	****	ns	****	ns	****	ns	****	ns	****
+PLPM	ns	****	ns	****	ns	****	ns	****	ns	****	ns	****
+PLPM+ L-Leucine	ns	****	ns	****	ns	****	ns	****	ns	****	ns	****
+PLPM+ L-Methionine	*	***	ns	****	ns	****	ns	****	****	****	****	****
+LM	****		****		****		****		****		****	
+LM+SVA	****	ns	****	ns	****	****	****	****	****	ns	****	****

There were no other significant differences in effect of the photo-amino acid media tested between RagAC^{GTP}-GFP and GFP-only constructs. Transfection at two hours post media change, and after a 12- hour leucine and methionine starvation resulted in higher proportion of expression in photo-amino acid medium relative to wild-type than 12h pre-incubation in expression media. This was mostly due to an increase in protein expression in full media when the cells were grown for an additional 12 hours before transfection, possibly increasing the number of cells at transfection time relative to the other media.

4.3.3 Quantification of photo-amino acid incorporation

For initial testing of photo-amino acid incorporation GFP-only construct was used. To estimate the percentage of photo-leucine and photo-methionine incorporation into peptides of the protein of interest (POI), they were purified with Ni-NTA magnetic beads and digested with trypsin prior to MS/MS analysis. The overall percentage of incorporation of photo-amino acids was calculated based on the sum of intensity of all POI peptides that contained photo-leucine (or photo-methionine) over the sum of

intensity of all leucine- (methionine-) containing peptides of the protein of interest. The possibility of unintended photo-amino acid activation and reaction with water was also investigated by including the mass shift for this modification into the analysis.

Previous experiments showed that supplementation with L-amino acids increased protein yield in photo-amino acid medium. To estimate the effect additional L-amino acids have on photo-amino acid incorporation HMF-tagged GFP was expressed in presence of L-photo-leucine and L-photo-methionine with, and without supplementation with 25-400 µM of L-leucine and L-methionine (1:80-1:10 ratio).

Approximately 6% of leucines and up to 37% of methionines in GFP were substituted with their photo-reactive counterparts. However, introduction of L-methionine during expression decreased incorporation of photo-methionine ~4-fold already at 25 µM (fig 4.3.3.1). The already lower incorporation of photo-leucine did not experience similar reductions. In fact, the incorporation rate increased with addition of L-leucine and L-methionine. This was in part due to improved protein yield leading to identifications of more peptides with photo-leucine in supplemented conditions. As photo-methionine inclusion in the expression media dramatically reduced protein yields, and introduction of L-methionine to remedy this resulted in poor incorporation rates, all further experiments were done with photo-leucine only.

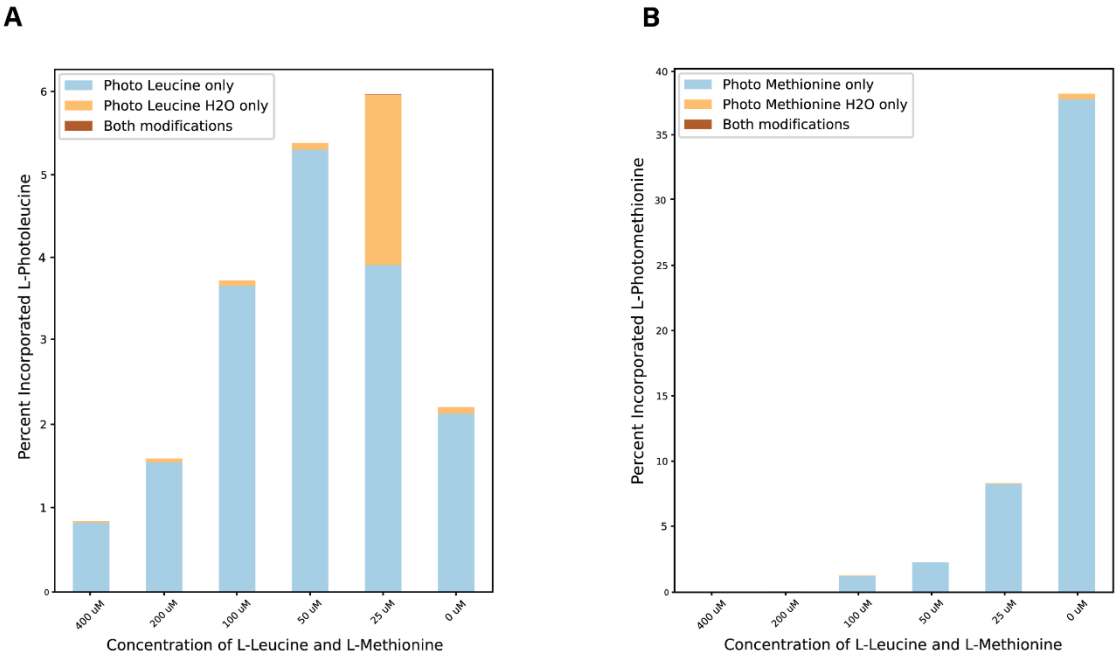

Figure 4.3.3.1 Supplementation of Photo-leucine medium with L-Leucine and L-Methionine in concentration range between 0-400 µM. (A) Percent incorporation of Photo-Leucine. (B) Percent incorporation of Photo-Methionine. Values reported here are average across all leucine or methionine containing peptides within each sample in same experiment.

To test the influence of expression time on photo-leucine incorporation, samples transfected with the GFP-only construct were harvested at 30, 44, and 54 hours post transfection. The percentage of photo-leucine incorporation was calculated as described above. After the initial search in MaxQuant 1.6.7.0 it became clear that there were extreme differences in identification rates of photo-leucine containing peptides even between biological replicates, as shown in fig. 4.3.3.2 A. Technical replicates also showed high variation. After manual inspection of the ion chromatograms from MS/MS it became clear that the peptides that were possibly present in all the samples were not being identified by the search algorithm either due to the quality of their MS2 spectrum, or simply because the MS2 spectrum was never acquired. When a “match between runs” search option (carries over identifications for precursor ions whose m/z, charge and elution time are within the specified window of the peptide identified in another run of the same experiment) was enabled in MaxQuant, the variation between the samples was reduced. Additionally, the incorporation rates calculated increased slightly overall, without substantially raising identifications in the samples that were grown in absence of photo-amino acids, suggesting that the use of this function is warranted and results in minimal false carry-over (fig 4.3.3.2). Full list of GFP peptides identified in the second search can be found in appendix IV.

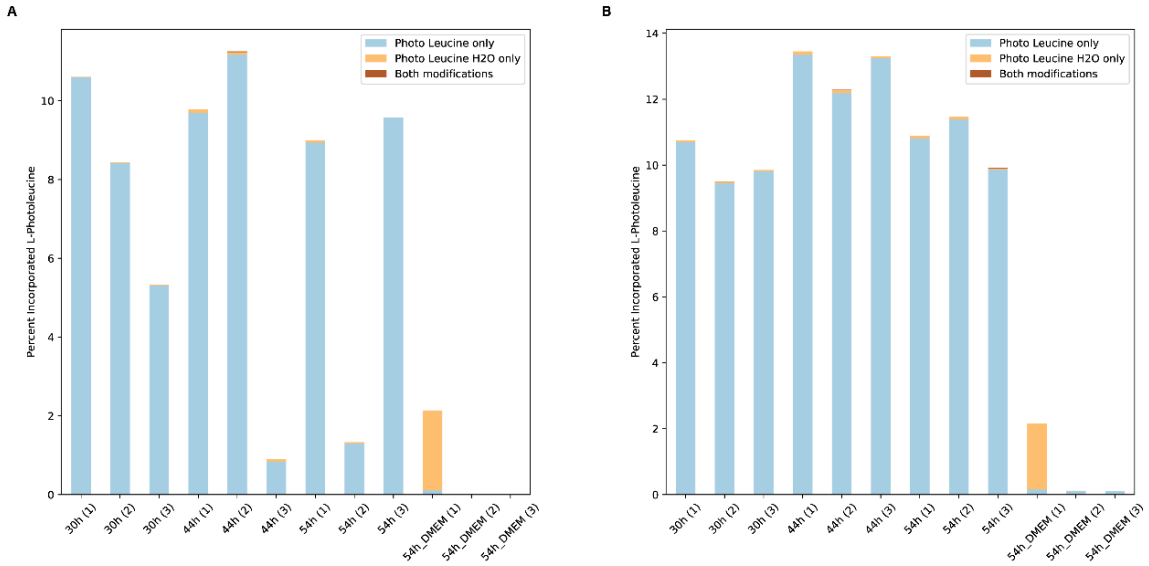

Figure 4.3.3.2 Effect of expression time on photo-leucine incorporation. (A) First MaxQuant search results with "match between runs" disabled. (B) Second search with MaxQuant with enabled "match between runs" option.

The percentage of incorporation of water-reacted photo-leucine identified in one of the photo-amino acid-free medium samples came from a single, presumably falsely identified, peptide with very high intensity that was never identified in any other sample. As evident from figure 4.3.3.2 B, expression time did not appreciably affect rates of photo-leucine incorporation. Since after at the 24–36-hour point, the cells expressing protein start to

lyse (fig 4.3.2.2), for future experiments the cells were harvested before or around the 40-hour mark to balance protein production with ease of harvesting intact cells.

Identification of photo-leucine sites was not consistent between the experiments, and the incorporation varied for each leucine position (0.1-10%) and some photo-leucine containing peptides were only found in one-or-two experiments. However, across all the optimisation experiments with more than 25 samples in total, only 7 leucines could not be observed as either leucine or photo-leucine, likely due to the properties of tryptic peptides on which they are located (fig. 4.3.3.3). For all other leucines, photo-leucine was identified as a variable modification in at least one sample. Generally, peptides with higher overall intensity were more consistently identified as having been substituted with photo-leucine. There was no clear relationship between codon usage, or leucine's position in the structure and efficiency of photo-leucine incorporation.

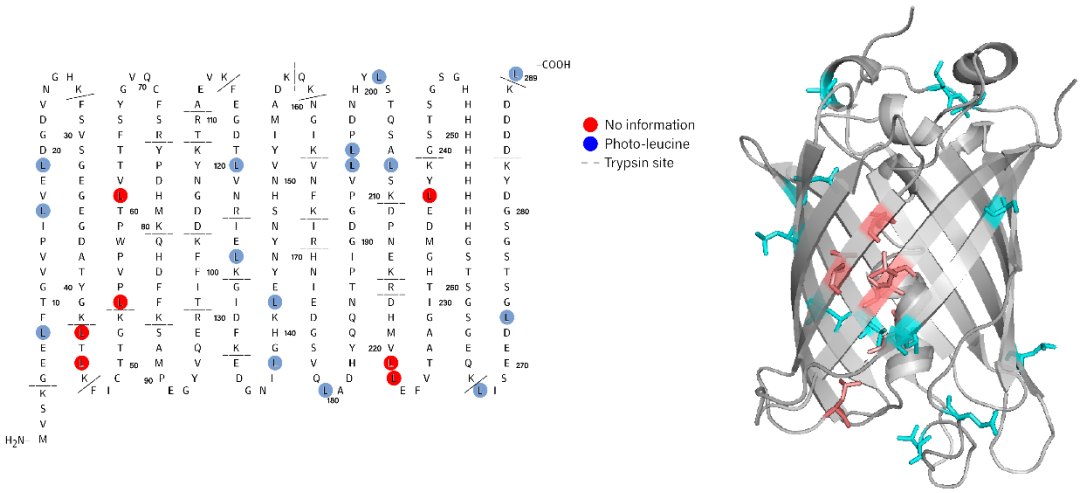

Figure 4.3.3.3 Photo-leucine incorporation at different positions in GFP. (Left) Position in sequence with trypsin binding sites highlighted. (Right) Position of photo-leucines in the known structure, with photo-leucines identified at least once coloured cyan, and ones that are never identified are coloured in salmon.

When expressed in photo-amino acid media, only 8 of the most abundant peptides of RagA and RagC could be found with a photo-leucine substitution, intensity information was available for 6 of them. As the overall incorporation rates could be underestimated due to absence of relevant identifications for some of the leucine-containing peptides, analysis of incorporation rates per peptide was deemed appropriate. Incorporation rates varied between the peptides from <1% to 4.4%. This was much lower than the observed incorporation rates for the peptides of GFP where average incorporation was ~14% with variability between 0.5-20%. However, incorporation of photo-leucine at these positions was mostly consistent between biological replicates (table 4.3.4.3).

