Control, use and misuse of cellular signalling networks

Pauline Juyoux

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Pauline JUYOUX

Thèse dirigée par **Andrew MAC CARTHY**
et codirigée par **Matthew BOWLER**,chercheur, EMBL

préparée au sein du **Laboratoire laboratoire européen de biologie moléculaire**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Contrôle et régulation des voies de signalisation cellulaires

Control, use and misuse of cellular signalling networks

Thèse soutenue publiquement le **30 juin 2021**,
devant le jury composé de :

Monsieur Frank SICHERI
PROFESSEUR, Université de Toronto, Rapporteur
Madame Yi JIN
CHARGE DE RECHERCHE, Université de Cardiff, Examinatrice
Madame Jane ENDICOTT
PROFESSEUR, Université de Newcastle, Rapporteure
Monsieur Matthew W. BOWLER
CHARGE DE RECHERCHE HDR, EMBL, Co-directeur de thèse
Monsieur Wojciech GALEJ
CHARGE DE RECHERCHE HDR, EMBL, Examinateur
Monsieur Mohamed-Ali HAKIMI
DIRECTEUR DE RECHERCHE, INSERM délégation Auvergne-Rhône-Alpes, Président

"Process of Acceleration"

A pursuit in understanding the world around us is crucial to create change, but how this knowledge is used and what change is created is the constant friction of civilization.

Li-Hill – Grenoble Street Art Fest 2019

Picture: Cédric Dulou

Acknowledgement & Remerciements

I would say that the Acknowledgement is one of the most important parts of a thesis as not a word of this manuscript would have been possible without the help, support and collaboration of the people I would like to thank.

First of all, I would like to thank Matthew Bowler, for being the greatest supervisor I could have hoped for, as well as for setting an example on how to be a great scientist and great human being. Thank you for giving me the opportunity to work on these interesting and challenging projects. For teaching me, with a lot of patience, about structural biology techniques and theory. For your enthusiasm and continuous encouragement. For opening my views on collaborations and open science. For your trust.

I would also like to thank Erika Pellegrini. For teaching me experimental techniques and transmitting very useful lab tips and tricks. For your interest and collaboration on these projects that you initiated during your PhD. For saving a big part of my project I had given up hope on. For great discussions. For being a friend as well as a great supervisor and colleague. For your mentoring. For setting an example I aspire to follow.

I want to thank Andrew McCarthy, for welcoming me in his group and lab. For trusting me and giving me independence and freedom. For allowing me to participate in many conferences, courses and activities.

I would like to thank the members of my jury: Prof. Jane Endicott, Prof. Franck Sicheri, and Dr Yi Jin, for taking the time to read and evaluate my work; as well as the members of my thesis advisory committee, Ali Hakimi, Toby Gibson and Wojtek Galej to have followed me throughout these years of my PhD.

I had the chance to work with, and supervise, three undergraduate students during my PhD, who helped me on several projects, and that made me grow as a scientist and supervisor. First of all, I would like to thank Jill von Velsen. You welcomed me into the lab and I believe you taught me at least as much as I taught you, if not more. I admire your passion and values. Thank you for keeping me motivated and interested. Your friendship means a lot to me. Then I would like to thank Gabrielle Admans. Your placement was short but intense. Your motivation, interest and skills gave the project a considerable boost. And you made summer 2019 way funnier. I would also like to thank Jack Clark. Between my injury and COVID-19, I could not supervise you as well as I wished to, but I was proud to see how you gained in autonomy and kept the project going.

I would like to express my gratitude to two people whose help has been crucial to get me started in the lab: Francesco Bisiak, a fellow predoc at the time, who took the time to teach me during his last year; and Vincent Mariaule, a walking manual and a great and compassionate colleague, I tried to learn as much as possible from you, and I am eternally grateful.

To Joanna Wandzik and Anne-Sophie Humm, your support and your friendship helped me hold on during difficult periods, but more than anything, made my journey so special. Thank you so much for everything.

I wish to thank Martha Brennich, Jens Hoffman and Shibom Basu from the McCarthy group for discussions and inputs.

I would like to acknowledge our collaborators who performed some of the experiments presented here: Laurence Braun and Alexandre Bougdour from Ali Hakimi's team at the IAB Grenoble who performed the cellular experiments with *Toxoplasma;* and Mark Tully from the ESRF Grenoble for collecting and processing some of the SAXS data.