4.3.4 UV-crosslinking of proteins with photo-leucine

As it would be difficult to measure any effects of cross-linking in GFP, RagAC^{GTP} construct was used as the model protein to evaluate cross-linking efficiency of photo-leucine. The proteins were expressed in photo-leucine containing medium while scaling up the expression volume to compensate for the lower yields. Similar to the experiments with GFP, RagAC was enriched by pull-down using magnetic Ni-NTA beads. Due to high volume of lysate and inability to use detergent for the bead washing steps, beads adhered to the tubes and resulted in non-reproducible elution (fig.4.3.4.1). Additionally, ~80 kDa band was always identified in the pull-downs and boiled beads, suggesting that the separation of the RagA and RagC at the P2A site during synthesis may be inefficient. Absence of the 110 kDa band, however, suggests that GFP was completely separated during synthesis.

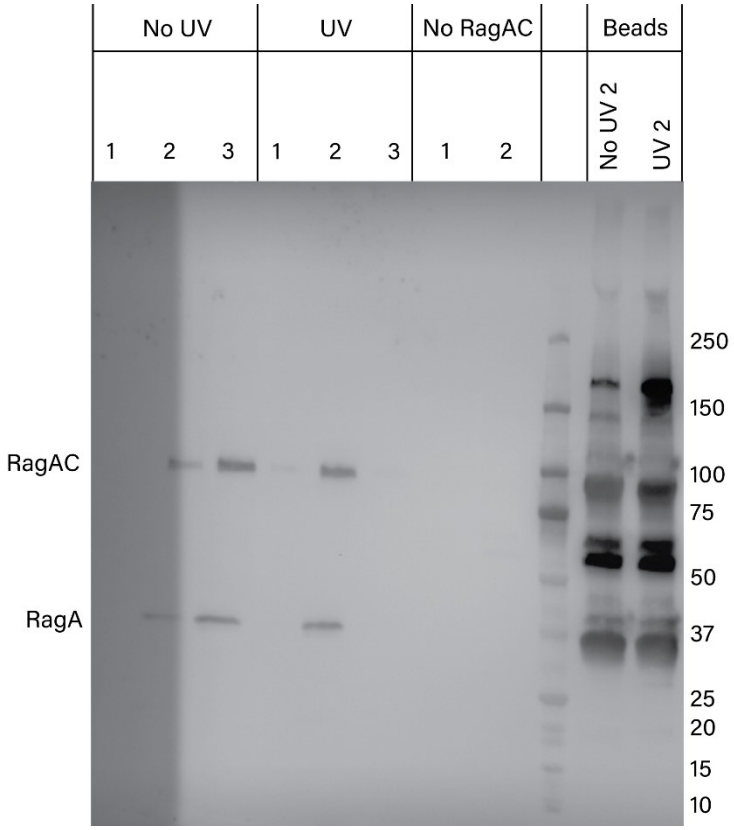

Figure 4.3.4.1 Anti-His Western blot analysis of RagAC pull-down elution and remaining protein on the beads. Bands corresponding to RagA and RagAC dimer are highlighted. MW ladder key is labelled on the right.

To overcome the irreproducibility of elution steps, an on-bead trypsin digestion protocol was used instead. Tandem mass spectrometry analysis of the peptides generated this way had identified ~3000 proteins across the samples, with RagAC consistently being in the Top18 proteins of the samples in which it was expressed. In contrast, ~1200-1800 proteins were identified in previous GFP pull-down experiments, with GFP having the highest MS1 intensity in all cases. 1479 proteins were identified in all samples, including those that did not express RagAC. 688 proteins were only identified in samples expressing RagAC, including GFP as a contaminant. Of these, 65 were found in at least two, and 16 in all three of the treated and untreated samples (table 4.3.4.1). 519 proteins were found only in UV

treated samples, of which 9 were found in all 3 replicates (table 4.3.4.2). Only 2 proteins were exclusively found in all untreated samples (TFa102, progranulin).

Table 4.3.4.1 Unique proteins only identified in pull-downs from cells expressing RagAC-GFP construct in photo-leucine media irrespective of UV treatment.

Gene	*Protein Name*	*Pathway*
COX6A1	Cytochrome c oxidase 6A1	Oxidative phosphorylation
GFP	GFP	NA (Co-expressed)
ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	Ketone body metabolism
HSD17B12	Very-long-chain 3 oxoacyl-CoA-reductase	Lipid biosynthesis
PPM1F	Protein phosphatase 1F	CaM-II deactivation, apoptosis
DDOST	Dolichyl-diphospho-oligosaccharide—protein glycosyltransferase 48kDa	Protein glycosylation
SLC25A	TCA transport protein	Mitochondrial CoA transport
NAA15	N-alpha-acetyltransferase 15	Protein stabilisation, apoptosis suppression
ZNF622	Zinc finger protein 622	Apoptosis
NCLN	Niacalin	Protein stabilisation
SFXN1	Sideroflexin	L-serine transport
TMCO1	Calcium load-activated calcium channel	Calcium homeostasis
MT-CO2	Cytochrome c oxidase subunit 2	Oxidative phosphorylation
HSPE1	10 kDa Heat shock protein	Protein folding
LARS	Leucine-t-RNA-ligase	Leucine-t-RNA synthesis, mTOR signalling
UQCRC1	Cytochrome b-c1 complex subunit 1	Oxidative phosphorylation

Table 4.3.4.2 Proteins identified exclusively in UV-treated samples.

Gene	*Protein Name*	*Pathway*
PGAM2	Phosphoglycerate mutase	Glycolysis, gluconeogenesis
MTPN	Myotrophin	Muscle development
PSAT1	Phosphoserine aminotransferase	Serine biosynthesis
PRDX1	Peroxyredoxin	Oxidative stress regulation
ACO1	Aconitate dehydratase	TCA cycle, iron uptake
PFN1	Profilin-1	Actin polimerisation
MPST	3-mercaptopyruvate sulfur stransferase	Cysteine catabolism
AKR1B1	Aldose reductase	Glycerolipid metabolism
PEBP1	Phosphatidylethanolamine-binding protein 1	MAPK, GSK-3 signalling

Overall, for proteins identified in more than one sample, there was good correlation of intensity between all biological replicates (average Pearson's r= 0.98) and between UV-treated and untreated samples (average Pearson's r=0.95). Combined, these results suggest that while the experiment was reproducible, and the UV-treatment largely did not affect the proteome and pull-down results. The washing procedure was also not stringent enough to yield a clean sample for further cross-linking analysis or evaluation of binding partners. While the intensity of the majority of the co-purified proteins was not significantly different between the treated and untreated samples, 66 proteins were enriched in the UV-treated sample (Q-value (P-value adjusted for multiple testing) < 0.025 and >2-fold change). None of these were known interaction partners of RagAC. Top 10 enriched proteins by Q-value are listed in table 4.3.4.3. Full list of identified proteins and SafeQuant output are provided in appendices VI.I and V.II.

Table 4.3.4.3 Top 10 proteins enriched in UV-treated sample

Gene	*Protein name*	*Pathway*	*Q-value*	*Log2 ratio*
NTMT1	N-terminal Xaa-Pro-Lys N-methyltransferase 1	DNA-damage repair	7.16E-05	4.61
PGK1	Phosphoglycerate kinase 1	Glucose metabolism	7.16E-05	3.27
ACTB	Actin, cytoplasmic 1	Cytoskeleton	0.00012	2.60
ANP32B	Acidic leucine-rich nuclear phosphoprotein 32 family member B	Histone binding	0.00042	2.64
ACTG1	Actin, cytoplasmic 2	Cytoskeleton	0.00042	1.79
ENO1	Alpha-enolase	Glycolysis	0.00045	1.93
FH	Fumarate hydratase, mitochondrial	TCA cycle	0.00046	2.47
PPIA	Peptidyl-prolyl cis-trans isomerase A	Protein folding	0.00077	1.77
PLS3	Plastin 3	Cytoskeleton	0.00078	1.86
MDH2	Malate Dehydrogenase 2	TCA cycle	0.00125	2.23

The extent of photo-activation upon UV treatment was also analysed. For four out of seven of the RagAC peptides which contained photo-leucine in non-UV-treated samples, a form of the UV-activated and modified leucine (reacted with water or self-inserted) was also identified in the UV treated sample. In all cases there was a notable reduction in abundance of the intact photo-leucine in the UV-treated samples if it was identified at all (table 4.3.4.4, MaxQuant search results in appendix V).

Table 4.3.4.4 Photo-leucine containing peptides in RagA and RagC

Leucine position	*No UV 1 (%)*	*No UV 2 (%)*	*No UV 3 (%)*	*UV 1 (%)*	*UV 2 (%)*	*UV 3 (%)*
A:110 Photo-Leu	2.72	2.26	3.00	-	-	-
Photo-Leu-UV				-	0.33	0.33
Photo-Leu_H2O				-	-	-
A:39 Photo-Leu	4.45	3.60	3.16	0.28	0.31	0.19
Photo-Leu-UV				1.27	0.99	1.55
Photo-Leu_H2O				0.71.	1.36	1.60
A:151 Photo-Leu	-	4.41	-	-	-	-
Photo-Leu-UV						
Photo-Leu_H2O						
C:140 Photo-Leu	0.77	0.97	0.84	-	0.25	0.34
Photo-Leu-UV						
Photo-Leu_H2O						
C:351 Photo-Leu	3.48	3.59	3.43	-	-	-
Photo-Leu-UV						
Photo-Leu_H2O						
C:91 Photo-Leu	1.36	1.04	1.36	-	-	-
Photo-Leu-UV				-	-	-
Photo-Leu_H2O				-	0.41	-
C:183 Photo-Leu	2.21	2.51	2.35	-	0.001	-
Photo-Leu-UV				-	0.27	0.49
Photo-Leu_H2O				0.29	0.19	0.13

The UV-activation of the photo-leucine under the conditions tested was good, as evidenced by scarcity (on average <15%) of the remaining diazirine group and at least partial conversion of some residues to quenched products in most UV-treated samples. The highest photo-leucine retention rate of around 38% in two out of three samples was observed for RagC L140. This residue is located at the core of the protein. As none of the UV-modifications on photo-leucine were present in the non-UV-treated sample, there is no evidence of premature quenching or reactivity of incorporated photo-leucines with normal handling of the samples.

At some positions (RagA39, RagC91), the sum of percentages of intact photo-leucine, activated photo-leucine that reacted with water, and activated self-reacted photo-leucine in UV-treated samples is similar to that of the intact photo-leucine in untreated samples. In others, however, this sum remained but a fraction of the photo-leucine identified in untreated samples (RagA110, 151, RagC140, 351, 91, 183), indicating that the rest of the photo leucine may have been cross-linked to either another solvent molecule or a nearby amino-acid. To explain the difference in the proportion of UV-modified photo-leucines in UV treated samples between peptides, I analysed the location of these leucines in the protein structure, fig. 4.3.4.2.

Residues 39 of RagA and 91 of RagC are the most solvent accessible residues and are not known to be involved in interprotein interactions to mTORC1 components or Ragulator.

While RagA110, RagC140, and RagC183 are the least solvent accessible. Interestingly, residue 351 in RagC, which is located at the interface with its binding partners LAMTOR2 and LAMTOR1, should be solvent accessible in the dimer alone. Since no reaction products or intact photo-leucine were observed in the UV-treated samples, it is possible that the photo-leucine at this position may be fully cross-linked to its binding partner. An increase in 160-170 kDa (size of RagAC-Ragulator complex, or 2x RagAC) product after UV treatment support this hypothesis. However, LAMTOR1 and LAMTOR5 were the only Ragulator complex proteins identified in the samples. While LAMTOR1 was identified in all the samples (including no RagAC control), LAMTOR5 was only identified in one of the UV-treated samples. To investigate whether L351 on RagC participates in cross-linking, a more stringent wash protocol would be required to remove most contaminants and allow for further cross-linking mass spectrometry analysis.

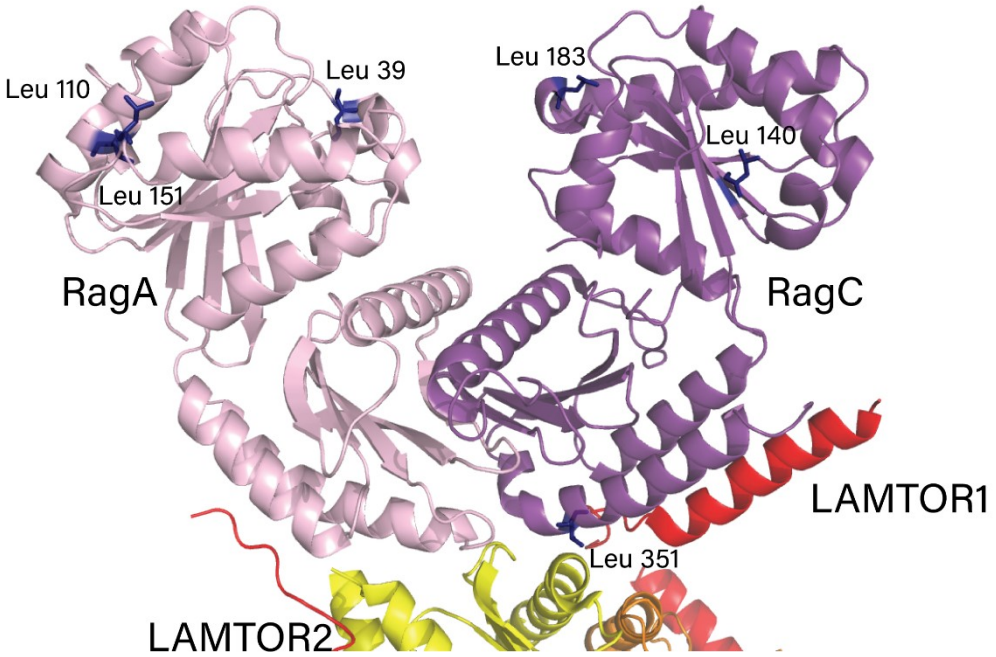

Figure 4.3.4.2 Leucines substituted with photo-leucine mapped onto the Raptor-RagAC^{GTP}-Ragulator structure (PDB ID: 6U62) Residue L91 Is not shown as it is not resolved in this structure.

5 Discussion

5.1 The effects of R42H mutation on hexokinase 2

As predicted by the loss of R42-D272 salt bridge hexokinase 2 R42H was overall less stable than the wild-type protein, but at high concentrations and in presence of G-6-P and glucose had a similar structure and stability to wild-type. A major difference in flex angle of the helix α_{13} cannot be explained by differences in ligand binding or dimerization behaviour differences between the wild-type and the mutant. However, it is worth noting, that hexokinase 1, which has an inactive N-terminal domain despite having the same active site conformation in complex with glucose-6-phosphate and glucose as hexokinase 2, has the most similar conformation of this helix. Molecular dynamics simulations and mutation studies have shown that interactions between the small subdomain of the N-terminal domain and helix α_{13} are required for catalytic activity of the N-terminus of hexokinase 2[205]. The absence of R468, one of the key players in these interactions is thought to be responsible for the inactivity of the N-terminal domain of hexokinase 1[205]. It is possible that the destabilisation caused by R42H mutation may also influence these dynamics and contribute to the decrease in activity and differences in linker helix conformation. Unfortunately, it is impossible to infer this relationship from the crystal structure directly as it is unclear whether this difference in dynamics influences the position and flex of the connecting helix in the structure. The effects of crystallization conditions and crystal packing on this angle also cannot be ruled out at this time.

The dimerization of the R42H mutant in absence of ligands was impaired relative to that of the wild-type enzyme. However, as the concentrations required for stabilisation of the wild-type protein without ligands were also relatively high, and, fully stabilised HK2 could not be obtained even at highest available concentrations, this is unlikely to have significant physiological implications.

ATP binding did not have a substantial stabilising effect on the wild-type protein at concentrations below 5 mM, while the shape of the denaturation curve changed for the mutant with evidence of stabilisation at concentrations around the previously reported K_d for the N-terminal domain (0.45 mM). This indicates ATP binding, but does not provide insight into differences to the wild-type enzyme, which shows the “stabilised” behaviour even in absence of ligands. At concentrations above 5 mM, the effect of ATP was comparable to the wild-type and there is no evidence to suspect any differences in affinity of the mutant for ATP.

Like ATP binding, glucose and glucose-6-phosphate binding did not appreciably affect the overall Tm of either the wild-type or the mutant enzymes at concentrations below the previously determined K_d of 0.13 mM, but also improved stability of the region in the R42H mutant that was more susceptible to thermal denaturation. However, unlike with ATP

binding, further increases in concentration of either glucose of glucose-6-phosphate led to substantial increases in T_m of both, the mutant and the wild-type enzymes.

As, unlike ATP, both glucose and glucose-6-phosphates are known to induce large conformational changes and subsequent dimerization of the hexokinase, it is likely that this T_m increase comes from dimerization events rather than simple stabilisation from ligand binding, like that observed for the less-stable part of the R42H mutant. Consistent with the hypothesis that this change comes from dimerization of the hexokinase, the effect of the mutation is less apparent with increasing concentrations of glucose and G-6-P to a point that the observed Tm is nearly identical for the wild-type and the mutant at G-6-P concentrations above 10 mM and at Tm above 49˚ C. It should be noted that a Tm plateau could not be reached even at 160 mM glucose and G-6-P, while glucose concentrations above 40 mM also had a stabilising effect on lysozyme, which is not known to bind glucose. Therefore, at concentrations higher than 40 mM, the dimerization effect is likely compounded with non-specific stabilising effects of glucose and G-6-P on protein stability.

As concentration-dependence of glucose binding on overall Tm of both, R42H mutant and wild-type HK2, is unchanged, there is no evidence for overall reduced affinity of HK2 R42H for glucose, or glucose-6-phosphate *in vitro.* Similarly, any effect of the mutation on dimerization capability upon glucose and G-6-P binding is unlikely. However, its influence on ATP binding, or differential effects on substrate binding or catalysis in the N-terminus, which can contribute to reduced activity cannot be ruled out as the effect of ATP and glucose-binding at lower concentrations could not be observed for wild-type protein with this method.

The similarities of the crystal structure and apparent overall affinities to its ligands and product of the R42H mutant to the wild-type hexokinase 2 are akin to those between the hexokinase 1 and hexokinase 2. Yet, the N-terminal domain of hexokinase is inactive through the changes in dynamics that prevent catalysis, and it is possible that the R42H mutation has a similar effect on the N-terminus. To further understand the differential effects of this mutation on activity of hexokinase 2 N- and C-terminal domains it would be useful to test the activity of a double R42H/S603A mutant. Additionally, measuring ligand affinity of truncated N-terminal (plus helix α_{13}) mutant, with isothermal titration calorimetry would help to confirm if ligand affinity of the N-terminus is affected separately from stabilising events. To confirm the hypothesis that the dramatic increase in stability at higher concentrations of glucose and glucose-6-phosphate is due to dimerization, stability analysis of a quadruple R42H/E280A/R283A/G284Y mutant would be useful. Otherwise, since the protein concentration required for dimerization in presence of glucose and glucose-6-phosphate is low, it may be possible to confirm this through SEC-MALS, AUC or mass photometry analysis.

5.2 RagAC-mTORC1 binding

When loaded with native nucleotides, all tested RagAC mutant combinations reduced the elution time of Raptor on the size exclusion chromatography runs in comparison to Raptor only elution. As expected, the $RagC^{GDP}$ mutant and $RagA^{GTP}$-$RagC^{GDP}$ mutant combinations, had a higher effect on Raptor peak elution time than the wild-type combination, suggesting a stronger interaction of those to variants with Raptor outright. Since the shift in elution peak was not characteristic of the expected nearly 2-fold size increase, and both SEC-MALS and AUC failed to record such a mass increase, it is possible that the observed interaction is transient. Although in this experiment the individual nucleotide state of each Rag protomer is not precisely determined, it is possible to deduce that after the nucleotide exchange with 4:1 GTP:GDP ratio both mutant combinations would likely have a $RagA^{GTP}$-$RagC^{GDP}$ conformation. In the case with WT-WT protein, it is a little more complicated. However, based on the knowledge that RagA is likely an inactive GTPase without a GAP, and that RagC may spontaneously hydrolyse GTP, especially when RagA is primed with GTP, it is possible that the introduction of GTP during nucleotide exchange improves Raptor binding through this mechanism[76]. This is also supported by the fact that the exchange with GTPγS did not have a similar effect on the WT-WT RagAC. Again, while we did see a decrease in elution time of Raptor when combined with RagAC and Ragulator that indicated specific binding, the degree of the change was not in line with the expected molecular weight for this complex either. As both methods used to calculate the size of the complex rely on separation, which may influence the binding equilibrium, we have decided not to further purify the fully assembled Ragulator-RagAC-mTORC1 complex prior to grid preparation and imaging. This strategy has provided satisfactory results and led to atomic-resolution reconstruction of the full lysosomal mTORC1 complex.

These results have highlighted that the interactions involved in formation of this complex are likely transient, and at least these defined components do not form a stable high affinity complex resistant to commonly used separation techniques. This combined with the fact that the complex is now well-understood make it an ideal test-case scenario for application development of auxiliary techniques for complexes involved in cell signalling networks such as cross-linking mass spectrometry.

5.3 Ultra-short and zero-length cross-linking for structural studies

In this work I have split my efforts between development of an *in vitro* chemical cross-linking mass spectrometry pipeline using 1,1-carbonyldiimidazole as the main cross linker, and an *in vivo* pipeline using previously developed method of metabolic labelling with photo-leucine and photo-methionine for evaluation of transient complexes in signalling networks.

5.3.1 *In vitro* cross-linking with CDI

Similar to what has been shown before for chemical cross-linkers, introduction of CDI in high excess at physiological conditions to an unstable complex such as Ragulator-RagAC-mTORC1 resulted in formation of a cross-linked product. Albeit the efficiency of this higher-molecular weight product generation was low. Most of the protein present in solution remained uncrosslinked to any neighbours. The efficiency of internal cross-links, as evidenced by the pattern of cross-linked RagC band was comparable to what could be achieved with an NHS-based cross-linker BS3 (bis(sulfosuccinimidyl)suberate). Curiously, unlike RagA, at higher cross-linker:protein concentrations the band corresponding to RagC became very faint indicating either extensive internal or inter-protein cross-link generation. This was in line with our expectation that a cross-link between lysines and serines or threonines in the Raptor claw region within residues 917-936 and lysines or serines within RagC G-domain may be formed. EDC and EDC+NHS cross-linking, on the other hand, failed to result in any visible cross-link formation on the gel despite the presence of proximal lysines and aspartates at the binding site of RagAC and Raptor.

Despite all efforts to maximise cross-link identification through employing different digestion strategies and size-exclusion fractionation-based enrichment of cross-linked peptides, identification of cross-link sites within these samples remained a challenge. While size exclusion chromatography of tryptic peptides yielded similar cross-linked peptide elution profiles to those described previously for this technique, it did not provide a comparable level of separation from common (unlinked) peptides to strong cation exchange chromatography SCX[214, 215]. Nonetheless, size exclusion chromatography proved useful in sample simplification through fractionation.

The combined effects of a neutral chaotropic reagent urea and an anionic detergent sodium deoxycholate, have proven useful in reducing the overall size of the peptides obtained after tryptic digestion and thus providing an alternative, larger dataset of identifiable peptides that has a 40-50% overlap with the dataset obtained with urea only. Along with employment of alternative proteases, this could be an additional layer of information to improve the number and quality of cross-link identifications that could be used as modelling restraints in the future[214, 216].

The main challenge in cross-link site identification turned out to be software choice and data quality. After initial struggles of identification of reliable cross-links, I have tested three available software packages for cleavable cross-linkers: two different versions of the Merox software and XLinkX, each with a different scoring algorithm[163, 164, 166, 217, 218]. Remarkably, the two versions of Merox (1.6.6.6 and 2.0.1.4) had very little overlap between the exact cross-link matches and scores, while Merox 2 and XlinkX had a substantial number of identical cross-link identifications whose individual match scores also correlated well.