I would like to acknowledge all the facilities and their managers for providing the structure and training for the experiments. Thanks to Michael Hons and Sarah Schneider for the Grenoble EM facility; to Felix Weiss and Wim Hagen from the Heidelberg EM facility for data collection and advice; to Alice Auber for the Eukaryotic Expression Facility; to Audrey Spiellman, Sylvie Rhodes and Laeticia Carraud of the Lab Service for making everything easier; to Ombelline Pessey, for radioactivity training and moral support with ACL surgery; to Mandy Rettel and Dominic Helm from the Proteomics Core facility who ran the MS experiments; to Kim Remans, Kathryn Perez and Karine Lapouge from the PEP core for help with protein expression and ITC measurements.

I wish to thank in particular the HTX team for a great service and platform, last minute plate setup, help during data collection, but also very importantly for friendship and support. Guillaume Hoffman, Andrea Pica, Irina Cornaciu, Peter Murphy, Florine Dupeux, Raphaël Bourgeas and the others, thank you for all the nice moments.

The EMBL administration is full of dedicated people making our life easier. I wish to thank in particular Elise Bralet, Jennyfer Foray, Monika Larchner and Patricia Cabezas. Thanks to EMBL for trying to make this place safer and open about mental health.

Overall, I am grateful for all my colleagues that make the EMBL such a particular and international place to work in. To my fellow PhD students, I wish we all find our place in science and make a difference.

Muchissimas gracias a Estefanía Lucendo y Ally García Jareño. Porque mi aventura en ciencia empezó antes de esta tesís, y os debo muchissimo. Gracias por enseñarme tanto, y para armarme con trucos y ideales tanto en el laboratorio como sobre lo que debería ser la ciencia. Gracias por su amistad.

Merci à mes ami·es et ma famille, pour avoir essayé de comprendre ce que je faisais. Pour me demander comment allaient mes protéines. Pour m'avoir écoutée sans toujours comprendre de quoi je parlais. Pour votre fierté et vos encouragements qui m'ont beaucoup aidée. And last but not least, merci à Arthur. Il faudrait des pages pour tout détailler. Merci pour ton soutien, ton aide précieuse et tes encouragements. Merci pour tout.

Table of Contents

ACKNOWLEDGEMENT & REMERCIEMENTS 7

TABLE OF CONTENTS 9

ABSTRACT 15

RESUME 16

GENERAL AUDIENCE ABSTRACT 19

RESUME DE THESE VULGARISE 19

INTRODUCTION 21

SIGNALLING PATHWAYS 21
Kinases at the centre of cell signalling 21
The canonical MAPK pathways 22
The multiple roles of p38α in cell signalling 24
p38α cellular functions 24
MKK6, the activator of p38α 25
p38α and diseases 26
Structural details of the p38α and MKK6 kinases 27
Protein kinase shared structural characteristics 27
MAPK specificities 31
Docking site and KIM motifs 31
Hydrophobic pocket/DEF site/lipid-binding site 34
MAPK extension 34
MAP2K specificities 34
p38α activation 35
Canonical p38α activation by MKK6 35
Alternative activation mechanisms of p38α 35
PTMs that influence p38 activation and activity 36
p38α regulation and signal termination 36
Kinase inhibitors and the “specificity” challenge 36
HOST-PATHOGEN INTERACTIONS 38
SLiMs mimicry and interactions with host signalling pathways 38
Toxoplasma gondii 38
Toxoplasmosis 38
Cellular invasion mechanism 39
Host immune responses 40
T. gondii effectors 41
Therapeutic outcomes 43
AIMS OF THIS THESIS 45