Upon closer examination of search, the overlap between identified peptide pairs was high across the three software packages, indicating a potential issue with cross-link position identification. There are two causes for this: differences in scoring algorithms, and poor overall quality of sequencing data. The differences in scoring algorithms may mean that the theoretical spectra of the same peptides cross-linked at different positions may become the top scoring match for the same spectrum. These differences are also responsible for differential treatment of special cases like homeotypic links, in which the same peptide is cross-linked to itself, and produces only one strong set of reporter ions and fragment ions of only one peptide on the MS2 level, while the calculated parent ion mass corresponds to double of the peptide plus linker. If this happens, it is likely that the MS2 data is incomplete and, therefore, does not allow to unambiguously identify the cross-link site, or even the peptides involved. This could be either due to poor fragmentation producing spectra with too few abundant ions, or simply low sample amount and low abundance of cross-linked peptides, which results in low signal-to-noise ratios in MS2 spectra preventing confident matches.

Since CDI is a cleavable cross-linker, and the MS2 spectra which are obtained from cross-linked peptides contain specific reporter ions, it was easy to assess the quality of data and data analysis by calculating the efficiency of cross-linked peptide identification. From the data analysis with XlinkX pipeline it became apparent that <10% of the MS2 spectra which contained reporter ions could be matched to a theoretical cross-linked spectrum regardless of the digestion procedure or SEC fraction. These findings highlight that even though we can identify a large number of cross-links in these samples, there are improvements to be made in data collection and data-analysis strategies to allow us to access all information from such studies. As outlined above, efficient fragmentation and high signal-to-noise ratios can help push the efficiency of peptide identifications. In this study we have opted for the faster data acquisition mode and only recorded MS2 scans after a single collision CID fragmentation event at a set energy level. However, it was shown that highly charged peptides cross-linked with non-cleavable linkers may benefit from differential fragmentation strategies that include electron transfer dissociation (ETD) and higher energy collisional dissociation (HCD) fragmentation to achieve the best quality spectra[219]. Similar approaches could also be useful for cleavable cross-linkers like CDI, so such data dependent acquisition methods could be developed in the future. Use of a combination of low energy collision induced dissociation (CID) MS2 with higher energy HCD MS3 data acquisition, or a stepped energy HCD MS2 fragmentation strategies as previously described for DSSO linker could be beneficial as well[166, 220].

To improve sample quality and cross-linked peptide abundance within the analysed sample, it could perhaps be useful to combine the benefits of the SCX-SEC 2D chromatography together with alternative protease digestion in varied conditions. An obvious caveat to this kind of method development would be an exponential increase in sample requirement. As both, SCX and SEC, can generate a large number of fractions, the

amount of starting material for analysis would need to allow for at least 15-25-fold sample dilution through this type of fractionation (3-5 SCX fractions by 5 SEC fractions) plus any losses through initial digestion a final desalting. If the SCX fractions of cross-linked peptides account for ~1/3 of the total sample, the minimum amount of protein to be processed per sample to obtain 10 μg peptide per fraction would approach 0.5 mg. Considering that a cross-linking experiment would rarely consist of one sample, but would rather comprise several conditions with replicates, alternative proteases and digestion conditions, the sample requirement for the set of experiments could quickly grow to tens of milligrams. This is on par, or more material than what is typically required for high resolution structural studies including optimisations. As producing and purifying such complexes, or even just their constituents in such amounts can be challenging, and requires a lot of resources, other approaches should also be explored.

The issue for cross-link identification in purified protein samples is not only with sample complexity, but also with the dynamic range. Even after enrichment procedures, the most abundant ions that enter the mass spectrometer are those of common ions. This severely limits how much total sample can be injected at once and therefore, limits how much cross-linked peptide can be analysed at once regardless of the total amount available for analysis. The sample requirement for purified complexes can be reduced, and the need for inefficient off-line sample fractionation and enrichment steps can be removed, or at least minimised by applying state-of-the-art in-line gas chromatography techniques. High filed asymmetric waveform ion mobility spectrometry (FAIMS) can allow precise low charge-state ion filtering directly prior to injection into the mass spectrometer. This could replace the strong cation exchange chromatography separation steps and increase the amount of sample that can be injected per SEC or C18 fraction without overloading the system with common peptide ions, while simplifying the sample preparation workflow[221, 222]. Due to low abundance of cross-linked peptide ions in the samples and therefore low signal, the cross-linking experiments can also benefit from improved noise estimation techniques, which would allow for more data points to be used for spectrum matching, further improving the accuracy and sensitivity of cross-link spectrum identification[223].

Despite all the limitations discussed above, I could identify over 60 cross-links within the Ragulator-RagAC-Raptor complex at 1% FDR with at least two software packages. However, when mapped to the cryo-EM and crystal structures of the complex, the Cα-Cα distances of identified cross-links did not neatly fall in line with previously reported maximum values of 18, or even 25Å[175, 176]. The distances could not be calculated for 23% of all unique cross-links, and an additional 21% of the links had calculated distances below the reported 18Å, leaving most of the links in violation of expected maximum distances. While some of the violating cross-links could be explained by potential minor flexibility of the proteins or lack of protein-protein interactions, many would be impossible to satisfy without completely unfolding the proteins. As CDI is rapidly degraded

in aqueous buffer (evidenced by lack of cross-linking activity when the stock is dissolved in buffer prior to reaction or poor mixing) and violating cross-links are equally represented among intra- and inter-protein cross-links, it is unlikely that this kind of the cross-linking happens after protein denaturation. Another possible explanation could be false identification of such cross-links. However, many of these were identified independently by two different algorithms irrespective of score cut-off. In fact, there was no clear correlation between cross-link scores and Cα distances, indicating that in this instance, the algorithms could not differentiate between the likely true cross-links and those between distal residues. Although some short peptides appeared to cross-link to multiple other peptides on different proteins, it would not be possible to satisfy many of these cross-links by any means. Despite all efforts I was unable to find universal criteria for cross-link-spectrum-match exclusion that targeted low molecular weight peptides and peptides with longer distances over more plausible ones with longer, more robust peptides for identification. It is possible that at least some of these long cross-links are real and come from alternative conformations and arrangements. Although care was taken to avoid aggregation and premature unfolding, this cannot be completely ruled out either, further highlighting the need for sample homogeneity for successful interpretation of results.

5.3.2 In vivo cross-linking mass spectrometry for photo-amino acid labelled proteins

Unlike chemical cross-linking *in vitro*, incorporating cross-linkable groups into proteins during synthesis would allow for reduced requirement of prior knowledge about the protein, complex and binding. However, incorporation of unnatural amino acids may have unpredictable effects of protein expression and stability, especially when the incorporation is proteome-wide and not targeted to a specific protein or location(s) within proteins. When protein production in photo amino acid containing medium was assessed in HEK 293T cells under the conditions of transient transfection, protein production and cell proliferation were significantly affected. This is at odds with previously reported apparent tolerance for photo-amino acids[190]. However, in that study the authors measured overall protein production and cell growth and did not measure the effect of photo-amino acids on specific protein production after transient transfection. In line with the results of that study the effect of photo-amino acids on cell proliferation was mild in the first 24 hours of transient expression. The negative effect of photo-amino acid medium only became significant at later time timepoints. Majority of this effect came from photo-methionine inclusion. However, while photo-leucine alone also produced some effect on cell proliferation, inclusion of both, photo methionine and photo-leucine when supplemented with L-methionine, but not L-leucine at a 1:100 ratio improved cell proliferation and survival to levels not significantly different to that in full medium. This suggests that photo-methionine may not be able to stimulate metabolic processes in the same way L-methionine would.

One of the possible mechanisms of this reduction in cell proliferation may be through inhibition of the mTORC1 axis. Due to the missing sulphur atom, L-photo-methionine may not be converted to S-adenosylmethionine, which signals methionine sufficiency to mTORC1 complex[43]. S-adenosyl methionine is also an important substrate of histone and DNA methylases and a precursor to glutathione and another important methionine metabolite S-adenosylhomocysteine that participates in folate and sulphur metabolism. A disturbance in the ratio of these two compounds, as well as the processes they regulate may contribute to observed slowdown in cell proliferation and reduced survival after prolonged incubation. Additionally, SAM is thought to be required for stress response activation[224].

In line with these observations and knowledge, we saw a decrease in protein production in photo-amino acid medium regardless of expression construct used, especially when photo-methionine is included. As with cell proliferation, we could rescue the protein production by addition of 1:100 ratio of L-methionine, therefore it is likely that the photo-methionine induced reduction in protein synthesis to methionine starvation levels is likely due to disruption of SAM-mediated signalling. Therefore, supplementation with L-methionine should always be considered when higher yields are desired. It is also worth mentioning that for immediate transfection (with or without a starvation step) photo-amino acids (and leucine- and methionine- starvation) were better tolerated when the 'active' RagAC dimer mutant was expressed instead of the GFP-only construct. The effect disappeared when the cells were pre-incubated with photo-amino acids prior to transfection. Although the effect was not statistically significant, this trend is in line with the hypothesis that photo-amino acids may influence the state of mTORC1 signalling through the amino acid sensing pathway. Further studies would be required to explore this.

Notably, as with cell proliferation, inclusion of photo-methionine in the medium actually mildly improved protein production relative to photo-leucine only samples irrespective of transfection and starvation modifications to the protocol. While the difference between those two conditions was not significant, it was enough to consistently result it significantly better protein expression than in starvation medium, while photo-leucine only samples did not provide significantly higher protein expression. This was puzzling as the photo-leucine only samples contained the same amount of L-methionine as full media, and 100x more methionine than medium supplemented with both photo-amino acids.

L-methionine is typically one of the least abundant amino acids in cell culture media used at 4-fold lower concentration compared to most other amino acids, and 10-fold less than what was used in this work and in the available photo-methionine publications. However, in the original trial experiments, I have also expressed these proteins in standard DMEM medium and did not note the difference in expression level compared to the supplemented media used here. The only difference between the photo-leucine only medium and double photo-amino acid medium supplemented with L-methionine is the

amount of methionine available for S-adenosylmethionine conversion, as well as its ratio to the methionine available for protein synthesis. Methionine toxicity has been reported for human cancer cells (albeit at much higher concentrations), as well as with system-wide effects due to metabolic reprogramming of liver cells, although the mechanisms of this toxicity are poorly understood[225-230]. It is possible that accumulation of homocysteine is responsible for the toxic effect which may be mediated via p53 axis[225, 226, 231]. HEK293T cells are also tumorigenic with a wildtype p53, and the energy requirement of cells overexpressing protein may be similar to that of cancer cells. It is possible that together with additional stress from potential misfolding due to photo-amino acid incorporation, similar cytotoxicity pathway could be involved in HEK293T cells here. Further proteomic studies would be required to understand the full extent of the impact of photo-amino acid incorporation on the health of the cell and signalling status.

Nevertheless, irrespective of the toxicity mechanism, it may be wise to optimise L-methionine concentration further to improve cell survival and maximise protein production when working with photo-amino acids. Since the metabolic requirements are different between cell types, it would likely be helpful to do this for each new cell type.

Since identification of photo-leucine and photo-methionine containing peptide was not possible for every peptide, the obtained incorporation values may not reflect the accurate overall level of incorporation in the protein of interest. Instead, they provide an estimate, while allowing to obtain some position-specific information to evaluate the uniformity of incorporation along the protein. Indeed, for both test cases the photo-amino acids were not distributed equally between the amino acids in the sequence. This did not depend on codon usage, but rather reflected the location of the residue within the protein. Buried leucines were less often found replaced by a photo-leucine in comparison with leucines on the surface, suggesting that, at least at certain positions, photo-amino acids may interfere with protein folding. This finding is also in line with the reduced apparent yields, as misfolded proteins are typically targeted for degradation and, in case of GFP do not produce any fluorescence. While it is unclear how much this contributes to the yield reduction compared to the effect of photo-amino acids on cell signalling, both mechanisms are likely involved.

The observed rates of UV activation were in line with previously published values, again with some variations according to residue position. Buried residues are more difficult to activate with UV compared to the surface residues. While not generally a concern, this could become an issue when targeting stable and large buried protein-protein interfaces.

Due to low protein yields and relatively low photo-leucine incorporation rates, the simplest pull-down protocol was chosen to minimise sample loss and maximise identifications. Likewise, on-bead digestion was performed due to inefficient and inconsistent elution with imidazole, particularly of cross-linked samples. This yielded reproducible samples. However, as a result, the background of identifications was high,

with few proteins significantly enriched in cross-linked samples compared to untreated samples, and a relatively low fold-change. None of the identified enriched proteins were known RagAC interactors. Although most are involved in nutrient metabolism, signalling pathways and stress response. As no denaturation step was used in pull-down, it is possible that many of the true interactors with hydrophobic surfaces had remained bound to RagAC during purification even without cross-linking, and some of the lower abundant interaction partners fell under identification threshold in untreated, and at least some of the treated samples.

To allow for more stringent selection of cross-linked peptides, on bead digestion could be performed with a more specific resin (antibody based, to avoid competition effects with interacting proteins) and used with a more stringent wash protocol (in presence of denaturant and/or detergent, sodium deoxycholate being the perfect candidate due to ease of integration into MS workflow). However, to be used for cross-linking studies, this would require substantial upscaling of the experiment to account for low protein production (3-5x), photo-leucine incorporation (25-50x) and incomplete UV activation (1.2x). Generally, one well of the 6-well plate containing c. $1x10^6$ cells should be enough for interaction detection under normal conditions. In the worst-case calculation $6x10^8$ cells would be required for analysis. This is the equivalent of 30 large 15 cm dishes of adherent cell culture in ~1 L of photo amino acid medium per sample. This is 30 times more than what was used for the experiments described in this thesis. Additionally, it would be impractical to use the Stratalinker that can process 3 plates per 30 minutes for this, so an alternative arrangement would be required. Overall, the required scale of the experiment could be achieved by in-house large-scale synthesis of photo-amino acids to obtain gram amounts, as well as the use of alternative irradiation means with higher energy light and larger area capacity to reduce the UV cross-linking time. At this stage, however, the sample purity and protein amount obtained was too low to consider SEC enrichment and cross-link identification. Additionally, cross-linking experiments on co-expressed proteins would benefit from and inclusion of a double P2A/T2A site between the proteins as any proteins attached through translation may not fold correctly, aggregate and result in identification of false cross-links.

As shown here, photo-amino acid inclusion and target protein overexpression likely change the state of cell signalling, both should be taken into account when designing future experiments and drawing conclusions from identifications obtained with this method. For example, RagAC pull-down experiments could be performed after re-introduction of full AA and serum medium for the last 2-4 hours prior to UV irradiation for pull-down of proteins associated with mTORC1 activation.

6 Conclusion

Protein-protein interactions within cell signalling networks are complex and challenging to study. In this work I have used several approaches to examine distinct complexes within glucose metabolism and mTORC1 network.

Using X-ray crystallography, I have determined the structure of a naturally occurring mutation in Hexokinase 2, which has a reduced activity and promotes metabolic reprogramming in Mexican cavefish. The resulting structure was largely same as the wild-type, highlighting the need for use of supplementary biophysical techniques in order to understand the effect of the mutation. Differential scanning fluorimetry has shown that the resulting protein is less stable than the wild-type, with lesser propensity to dimerise in absence of ligands, but largely unaffected ligand binding. However, there were some differences in the dynamics of unfolding based on ligand binding. To investigate this further, experiments with truncated N- and C-terminal domains would be required, as both hexokinase 2 domains are known to bind and turn over substrates which complicates analysis of the effects on individual domains and inter-domain crosstalk.

Within mTORC1 amino acid signalling branch, the mechanism of recruitment of mTOR to the lysosome and the architecture of the tethering activation complex have long been of interest. However, the highly flexible RagAC dimer and its weak interaction interface have remained elusive. Expression of active variants, separate high-yield purification of all components of the megacomplex from insect cells or *E. coli,* together with simplified grid preparation protocol have enabled structure determination of the entire lysosomal mTORC1 assembly with single particle cryo-EM. Our structure supports the observations made in two other recent publications with structures of the Ragulator-RagAC-Raptor and RagAC-mTORC1 but highlights the importance of inclusion of all complex members to form a more stable complex and allow for cryo-EM map reconstruction at a higher resolution.

Both X-ray and cryo-EM structure determination provide us with snapshots of these protein complexes. In some cases, cryo-EM can provide information about distinct states of the system through classification. However, resolving continuous movement and capturing locations of disordered residues is still difficult. The structures presented in this work are good examples of each of these cases. Cross-linking mass spectrometry is one of several potential avenues for exploration of such invisible states and structural elements. In the second part of this work, I have used the RagAC interaction network as a model to develop zero-length chemical and translationally incorporated cross-linking workflows to study similar dynamic complexes *in vitro* and *in vivo*.

The *in vitro* cross-linking results obtained with CDI as a cross-linker have shown that current methodologies do not allow for efficient and reliable identification of cross-links for flexible proteins and complexes. Sample preparation methods, MSMS analysis

methods and software all have substantial effects on the final outcomes. The results shown in this work highlight the need for further improvements in scoring functions within cross-link identification software packages as well as optimisation of mass spectrometry methods that would require less material and improve the data quality for unambiguous cross-link position identification. However, even if perfect data and identifications of cross-links could be achieved, there are still some major obstacles to obtaining useful information from such experiments: preferential cross-linking of more abundant states, links that are incompatible with each other for flexible and non-homogeneous samples, and aggregation. Given the very small improvement in plausible Cα-Cα distances and a reduced cross-linking efficiency and resulting shortage of data points, it is difficult to recommend the use of carbonyldiimidazole over linkers like DSSO with many more established protocols and applications for a variety of targets. However, CDI cross-linking would still be a useful and feasible strategy to study well-behaved small complexes that may require the reactivity towards hydroxyl groups to cover a particular area ('blind spot') that is lysine-poor or is in a tight spot where the bulkier cross-linker may not be able to access.

For photo-amino acid cross-linking, a substantial reduction in protein yield and low incorporations were observed, therefore any cross-linking mass-spectrometry would face similar issues as chemical cross-linking with CDI with regards to sample quantity and purity requirements, and mass spectrometry methods. More efficient cross-linked peptide separation without sample dilution or extensive fractionation prior to mass spectrometry measurement would be paramount to successfully identify diverse and low abundant cross-links with this method. Additionally, my findings show that inclusion of photo-amino acids, as well as target protein overexpression can influence the state of cellular metabolism and signalling. Therefore, further studies would help to better understand the influence of photo-amino acids on the cellular proteome, so that protocols can be adapted to each experiment depending on the network and interactions of interest. This is especially important as this method could provide the advantage in terms of increasing throughput and allow for unbiased study design of unprecedented depth with little prior knowledge of the system. At the very least it would allow to narrow the scope and reduce the resource requirements for low throughput methods that would anyway be needed to ascertain the biological relevance of identified interactions.

7 References

1. Lin, S.C. and D.G. Hardie, *AMPK: Sensing Glucose as well as Cellular Energy Status.* Cell Metab, 2018. **27**(2): p. 299-313.
2. Hawley, S.A., et al., *Characterization of the AMP-activated protein kinase kinase from rat liver and identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase.* J Biol Chem, 1996. **271**(44): p. 27879-87.
3. Shaw, R.J., et al., *The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress.* Proc Natl Acad Sci U S A, 2004. **101**(10): p. 3329-35.
4. Boudeau, J., et al., *MO25alpha/beta interact with STRADalpha/beta enhancing their ability to bind, activate and localize LKB1 in the cytoplasm.* EMBO J, 2003. **22**(19): p. 5102-14.
5. Hawley, S.A., et al., *Complexes between the LKB1 tumor suppressor, STRAD alpha/beta and MO25 alpha/beta are upstream kinases in the AMP-activated protein kinase cascade.* J Biol, 2003. **2**(4): p. 28.
6. Woods, A., et al., *LKB1 is the upstream kinase in the AMP-activated protein kinase cascade.* Curr Biol, 2003. **13**(22): p. 2004-8.
7. Gowans, G.J., et al., *AMP is a true physiological regulator of AMP-activated protein kinase by both allosteric activation and enhancing net phosphorylation.* Cell Metab, 2013. **18**(4): p. 556-66.
8. Hardie, D.G., et al., *AMP-activated protein kinase: an ultrasensitive system for monitoring cellular energy charge.* Biochem J, 1999. **338 (Pt 3)**: p. 717-22.
9. Corton, J.M., et al., *5-aminoimidazole-4-carboxamide ribonucleoside. A specific method for activating AMP-activated protein kinase in intact cells?* Eur J Biochem, 1995. **229**(2): p. 558-65.
10. Garcia, D. and R.J. Shaw, *AMPK: Mechanisms of Cellular Energy Sensing and Restoration of Metabolic Balance.* Mol Cell, 2017. **66**(6): p. 789-800.
11. Hardie, D.G., B.E. Schaffer, and A. Brunet, *AMPK: An Energy-Sensing Pathway with Multiple Inputs and Outputs.* Trends Cell Biol, 2016. **26**(3): p. 190-201.
12. Gonzalez, A., et al., *AMPK and TOR: The Yin and Yang of Cellular Nutrient Sensing and Growth Control.* Cell Metab, 2020. **31**(3): p. 472-492.
13. Saxton, R.A. and D.M. Sabatini, *mTOR Signaling in Growth, Metabolism, and Disease.* Cell, 2017. **168**(6): p. 960-976.
14. Winter-Vann, A.M. and G.L. Johnson, *Integrated activation of MAP3Ks balances cell fate in response to stress.* J Cell Biochem, 2007. **102**(4): p. 848-58.
15. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.
16. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.
17. DeBerardinis, R.J., et al., *The biology of cancer: metabolic reprogramming fuels cell growth and proliferation.* Cell Metab, 2008. **7**(1): p. 11-20.
18. Lunt, S.Y. and M.G. Vander Heiden, *Aerobic glycolysis: meeting the metabolic requirements of cell proliferation.* Annu Rev Cell Dev Biol, 2011. **27**: p. 441-64.

19. Palm, W. and C.B. Thompson, *Nutrient acquisition strategies of mammalian cells.* Nature, 2017. **546**(7657): p. 234-242.
20. Cassim, S., et al., *Warburg and Beyond: The Power of Mitochondrial Metabolism to Collaborate or Replace Fermentative Glycolysis in Cancer.* Cancers (Basel), 2020. **12**(5).
21. Moschetta, M., et al., *Therapeutic targeting of the mTOR-signalling pathway in cancer: benefits and limitations.* Br J Pharmacol, 2014. **171**(16): p. 3801-13.
22. Magaway, C., E. Kim, and E. Jacinto, *Targeting mTOR and Metabolism in Cancer: Lessons and Innovations.* Cells, 2019. **8**(12).
23. Heitman, J., N.R. Movva, and M.N. Hall, *Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast.* Science, 1991. **253**(5022): p. 905-9.
24. Sabers, C.J., et al., *Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells.* J Biol Chem, 1995. **270**(2): p. 815-22.
25. Sabatini, D.M., et al., *RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs.* Cell, 1994. **78**(1): p. 35-43.
26. Brown, E.J., et al., *A mammalian protein targeted by G1-arresting rapamycin-receptor complex.* Nature, 1994. **369**(6483): p. 756-8.
27. Chiu, M.I., H. Katz, and V. Berlin, *RAPT1, a mammalian homolog of yeast Tor, interacts with the FKBP12/rapamycin complex.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12574-8.
28. Roustan, V., et al., *An evolutionary perspective of AMPK-TOR signaling in the three domains of life.* J Exp Bot, 2016. **67**(13): p. 3897-907.
29. Cornu, M., V. Albert, and M.N. Hall, *mTOR in aging, metabolism, and cancer.* Curr Opin Genet Dev, 2013. **23**(1): p. 53-62.
30. Aylett, C.H., et al., *Architecture of human mTOR complex 1.* Science, 2016. **351**(6268): p. 48-52.
31. Sarbassov, D.D., et al., *Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton.* Curr Biol, 2004. **14**(14): p. 1296-302.
32. Stuttfeld, E., et al., *Architecture of the human mTORC2 core complex.* Elife, 2018. **7**.
33. Avruch, J., et al., *Activation of mTORC1 in two steps: Rheb-GTP activation of catalytic function and increased binding of substrates to raptor.* Biochem Soc Trans, 2009. **37**(Pt 1): p. 223-6.
34. Mahoney, S.J., et al., *A small molecule inhibitor of Rheb selectively targets mTORC1 signaling.* Nat Commun, 2018. **9**(1): p. 548.
35. Parmar, N. and F. Tamanoi, *Rheb G-Proteins and the Activation of mTORC1.* Enzymes, 2010. **27**: p. 39-56.
36. Sato, T., et al., *Specific activation of mTORC1 by Rheb G-protein in vitro involves enhanced recruitment of its substrate protein.* J Biol Chem, 2009. **284**(19): p. 12783-91.
37. Yang, H., et al., *Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40.* Nature, 2017. **552**(7685): p. 368-373.
38. Tamai, T., et al., *Rheb (Ras homologue enriched in brain)-dependent mammalian target of rapamycin complex 1 (mTORC1) activation becomes indispensable for*