CHAPTER I: CHARACTERIZATION OF THE ACTIVATION OF P38A BY ITS UPSTREAM MAP2K MKK6 47

INTRODUCTION 47
MATERIALS AND METHODS 51
Protein expression and purification 51
Recombinant protein constructs 51
Plasmid amplification 54
Protein expression in *E. coli* 54
Protein expression in insect cells 54
Protein purification 54
SDS-PAGE gels 55
Buffer optimization (Thermofluor assay) 56
Mass spectrometry 56
Intact mass by Q-TOF MS (Quadrupole time-of-flight Mass Spectrometry) 56
Digestion and PTMs analysis by LC-MS/MS (Liquid chromatography coupled to tandem Mass Spectrometry) 56
Sample preparation 56
LC-MS/MS 57
Data analysis 57
SAXS 57
X-ray crystallography 58
Crystallization 58
Negative stain electron microscopy 58
Grafix crosslinking 58
Grid preparation 58
Data collection and processing 58
CryoEM 59
Grid preparation 59
Screening and data collection 59
Data processing 59
RESULTS 61
How to express and stabilize a functional MKK6/p38α complex in the transition state 61
Engineering of an MKK6/GRA24 chimera to stabilize the interaction with p38α 61
Expression and purification optimization 62
Strategy 1: separate expression in *E. coli* and co-purification 62
Strategy 2: Separate expression in insect cells (MKK6DDGRA) and *E. coli*, separate initial purification and co-purification by gel filtration 63
Control of the MKK6DDGRA/p38α complex phosphorylation state and transition states analogues 64
An active and functional complex 66
p38α 67
mMKK6DDGRA 68
Towards a better structural understanding of the MKK6/p38α interaction 70
Attempts to solve the complex structure by X-ray crystallography 70
The apo mMKK6DDGRA/p38α complex is present and stable in solution, but flexible 71
Low resolution negative stain EM map of the crosslinked mMKK6DDGRA/p38α complex 73

The mMKK6DDGRA/p38α T180V complex with ADP and AlF_4^- is in a more compact conformation 75
Preliminary cryoEM map of the MKK6DDGRA/p38α T180V complex with ADP + AlF_4^- 78
Specimen optimization 78
Data processing of the Krios dataset 80
Architecture of the MKK6DDGRA/p38α T180V complex 82
Novel interaction interface between p38α hydrophobic pocket and MKK6 αG-helix 84
MKK6 αC-helix conformation and N-terminal extension 85
Focus on the active site and activation loops conformations 87
DISCUSSION AND PERSPECTIVES 92
The strategy to stabilize a transient interaction is crucial 92
The benefits of an integrative structural biology approach 92
The cryoEM resolution revolution has opened new possibilities for the study of small objects 93
New insights into the interaction between a MAPK and its upstream MAP2K 94
Novel interaction interface between p38α hydrophobic pocket and MKK6 αG-helix 95
Transition state or post-catalytic state? 98

CHAPTER II: NOVEL MECHANISM OF ACTIVATION OF P38A BY MKK6 USING ADP AS A SOURCE OF PHOSPHATE 99

INTRODUCTION 99
MATERIALS AND METHODS 102
Protein expression and purification 102
Recombinant protein constructs 102
Nucleotides and analogues 104
Phosphorylation assays with native-PAGE gels 104
Phosphorylation assays with radiolabelled nucleotides 104
RESULTS 106
p38α is phosphorylated on its activation loop Thr180 and Tyr182 residues by MKK6DD in the presence of ADP 106
ADP can be used as a phosphoryl source 106
ADP is less efficient than ATP to phosphorylate p38α 108
p38α phosphorylation with ADP is due to MKK6DD activity 110
The use of β-phosphate is not required to achieve dual phosphorylation of p38α . 114
MKK6 is able to use the β-phosphate of ADP and ATP to phosphorylate p38α 115
Direct evidence of β-phosphate incorporation 115
Is the nucleophilic attack of β-phosphate dependent on the nucleotide concentration? 116
Can other kinases also use ADP as a source of phosphate? 118
DISCUSSION AND PERSPECTIVES 120
ADP kinases are rare and cannot use ATP 120
The structural flexibility of the MKK6/p38α catalysis conformation might have a role in the catalysis of both nucleotides 121
The advantages of using both ADP and ATP for dual-phosphorylation 122
Processive or distributive? 122
ADP phosphorylation could help maintain active signalling in stress conditions 123

MKK6 could use ADP instead of ATP to orientate the p38α signalling to specific responses 124
Is the ability to use ADP for phosphorylation shared by other MAP2Ks or other kinases? 124