cardiac hypertrophic growth after early postnatal period. J Biol Chem, 2013. **288**(14): p. 10176-87.
39. Nicklin, P., et al., *Bidirectional transport of amino acids regulates mTOR and autophagy.* Cell, 2009. **136**(3): p. 521-34.
40. Sancak, Y., et al., *The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1.* Science, 2008. **320**(5882): p. 1496-501.
41. Hara, K., et al., *Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism.* J Biol Chem, 1998. **273**(23): p. 14484-94.
42. Meng, D., et al., *Glutamine and asparagine activate mTORC1 independently of Rag GTPases.* J Biol Chem, 2020. **295**(10): p. 2890-2899.
43. Gu, X., et al., *SAMTOR is an S-adenosylmethionine sensor for the mTORC1 pathway.* Science, 2017. **358**(6364): p. 813-818.
44. Chantranupong, L., et al., *The CASTOR Proteins Are Arginine Sensors for the mTORC1 Pathway.* Cell, 2016. **165**(1): p. 153-164.
45. Saxton, R.A., et al., *Mechanism of arginine sensing by CASTOR1 upstream of mTORC1.* Nature, 2016. **536**(7615): p. 229-33.
46. Zhou, Y., et al., *Methionine Regulates mTORC1 via the T1R1/T1R3-PLCbeta-Ca(2+)-ERK1/2 Signal Transduction Process in C2C12 Cells.* Int J Mol Sci, 2016. **17**(10).
47. Bar-Peled, L., et al., *A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1.* Science, 2013. **340**(6136): p. 1100-6.
48. Kim, J. and E. Kim, *Rag GTPase in amino acid signaling.* Amino Acids, 2016. **48**(4): p. 915-928.
49. Kira, S., et al., *Dynamic relocation of the TORC1-Gtr1/2-Ego1/2/3 complex is regulated by Gtr1 and Gtr2.* Mol Biol Cell, 2016. **27**(2): p. 382-96.
50. Peli-Gulli, M.P., et al., *Amino Acids Stimulate TORC1 through Lst4-Lst7, a GTPase-Activating Protein Complex for the Rag Family GTPase Gtr2.* Cell Rep, 2015. **13**(1): p. 1-7.
51. Petit, C.S., A. Roczniak-Ferguson, and S.M. Ferguson, *Recruitment of folliculin to lysosomes supports the amino acid-dependent activation of Rag GTPases.* J Cell Biol, 2013. **202**(7): p. 1107-22.
52. Powis, K. and C. De Virgilio, *Conserved regulators of Rag GTPases orchestrate amino acid-dependent TORC1 signaling.* Cell Discov, 2016. **2**: p. 15049.
53. Huang, H., et al., *Role of glutamine and interlinked asparagine metabolism in vessel formation.* EMBO J, 2017. **36**(16): p. 2334-2352.
54. Jung, J., H.M. Genau, and C. Behrends, *Amino Acid-Dependent mTORC1 Regulation by the Lysosomal Membrane Protein SLC38A9.* Mol Cell Biol, 2015. **35**(14): p. 2479-94.
55. Rebsamen, M., et al., *SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1.* Nature, 2015. **519**(7544): p. 477-81.
56. Wang, S., et al., *Metabolism. Lysosomal amino acid transporter SLC38A9 signals arginine sufficiency to mTORC1.* Science, 2015. **347**(6218): p. 188-94.
57. Simon A. Fromm, R.E.L., James H. Hurley, *Structural mechanism for amino acid-dependent Rag GTPase switching by SLC38A9.* bioRxiv 2020.
58. Zhou, Y., et al., *Methionine and valine activate the mammalian target of rapamycin complex 1 pathway through heterodimeric amino acid taste receptor*

(TAS1R1/TAS1R3) and intracellular Ca(2+) in bovine mammary epithelial cells. J Dairy Sci, 2018. **101**(12): p. 11354-11363.
59. Jewell, J.L., et al., *Metabolism. Differential regulation of mTORC1 by leucine and glutamine.* Science, 2015. **347**(6218): p. 194-8.
60. Bernfeld, E., et al., *Phospholipase D-dependent mTOR complex 1 (mTORC1) activation by glutamine.* J Biol Chem, 2018. **293**(42): p. 16390-16401.
61. Cheng, Z., et al., *Rab1A promotes proliferation and migration abilities via regulation of the HER2/AKT-independent mTOR/S6K1 pathway in colorectal cancer.* Oncol Rep, 2019. **41**(5): p. 2717-2728.
62. Thomas, J.D., et al., *Rab1A Is an mTORC1 Activator and a Colorectal Oncogene.* Cancer Cell, 2016. **30**(1): p. 181-182.
63. Xu, B.H., et al., *Aberrant amino acid signaling promotes growth and metastasis of hepatocellular carcinomas through Rab1A-dependent activation of mTORC1 by Rab1A.* Oncotarget, 2015. **6**(25): p. 20813-28.
64. Sancak, Y., et al., *Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids.* Cell, 2010. **141**(2): p. 290-303.
65. Nada, S., et al., *p18/LAMTOR1: a late endosome/lysosome-specific anchor protein for the mTORC1/MAPK signaling pathway.* Methods Enzymol, 2014. **535**: p. 249-63.
66. Sekiguchi, T., et al., *Novel G proteins, Rag C and Rag D, interact with GTP-binding proteins, Rag A and Rag B.* J Biol Chem, 2001. **276**(10): p. 7246-57.
67. Jeong, J.H., et al., *Crystal structure of the Gtr1p(GTP)-Gtr2p(GDP) protein complex reveals large structural rearrangements triggered by GTP-to-GDP conversion.* J Biol Chem, 2012. **287**(35): p. 29648-53.
68. Lagerstedt, J.O., et al., *Structure and function of the GTP binding protein Gtr1 and its role in phosphate transport in Saccharomyces cerevisiae.* Biochemistry, 2005. **44**(2): p. 511-7.
69. Anandapadamanaban, M., et al., *Architecture of human Rag GTPase heterodimers and their complex with mTORC1.* Science, 2019. **366**(6462): p. 203-210.
70. de Araujo, M.E.G., et al., *Crystal structure of the human lysosomal mTORC1 scaffold complex and its impact on signaling.* Science, 2017. **358**(6361): p. 377-381.
71. Yonehara, R., et al., *Structural basis for the assembly of the Ragulator-Rag GTPase complex.* Nat Commun, 2017. **8**(1): p. 1625.
72. Zhang, T., et al., *Structural basis for Ragulator functioning as a scaffold in membrane-anchoring of Rag GTPases and mTORC1.* Nat Commun, 2017. **8**(1): p. 1394.
73. Kim, E., et al., *Regulation of TORC1 by Rag GTPases in nutrient response.* Nat Cell Biol, 2008. **10**(8): p. 935-45.
74. Bar-Peled, L., et al., *Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1.* Cell, 2012. **150**(6): p. 1196-208.
75. Panchaud, N., M.P. Peli-Gulli, and C. De Virgilio, *Amino acid deprivation inhibits TORC1 through a GTPase-activating protein complex for the Rag family GTPase Gtr1.* Sci Signal, 2013. **6**(277): p. ra42.
76. Shen, K., A. Choe, and D.M. Sabatini, *Intersubunit Crosstalk in the Rag GTPase Heterodimer Enables mTORC1 to Respond Rapidly to Amino Acid Availability.* Mol Cell, 2017. **68**(3): p. 552-565 e8.

77. Shen, K., et al., *Architecture of the human GATOR1 and GATOR1-Rag GTPases complexes.* Nature, 2018. **556**(7699): p. 64-69.
78. Padi, S.K.R., et al., *Phosphorylation of DEPDC5, a component of the GATOR1 complex, releases inhibition of mTORC1 and promotes tumor growth.* Proc Natl Acad Sci U S A, 2019. **116**(41): p. 20505-20510.
79. Wolfson, R.L., et al., *KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1.* Nature, 2017. **543**(7645): p. 438-442.
80. Peng, M., N. Yin, and M.O. Li, *SZT2 dictates GATOR control of mTORC1 signalling.* Nature, 2017. **543**(7645): p. 433-437.
81. Chantranupong, L., et al., *The Sestrins interact with GATOR2 to negatively regulate the amino-acid-sensing pathway upstream of mTORC1.* Cell Rep, 2014. **9**(1): p. 1-8.
82. Tsun, Z.Y., et al., *The folliculin tumor suppressor is a GAP for the RagC/D GTPases that signal amino acid levels to mTORC1.* Mol Cell, 2013. **52**(4): p. 495-505.
83. Meng, J. and S.M. Ferguson, *GATOR1-dependent recruitment of FLCN-FNIP to lysosomes coordinates Rag GTPase heterodimer nucleotide status in response to amino acids.* J Cell Biol, 2018. **217**(8): p. 2765-2776.
84. Lawrence, R.E., et al., *Structural mechanism of a Rag GTPase activation checkpoint by the lysosomal folliculin complex.* Science, 2019. **366**(6468): p. 971-977.
85. Shen, K., et al., *Cryo-EM Structure of the Human FLCN-FNIP2-Rag-Ragulator Complex.* Cell, 2019. **179**(6): p. 1319-1329 e8.
86. Rogala, K.B., et al., *Structural basis for the docking of mTORC1 on the lysosomal surface.* Science, 2019. **366**(6464): p. 468-475.
87. Shen, K., et al., *Arg-78 of Nprl2 catalyzes GATOR1-stimulated GTP hydrolysis by the Rag GTPases.* J Biol Chem, 2019. **294**(8): p. 2970-2975.
88. Dubouloz, F., et al., *The TOR and EGO protein complexes orchestrate microautophagy in yeast.* Mol Cell, 2005. **19**(1): p. 15-26.
89. Powis, K., et al., *Crystal structure of the Ego1-Ego2-Ego3 complex and its role in promoting Rag GTPase-dependent TORC1 signaling.* Cell Res, 2015. **25**(9): p. 1043-59.
90. Zhang, T., et al., *Ego3 functions as a homodimer to mediate the interaction between Gtr1-Gtr2 and Ego1 in the ego complex to activate TORC1.* Structure, 2012. **20**(12): p. 2151-60.
91. Mu, Z., et al., *Structural insight into the Ragulator complex which anchors mTORC1 to the lysosomal membrane.* Cell Discov, 2017. **3**: p. 17049.
92. Su, M.Y., et al., *Hybrid Structure of the RagA/C-Ragulator mTORC1 Activation Complex.* Mol Cell, 2017. **68**(5): p. 835-846 e3.
93. Shen, K. and D.M. Sabatini, *Ragulator and SLC38A9 activate the Rag GTPases through noncanonical GEF mechanisms.* Proc Natl Acad Sci U S A, 2018. **115**(38): p. 9545-9550.
94. Rebsamen, M. and G. Superti-Furga, *SLC38A9: A lysosomal amino acid transporter at the core of the amino acid-sensing machinery that controls MTORC1.* Autophagy, 2016. **12**(6): p. 1061-2.
95. Fromm, S.A., R.E. Lawrence, and J.H. Hurley, *Structural mechanism for amino acid-dependent Rag GTPase nucleotide state switching by SLC38A9.* Nat Struct Mol Biol, 2020. **27**(11): p. 1017-1023.

96. Zoncu, R., et al., *mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase.* Science, 2011. **334**(6056): p. 678-83.
97. Zhang, C.S., et al., *Determining AMPK Activation via the Lysosomal v-ATPase-Ragulator-AXIN/LKB1 Axis.* Methods Mol Biol, 2018. **1732**: p. 393-411.
98. Inoki, K., et al., *Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling.* Genes Dev, 2003. **17**(15): p. 1829-34.
99. Carroll, B., et al., *Control of TSC2-Rheb signaling axis by arginine regulates mTORC1 activity.* Elife, 2016. **5**.
100. Demetriades, C., N. Doumpas, and A.A. Teleman, *Regulation of TORC1 in response to amino acid starvation via lysosomal recruitment of TSC2.* Cell, 2014. **156**(4): p. 786-99.
101. Menon, S., et al., *Spatial control of the TSC complex integrates insulin and nutrient regulation of mTORC1 at the lysosome.* Cell, 2014. **156**(4): p. 771-85.
102. Huang, J., et al., *The TSC1-TSC2 complex is required for proper activation of mTOR complex 2.* Mol Cell Biol, 2008. **28**(12): p. 4104-15.
103. Zhang, H.M., et al., *Moderate lifelong overexpression of tuberous sclerosis complex 1 (TSC1) improves health and survival in mice.* Sci Rep, 2017. **7**(1): p. 834.
104. Choo, A.Y., et al., *Glucose addiction of TSC null cells is caused by failed mTORC1-dependent balancing of metabolic demand with supply.* Mol Cell, 2010. **38**(4): p. 487-99.
105. Jiang, X., H.L. Kenerson, and R.S. Yeung, *Glucose deprivation in tuberous sclerosis complex-related tumors.* Cell Biosci, 2011. **1**: p. 34.
106. Rocktaschel, P., A. Sen, and M.Z. Cader, *High glucose concentrations mask cellular phenotypes in a stem cell model of tuberous sclerosis complex.* Epilepsy Behav, 2019. **101**(Pt B): p. 106581.
107. Zhang, C.S., et al., *Fructose-1,6-bisphosphate and aldolase mediate glucose sensing by AMPK.* Nature, 2017. **548**(7665): p. 112-116.
108. Yoon, I., et al., *Glucose-dependent control of leucine metabolism by leucyl-tRNA synthetase 1.* Science, 2020. **367**(6474): p. 205-210.
109. Dibble, C.C., et al., *TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1.* Mol Cell, 2012. **47**(4): p. 535-46.
110. Roberts, D.J., et al., *Hexokinase-II positively regulates glucose starvation-induced autophagy through TORC1 inhibition.* Mol Cell, 2014. **53**(4): p. 521-33.
111. Almacellas, E., et al., *Phosphofructokinases Axis Controls Glucose-Dependent mTORC1 Activation Driven by E2F1.* iScience, 2019. **20**: p. 434-448.
112. Orozco, J.M., et al., *Dihydroxyacetone phosphate signals glucose availability to mTORC1.* Nat Metab, 2020. **2**(9): p. 893-901.
113. Egan, D., et al., *The autophagy initiating kinase ULK1 is regulated via opposing phosphorylation by AMPK and mTOR.* Autophagy, 2011. **7**(6): p. 643-4.
114. Zhao, M. and D.J. Klionsky, *AMPK-dependent phosphorylation of ULK1 induces autophagy.* Cell Metab, 2011. **13**(2): p. 119-20.
115. Kim, J., et al., *AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1.* Nat Cell Biol, 2011. **13**(2): p. 132-41.
116. Khan, S.H. and R. Kumar, *Role of an intrinsically disordered conformation in AMPK-mediated phosphorylation of ULK1 and regulation of autophagy.* Mol Biosyst, 2012. **8**(1): p. 91-6.

117. Kim, M. and J.H. Lee, *Identification of an AMPK phosphorylation site in Drosophila TSC2 (gigas) that regulate cell growth.* Int J Mol Sci, 2015. **16**(4): p. 7015-26.
118. Inoki, K., T. Zhu, and K.L. Guan, *TSC2 mediates cellular energy response to control cell growth and survival.* Cell, 2003. **115**(5): p. 577-90.
119. de Martin Garrido, N. and C.H.S. Aylett, *Nutrient Signaling and Lysosome Positioning Crosstalk Through a Multifunctional Protein, Folliculin.* Front Cell Dev Biol, 2020. **8**: p. 108.
120. Baba, M., et al., *Folliculin encoded by the BHD gene interacts with a binding protein, FNIP1, and AMPK, and is involved in AMPK and mTOR signaling.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15552-7.
121. Takagi, Y., et al., *Interaction of folliculin (Birt-Hogg-Dube gene product) with a novel Fnip1-like (FnipL/Fnip2) protein.* Oncogene, 2008. **27**(40): p. 5339-47.
122. Xie, J. and C.G. Proud, *Signaling crosstalk between the mTOR complexes.* Translation (Austin), 2014. **2**(1): p. e28174.
123. Wang, L., et al., *PRAS40 regulates mTORC1 kinase activity by functioning as a direct inhibitor of substrate binding.* J Biol Chem, 2007. **282**(27): p. 20036-44.
124. Wiza, C., E.B. Nascimento, and D.M. Ouwens, *Role of PRAS40 in Akt and mTOR signaling in health and disease.* Am J Physiol Endocrinol Metab, 2012. **302**(12): p. E1453-60.
125. Jia, R. and J.S. Bonifacino, *Lysosome Positioning Influences mTORC2 and AKT Signaling.* Mol Cell, 2019. **75**(1): p. 26-38 e3.
126. Pu, J., T. Keren-Kaplan, and J.S. Bonifacino, *A Ragulator-BORC interaction controls lysosome positioning in response to amino acid availability.* J Cell Biol, 2017. **216**(12): p. 4183-4197.
127. Cinti, A., et al., *HIV-1 enhances mTORC1 activity and repositions lysosomes to the periphery by co-opting Rag GTPases.* Sci Rep, 2017. **7**(1): p. 5515.
128. Miller, C.J. and B.E. Turk, *Homing in: Mechanisms of Substrate Targeting by Protein Kinases.* Trends Biochem Sci, 2018. **43**(5): p. 380-394.
129. Ubersax, J.A. and J.E. Ferrell, Jr., *Mechanisms of specificity in protein phosphorylation.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 530-41.
130. Good, M.C., J.G. Zalatan, and W.A. Lim, *Scaffold proteins: hubs for controlling the flow of cellular information.* Science, 2011. **332**(6030): p. 680-6.
131. Valk, E., et al., *Multistep phosphorylation systems: tunable components of biological signaling circuits.* Mol Biol Cell, 2014. **25**(22): p. 3456-60.
132. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. **340**(6230): p. 245-6.
133. Fetchko, M. and I. Stagljar, *Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions.* Methods, 2004. **32**(4): p. 349-62.
134. Kerppola, T.K., *Visualization of molecular interactions by fluorescence complementation.* Nat Rev Mol Cell Biol, 2006. **7**(6): p. 449-56.
135. Lievens, S., et al., *Kinase Substrate Sensor (KISS), a mammalian in situ protein interaction sensor.* Mol Cell Proteomics, 2014. **13**(12): p. 3332-42.
136. Zolghadr, K., et al., *A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells.* Mol Cell Proteomics, 2008. **7**(11): p. 2279-87.
137. Glatter, T., et al., *An integrated workflow for charting the human interaction proteome: insights into the PP2A system.* Mol Syst Biol, 2009. **5**: p. 237.

138. Buljan, M., et al., *Kinase Interaction Network Expands Functional and Disease Roles of Human Kinases.* Mol Cell, 2020. **79**(3): p. 504-520 e9.
139. Hein, M.Y., et al., *A human interactome in three quantitative dimensions organized by stoichiometries and abundances.* Cell, 2015. **163**(3): p. 712-23.
140. Huttlin, E.L., et al., *Architecture of the human interactome defines protein communities and disease networks.* Nature, 2017. **545**(7655): p. 505-509.
141. Roux, K.J., et al., *A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells.* J Cell Biol, 2012. **196**(6): p. 801-10.
142. Rhee, H.W., et al., *Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging.* Science, 2013. **339**(6125): p. 1328-1331.
143. Liu, X., et al., *An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations.* Nat Commun, 2018. **9**(1): p. 1188.
144. Liu, X., et al., *Combined proximity labeling and affinity purification-mass spectrometry workflow for mapping and visualizing protein interaction networks.* Nat Protoc, 2020. **15**(10): p. 3182-3211.
145. Schopp, I.M., et al., *Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes.* Nat Commun, 2017. **8**: p. 15690.
146. O'Reilly, F.J. and J. Rappsilber, *Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology.* Nat Struct Mol Biol, 2018. **25**(11): p. 1000-1008.
147. Lento, C., et al., *Time-resolved ElectroSpray Ionization Hydrogen-deuterium Exchange Mass Spectrometry for Studying Protein Structure and Dynamics.* J Vis Exp, 2017(122).
148. Xiao, K., et al., *Revealing the architecture of protein complexes by an orthogonal approach combining HDXMS, CXMS, and disulfide trapping.* Nat Protoc, 2018. **13**(6): p. 1403-1428.
149. Biehn, S.E. and S. Lindert, *Accurate protein structure prediction with hydroxyl radical protein footprinting data.* Nat Commun, 2021. **12**(1): p. 341.
150. Espino, J.A. and L.M. Jones, *In Vivo Hydroxyl Radical Protein Footprinting for the Study of Protein Interactions in Caenorhabditis elegans.* J Vis Exp, 2020(158).
151. Rajabi, K., A.E. Ashcroft, and S.E. Radford, *Mass spectrometric methods to analyze the structural organization of macromolecular complexes.* Methods, 2015. **89**: p. 13-21.
152. Lensink, M.F., et al., *Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment.* Proteins, 2019. **87**(12): p. 1200-1221.
153. Busenlehner, L.S. and R.N. Armstrong, *Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry.* Arch Biochem Biophys, 2005. **433**(1): p. 34-46.
154. Percy, A.J., et al., *Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review.* Anal Chim Acta, 2012. **721**: p. 7-21.
155. Takamoto, K. and M.R. Chance, *Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 251-76.

156. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry.* Anal Chem, 1999. **71**(18): p. 3965-73.
157. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.
158. Korn, A.H., S.H. Feairheller, and E.M. Filachione, *Glutaraldehyde: nature of the reagent.* J Mol Biol, 1972. **65**(3): p. 525-9.
159. Hardy, P.M., A.C. Nicholls, and H.N. Rydon, *The nature of the cross-linking of proteins by glutaraldehyde. Part I. Interaction of glutaraldehyde with the amino-groups of 6-aminohexanoic acid and of alpha-N-acetyl-lysine.* J Chem Soc Perkin 1, 1976(9): p. 958-62.
160. Hermann, R., R. Jaenicke, and R. Rudolph, *Analysis of the reconstitution of oligomeric enzymes by cross-linking with glutaraldehyde: kinetics of reassociation of lactic dehydrogenase.* Biochemistry, 1981. **20**(18): p. 5195-201.
161. Leitner, A., et al., *Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics.* Mol Cell Proteomics, 2010. **9**(8): p. 1634-49.
162. Hermanson, G.T., *Bioconjugate Techniques*. Third Edition ed. 2013: San Diego, CA: Academic Press.
163. Gotze, M., et al., *A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions.* Anal Chem, 2019. **91**(15): p. 10236-10244.
164. Iacobucci, C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions.* Nat Protoc, 2018. **13**(12): p. 2864-2889.
165. Klykov, O., et al., *Efficient and robust proteome-wide approaches for cross-linking mass spectrometry.* Nat Protoc, 2018. **13**(12): p. 2964-2990.
166. Liu, F., et al., *Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification.* Nat Commun, 2017. **8**: p. 15473.
167. Madler, S., et al., *Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities.* J Mass Spectrom, 2009. **44**(5): p. 694-706.
168. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.
169. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* Angew Chem Int Ed Engl, 2018. **57**(22): p. 6390-6396.
170. Madler, S., et al., *Does chemical cross-linking with NHS esters reflect the chemical equilibrium of protein-protein noncovalent interactions in solution?* J Am Soc Mass Spectrom, 2010. **21**(10): p. 1775-83.
171. Fursch, J., et al., *Proteome-Wide Structural Probing of Low-Abundant Protein Interactions by Cross-Linking Mass Spectrometry.* Anal Chem, 2020. **92**(5): p. 4016-4022.
172. Merkley, E.D., et al., *Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances.* Protein Sci, 2014. **23**(6): p. 747-59.
173. Degiacomi, M.T., et al., *Accommodating Protein Dynamics in the Modeling of Chemical Crosslinks.* Structure, 2017. **25**(11): p. 1751-1757 e5.