CHAPTER III: INTERACTIONS BETWEEN *T. GONDII* GRA24 AND HUMAN MAPKS 126

INTRODUCTION 126
MATERIALS AND METHODS 129
Protein expression and purification 129
Recombinant protein constructs 129
Peptides 131
ITC 131
X-ray crystallography 131
Crystallization optimisation 131
Data collection 132
Structure determination and refinement 132
Cellular assays 132
Parasites and host cells 132
Reagents 132
Cell fractionation 133
Western blots 133
RESULTS 134
Interaction between the GRA24 KIM peptide and ERK1, p38α and JNK1 MAPKs 134
Crystal structure of ERK1 bound to GRA24 KIM1 peptide 136
Crystallization optimization 136
Data collection and structure determination 138
Does GRA24 induces ERK1/2 phosphorylation and activation during host cell infection by *T. gondii*? 145
DISCUSSION AND PERSPECTIVES 147
The GRA24 KIM1 peptide binds to ERK1 docking site in a similar fashion to p38α 147
GRA24 does not influence ERK1/2 phosphorylation during *T. gondii* infection 147
Does the interaction between GRA24 and ERK1/2 during *T. gondii* infection have an effect on ERK1/2 signalling? 147

CHAPTER IV: INTERACTIONS BETWEEN *T. GONDII* GRA16 AND HUMAN USP7 149

INTRODUCTION 149
MATERIALS AND METHODS 153
Protein expression and purification 153
Recombinant protein constructs 153
Peptides 155
Protein purification 155
Peptide microarrays 156
ITC 157
X-ray crystallography 157
Crystallization optimization 157
Data collection 157

Structure determination and refinement 158
Cellular assays 158
Cells 158
293-T-REx transfection 158
Immuno-precipitation assay 158
Western blots 158
RESULTS 159
SLIM identification 159
Identification of the potential binding regions of GRA16 to USP7 160
Expression and purification of USP7 domains 163
GRA16 peptide containing MATH2 docking motif has high affinity for USP7 C-terminal domain 164
Crystal structure of GRA16 peptide containing a MATH2 motif to USP7 N-terminal domain 168
Crystallization optimization 168
Data collection and structure determination 171
Comparison of the binding between different peptides and USP7 N-terminal domain 176
Deletion of GRA16 MATH2 motif abolishes its interaction with USP7 178
DISCUSSION AND PERSPECTIVES 180
Necessity to complement *in silico* and *in vitro* interaction studies with cellular assays 180
GRA16 interacts with USP7 N-terminal domain through a MATH2 motif 181
A tool to study the effects of GRA16 during infection 181
Toward the study of the GRA16-USP7-PP2A complex 182

CONCLUSIONS 185

PRESENTATION OF THE RESULTS 187

KEY RESOURCES TABLE 189

TABLE OF FIGURES 191

TABLE OF TABLES 201

LIST OF ABBREVIATIONS 203

APPENDIX 205

BIBLIOGRAPHY 217

Abstract

Key words: Cell signalling, phosphorylation, MAPK, *Toxoplasma gondii*, host-pathogen interactions, protein-protein interaction

Protein-protein interactions are at the centre of cell signalling. The MAPKs (Mitogen-Activated Protein Kinases) are a family of intracellular protein kinases that form signalling cascades that control fundamental cell mechanisms such as proliferation, differentiation, inflammation and cell death. Signals progress through kinases (MAP4K to MAP3K to MAP2K) which eventually activate a MAPK. MAPK activation is achieved through the double phosphorylation of a TxY motif in its activation loop, and modulates the expression of genes through transcription factors or other protein kinases. The docking specificity of MAP2Ks towards MAPKs is mediated by Kinase Interacting Motifs (KIMs). Many X-ray structures exist of the individual kinases and complexes with isolated KIMs. However, due to the transient nature of the interaction between kinases, the molecular details of selectivity and activation of a MAPK by its upstream component are unknown.

Two important members of the MAPK family involved in the inflammatory response are p38α, a MAPK, and its activating MAP2K, MKK6. Through them, we present the first model of the interaction between a MAPK and its upstream activating MAP2K trapped in transition state. The model reveals a face-to-face heterokinase dimer, allowing the activation loop of p38α to access the active site of MKK6. Most interactions are distal from the phosphorylation sites, including the well-characterised MAP2K KIM – MAPK docking site interaction and a novel interaction interface between the C-lobes of the two kinases. This previously unobserved configuration is favourable for catalysis, and, despite its limited resolution, the cryoEM structure of MKK6 and p38α paves the way for hypotheses on MAPK activation.