174. Bullock, J.M.A., et al., *Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry.* Structure, 2018. **26**(7): p. 1015-1024 e2.
175. Leitner, A., et al., *Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. 9455-60.
176. Hage, C., et al., *The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis.* Angew Chem Int Ed Engl, 2017. **56**(46): p. 14551-14555.
177. Debelyy, M.O., et al., *Chemical crosslinking and mass spectrometry to elucidate the topology of integral membrane proteins.* PLoS One, 2017. **12**(10): p. e0186840.
178. Mintseris, J. and S.P. Gygi, *High-density chemical cross-linking for modeling protein interactions.* Proc Natl Acad Sci U S A, 2020. **117**(1): p. 93-102.
179. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes.* J Struct Biol, 2011. **173**(3): p. 530-40.
180. Rampler, E., et al., *Comprehensive Cross-Linking Mass Spectrometry Reveals Parallel Orientation and Flexible Conformations of Plant HOP2-MND1.* J Proteome Res, 2015. **14**(12): p. 5048-62.
181. Matzinger, M. and K. Mechtler, *Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks in Vivo.* J Proteome Res, 2021. **20**(1): p. 78-93.
182. Dumas, A., et al., *Designing logical codon reassignment - Expanding the chemistry in biology.* Chem Sci, 2015. **6**(1): p. 50-69.
183. Chung, C.Z., K. Amikura, and D. Soll, *Using Genetic Code Expansion for Protein Biochemical Studies.* Front Bioeng Biotechnol, 2020. **8**: p. 598577.
184. Berg, M., et al., *An in vivo photo-cross-linking approach reveals a homodimerization domain of Aha1 in S. cerevisiae.* PLoS One, 2014. **9**(3): p. e89436.
185. Pettelkau, J., et al., *Reliable identification of cross-linked products in protein interaction studies by 13C-labeled p-benzoylphenylalanine.* J Am Soc Mass Spectrom, 2014. **25**(9): p. 1628-41.
186. Marx, D.C., et al., *SurA is a cryptically grooved chaperone that expands unfolded outer membrane proteins.* Proc Natl Acad Sci U S A, 2020. **117**(45): p. 28026-28035.
187. Yamano, K., et al., *Site-specific Interaction Mapping of Phosphorylated Ubiquitin to Uncover Parkin Activation.* J Biol Chem, 2015. **290**(42): p. 25199-211.
188. Hauser, M., et al., *Identification of peptide-binding sites within BSA using rapid, laser-induced covalent cross-linking combined with high-performance mass spectrometry.* J Mol Recognit, 2018. **31**(2).
189. Schwarz, R., et al., *Monitoring conformational changes in peroxisome proliferator-activated receptor alpha by a genetically encoded photoamino acid, cross-linking, and mass spectrometry.* J Med Chem, 2013. **56**(11): p. 4252-63.
190. Suchanek, M., A. Radzikowska, and C. Thiele, *Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells.* Nat Methods, 2005. **2**(4): p. 261-7.
191. Yang, T., et al., *Photo-lysine captures proteins that bind lysine post-translational modifications.* Nat Chem Biol, 2016. **12**(2): p. 70-2.

192. MacKinnon, A.L., et al., *Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation.* J Am Chem Soc, 2007. **129**(47): p. 14560-1.
193. Lossl, P., et al., *Analysis of nidogen-1/laminin gamma1 interaction by cross-linking, mass spectrometry, and computational modeling reveals multiple binding modes.* PLoS One, 2014. **9**(11): p. e112886.
194. Haupl, B., C.H. Ihling, and A. Sinz, *Combining affinity enrichment, cross-linking with photo amino acids, and mass spectrometry for probing protein kinase D2 interactions.* Proteomics, 2017. **17**(10): p. e1600459.
195. Henderson, T.A. and M.L. Nilles, *In Vivo Photo-Cross-Linking to Study T3S Interactions Demonstrated Using the Yersinia pestis T3S System.* Methods Mol Biol, 2017. **1531**: p. 47-60.
196. Black, D.J., et al., *Evaluating Calmodulin-Protein Interactions by Rapid Photoactivated Cross-Linking in Live Cells Metabolically Labeled with Photo-Methionine.* J Proteome Res, 2019. **18**(10): p. 3780-3791.
197. Kolbel, K., C.H. Ihling, and A. Sinz, *Analysis of peptide secondary structures by photoactivatable amino acid analogues.* Angew Chem Int Ed Engl, 2012. **51**(50): p. 12602-5.
198. Lossl, P. and A. Sinz, *Combining Amine-Reactive Cross-Linkers and Photo-Reactive Amino Acids for 3D-Structure Analysis of Proteins and Protein Complexes.* Methods Mol Biol, 2016. **1394**: p. 109-127.
199. Chung, C.T., S.L. Niemela, and R.H. Miller, *One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution.* Proc Natl Acad Sci U S A, 1989. **86**(7): p. 2172-5.
200. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors.* Nat Methods, 2006. **3**(12): p. 1021-32.
201. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
202. Ahrne, E., et al., *Critical assessment of proteome-wide label-free absolute abundance estimation strategies.* Proteomics, 2013. **13**(17): p. 2567-78.
203. Riddle, M.R., et al., *Insulin resistance in cavefish as an adaptation to a nutrient-limited environment.* Nature, 2018. **555**(7698): p. 647-651.
204. Shimobayashi, M., et al., *Diet-induced loss of adipose Hexokinase 2 triggers hyperglycemia.* bioRxiv, 2020.
205. Nawaz, M.H., et al., *The catalytic inactivation of the N-half of human hexokinase 2 and structural and biochemical characterization of its mitochondrial conformation.* Biosci Rep, 2018. **38**(1).
206. Aleshin, A.E., et al., *Nonaggregating mutant of recombinant human hexokinase I exhibits wild-type kinetics and rod-like conformations in solution.* Biochemistry, 1999. **38**(26): p. 8359-66.
207. Aleshin, A.E., et al., *Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation.* J Mol Biol, 2000. **296**(4): p. 1001-15.
208. Aleshin, A.E., et al., *The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose and glucose-6-phosphate.* Structure, 1998. **6**(1): p. 39-50.

209. Duy, C. and J. Fitter, *How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra.* Biophys J, 2006. **90**(10): p. 3704-11.
210. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.
211. Zahid, N.I., et al., *Fluorescence probing of the temperature-induced phase transition in a glycolipid self-assembly: hexagonal <--> micellar and cubic <--> lamellar.* Langmuir, 2012. **28**(11): p. 4989-95.
212. Fan, K., et al., *Hexokinase 2 dimerization and interaction with voltage-dependent anion channel promoted resistance to cell apoptosis induced by gemcitabine in pancreatic cancer.* Cancer Med, 2019. **8**(13): p. 5903-5915.
213. Tsai, H.J. and J.E. Wilson, *Functional organization of mammalian hexokinases: both N- and C-terminal halves of the rat type II isozyme possess catalytic sites.* Arch Biochem Biophys, 1996. **329**(1): p. 17-23.
214. Leitner, A., et al., *Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography.* Mol Cell Proteomics, 2012. **11**(3): p. M111 014126.
215. Fritzsche, R., et al., *Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis.* Rapid Commun Mass Spectrom, 2012. **26**(6): p. 653-8.
216. Mendes, M.L., et al., *An integrated workflow for crosslinking mass spectrometry.* Mol Syst Biol, 2019. **15**(9): p. e8994.
217. Gotze, M., et al., *Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis.* J Am Soc Mass Spectrom, 2015. **26**(1): p. 83-97.
218. Muller, M.Q., et al., *Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS.* Anal Chem, 2010. **82**(16): p. 6958-68.
219. Kolbowski, L., M.L. Mendes, and J. Rappsilber, *Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer.* Anal Chem, 2017. **89**(10): p. 5311-5318.
220. Stieger, C.E., P. Doppler, and K. Mechtler, *Optimized Fragmentation Improves the Identification of Peptides Cross-Linked by MS-Cleavable Reagents.* J Proteome Res, 2019. **18**(3): p. 1363-1370.
221. Wu, S.T., Y.Q. Xia, and M. Jemal, *High-field asymmetric waveform ion mobility spectrometry coupled with liquid chromatography/electrospray ionization tandem mass spectrometry (LC/ESI-FAIMS-MS/MS) multi-component bioanalytical method development, performance evaluation and demonstration of the constancy of the compensation voltage with change of mobile phase composition or flow rate.* Rapid Commun Mass Spectrom, 2007. **21**(22): p. 3667-76.
222. Bonneil, E., S. Pfammatter, and P. Thibault, *Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS).* J Mass Spectrom, 2015. **50**(11): p. 1181-95.
223. Zhurov, K.O., et al., *Distinguishing analyte from noise components in mass spectra of complex samples: where to cut the noise?* Anal Chem, 2014. **86**(7): p. 3308-16.
224. Ding, W., et al., *Stress-responsive and metabolic gene regulation are altered in low S-adenosylmethionine.* PLoS Genet, 2018. **14**(11): p. e1007812.

225. Kim, H.H. and C.S. Park, *Methionine cytotoxicity in the human breast cancer cell line MCF-7.* In Vitro Cell Dev Biol Anim, 2003. **39**(3-4): p. 117-9.
226. Benavides, M.A., et al., *Methionine inhibits cellular growth dependent on the p53 status of cells.* Am J Surg, 2007. **193**(2): p. 274-83.
227. Benavides, M.A., et al., *Suppression by L-methionine of cell cycle progression in LNCaP and MCF-7 cells but not benign cells.* Anticancer Res, 2010. **30**(6): p. 1881-5.
228. Benavides, M.A., et al., *L-methionine-induced alterations in molecular signatures in MCF-7 and LNCaP cancer cells.* J Cancer Res Clin Oncol, 2011. **137**(3): p. 441-53.
229. Martinez-Una, M., et al., *Excess S-adenosylmethionine reroutes phosphatidylethanolamine towards phosphatidylcholine and triglyceride synthesis.* Hepatology, 2013. **58**(4): p. 1296-305.
230. Benavides, M.A., et al., *L-Methionine inhibits growth of human pancreatic cancer cells.* Anticancer Drugs, 2014. **25**(2): p. 200-3.
231. Skovierova, H., et al., *The Molecular and Cellular Effect of Homocysteine Metabolism Imbalance on Human Health.* Int J Mol Sci, 2016. **17**(10).

Appendices (supplied electronically)

I. HK2 R42H PDB validation report

Preliminary PDB Validation report for the crystal structure.

II. RagAC mutant verification table

MaxQuant 'Peptide' table, filtered for peptides of interest.

III. CDI cross-link table

Unique cross-links identified with Merox 1.6.6.6 and 2.0.1.4, and XlinkX

IV. GFP peptide level photo-leucine incorporation

MaxQuant evidence table filtered for GFP.

V. RagAC photo-leucine incorporation

MaxQuant evidence table filtered for RagA and RagC

VI. RagAC Pull-down results (No UV/UV enrichment)

Results of proteomic analysis of RagAC pull-down from cell lysate. MaxQuant 'Protein Groups' table: samples 1-3: Transfected, untreated; samples 4-6: transfected, UV treated; samples 7-9: non-transfected, untreated.

SafeQuant Output: Samples 1-3 Untreated; Samples 4-6 UV treated.

Curriculum vitae Yana Demyanenko

I am passionate about proteomics and am working towards developing and implementing state-of-the-art sample preparation methods for studying protein-protein interactions and signalling networks.

Education

PhD in Biochemistry University of Basel, Maier Lab 2017-2021

Adapted available cross-linking mass spectrometry (XL-MS) workflows for use with ultra-short and zero-length MS-cleavable cross-linkers. Evaluated their use for characterization of protein-protein interfaces of large macromolecular complexes within mTORC1 signaling network in vitro.

Developed a workflow for position-specific analysis of photo-amino acid incorporation into recombinantly expressed proteins in HEK cells for in vivo structural studies.

MRes in Biochemistry University of Oxford, Kleanthous Lab 2015-2017

Used XL-MS to characterize and model the interaction of an antimicrobial protein colicin E9 and its receptor and translocator proteins at the outer membrane of E. coli. Gained experience handling LTQ Orbitrap and Q Exactive MS.

BSc Biotechnology Newcastle University 2011-2015

Upper Second Class Honours with Industrial placement year

Experience

Research Technician at Orla Protein Technologies 2013-2014

Managed lab equipment and stocks. Expressed and purified engineered proteins according to GMP. Produced functional protein-coated plastic and gold surfaces for our customers Performed batch quality control and worked on R&D projects. Responsible for research and production in absence of senior scientists

iGEM Team Member Summer 2013

Team gold for development and characterization of BioBricks for B. subtilis, which provide an inducible shape switch (from rod to cell wall-free L-form). Organized and led a rule-based modelling workshop on BioNetGen language.

Achievements

Proteomics in Cell Biology and disease	2020
Integrative modelling of biomolecular interactions (EMBO course, Poster)	2018
Annual PhD Retreat Organizer	2018-2019
PhD Fellowship for Excellence	2017-2021
INTO Newcastle Academic Achievement Student Prize	2011