We also discovered a novel non-canonical phosphorylation mechanism of p38α by MKK6 using ADP as a source of phosphate. We show that MKK6 is able to phosphorylate p38α using the β-phosphate of ADP and ATP *in vitro*, with lower efficiency than with the canonical nucleophilc attack of γ-phosphate of ATP. This alternative route of phosphorylation could confer an adaptive advantage to stress in case of ATP deprivation to maintain the signalling pathway.

Some pathogens hijack host signalling pathways, such as the MAPK pathway, to remodel the cellular response to their advantage. The intracellular parasite *Toxoplasma gondii* secretes effector proteins that mimic motifs present in some host proteins to directly act on their targets. We characterized how two of *T. gondii* effectors, GRA24 and GRA16, interact with their target protein.

GRA24 binds the host MAPKs p38α and ERK1/2 through a KIM motif; and GRA16 interacts with the host USP7 deubiquitinase through a MATH binding motif. Both motifs have high affinity for their targets and bind to known docking sites in competition with host interactors. We present the molecular details of the interaction between these pathogen motifs, and their host target, and propose some potential mechanism of action.

Résumé

Mots-clés : Signalisation cellulaire, phosphorylation, MAPK, *Toxoplasma gondii*, interactions hôtes-pathogènes, interactions protéine-protéine

Les interactions protéine-protéine sont au centre de la signalisation cellulaire. Les MAPKs (Mitogen-Activated Protein Kinases) sont une famille de protéines kinases intracellulaires qui forment des cascades de signalisation contrôlant des mécanismes cellulaires fondamentaux tels que la prolifération, la différenciation, l'inflammation et la mort cellulaire. Les signaux progressent à travers différentes kinases (MAP4K, MAP3K et MAP2K) qui finissent par activer une MAPK. L'activation des MAPKs se fait par la double phosphorylation d'un motif TxY situé dans sa boucle d'activation, et module l'expression de gènes par l'intermédiaire de facteurs de transcription ou d'autres protéines kinases. La spécificité de liaison des MAP2Ks aux MAPKs est médiée par des motifs nommés KIM (pour Kinase Interacting Motif). Il existe de nombreuses structures aux rayons X des kinases individuelles et en complexe avec des KIMs isolés. Cependant, en raison de la nature transitoire de l'interaction entre les kinases, les détails moléculaires de la sélectivité et de l'activation d'une MAPK par sa MAP2K sont inconnus.

Deux membres importants de la famille MAPK impliqués dans la réponse inflammatoire sont p38α, une MAPK, et sa MAP2K activatrice, MKK6. A travers eux, nous présentons le premier modèle de l'interaction entre une MAPK et sa MAP2K activatrice amont piégée à l'état de transition. Le modèle révèle un dimère de kinases face à face permettant à la boucle d'activation de p38α d'accéder au site actif de MKK6. La plupart des interactions sont éloignées par rapport aux sites de phosphorylation. En plus de l'interaction bien caractérisée entre le KIM de la MAPK2K et le "docking site" de la MAPK, une nouvelle interface d'interaction entre les lobes C des deux kinases a été indentifiée. Cette configuration jusqu'alors inobservée est favorable à la catalyse et, malgré sa résolution limitée, le modèle cryoEM de MKK6 et p38α ouvre la voie à des hypothèses sur l'activation des MAPK.

Nous avons également découvert un nouveau mécanisme de phosphorylation non canonique de p38α par MKK6 utilisant l'ADP comme source de phosphate. Nous montrons que MKK6 est capable de phosphoryler p38α en utilisant le β-phosphate de l'ADP et de l'ATP *in vitro*, avec une efficacité moindre qu'avec l'hydrolyse canonique du γ-phosphate de l'ATP. Cette voie alternative de phosphorylation pourrait conférer un avantage adaptatif au stress en cas de privation d'ATP pour maintenir la voie de signalisation.

Certains pathogènes détournent les voies de signalisation de leur hôte pour remodeler la réponse cellulaire à leur avantage. Le parasite intracellulaire *Toxoplasma gondii* sécrète des protéines effectrices qui imitent des motifs présents dans certaines protéines de l'hôte pour agir directement sur leurs cibles. Nous avons caractérisé comment deux effecteurs de *T. gondii*, GRA24 et GRA16, interagissent avec leur protéine cible.

GRA24 se lie aux MAPKs de l'hôte p38α et ERK1/2 par le biais d'un motif KIM ; et GRA16 interagit avec la déubiquitinase hôte USP7 par le biais d'un motif de liaison MATH. Les deux

motifs ont une haute affinité pour leur cible et se lient à des sites d’interaction connus en compétition avec les interacteurs de l'hôte. Nous présentons les détails moléculaires de l'interaction entre ces motifs pathogènes et leur cible et proposons un mécanisme d'action potentiel.

General audience abstract

To react to external stress and trigger the appropriate response, cells possess a complex signalling network of interacting proteins. In this thesis, we focus on the proteins MKK6 and p38α which play an important role in the inflammatory process.

When MKK6 is activated in response to stress, it in turn activates p38α which can then activate other proteins to trigger inflammation. We present the first molecular model of the interaction between MKK6 and p38α, revealing important details about this activation mechanism.

These signalling networks can be hijacked by pathogens. The parasite *Toxoplasma gondii* secretes proteins that mimic those of the host to remodel the cellular response to its advantage. We present how two of these proteins, GRA16 and GRA24, interact with their target proteins.

Résumé de thèse vulgarisé

Pour réagir aux stress extérieurs et déclencher les réponses appropriées, les cellules possèdent un réseau de signalisation complexe formé de protéines interagissant les unes avec les autres. Dans cette thèse, nous nous intéressons aux protéines MKK6 et p38α ayant un rôle important dans le processus inflammatoire.

Quand MKK6 est activée en réponse à un stress, elle active à son tour p38α qui peut ensuite activer d'autres protéines pour déclencher l'inflammation. Nous présentons le premier modèle moléculaire de l'interaction entre MKK6 et p38α révélant des détails importants sur ce mécanisme d'activation.

Ces réseaux de signalisation peuvent être détournés par des pathogènes. Le parasite *Toxoplasma gondii* sécrète des protéines imitant celles de l'hôte pour remodeler la réponse cellulaire à son avantage. Nous présentons comment deux d'entre elles, GRA16 et GRA24, interagissent avec leur protéine cible.

Introduction

Signalling pathways

Kinases at the centre of cell signalling

Communication and information transfer are crucial for correct cell, tissue and organism functions. Throughout evolution, different mechanisms have appeared to fulfil these important missions and form a highly complex signalling network where proteins are the main components.

Proteins, especially enzymes, can exist in a number of states. The transition from one state to the other (for example from inactive to active) is induced by conformational changes triggered by post-translational modifications (PTMs), protein complex assemblies, and/or transient protein-protein interactions. One of the most relevant PTMs in cell signalling is phosphorylation, which leads to crucial effects on the protein state, e.g. activation, inhibition or creation of new binding sites (Ardito *et al.*, 2017). Kinases are the enzymes catalysing such reaction and in addition, in the same cell signalling pathway, multiple kinases are involved either as enzyme or substrate.

Kinases are present in all organisms across evolution. They have the property to transfer a phosphate group from a donor (usually an ATP molecule) to a substrate. This is achieved by the correct placement of substrates for nucleophilic attack and the balancing of opposing charges accumulating on the substrates with positively charged amino acids and cationic co-factors (Ardito *et al.*, 2017). This enables kinases to achieve enormous rate enhancements for the cleavage of the phosphodiester bond. In prokaryotes, kinases are mostly involved in energy transduction but in the eukaryotes, protein kinases evolved as important modulators of function as the modifications can be easily reversed. The eukaryotic cell has evolved a complex and highly regulated signalling network to integrate a multitude of stimuli using phosphorylation catalysed by kinases as the tool to transduce the signals. They represent 1.7% of all human genes (Manning, 2002) and more than 538 human kinases (kinase.com) have been described so far. They can be classified into families based on evolutionary sequence conservation as represented in the human kinome phylogenetic tree (Figure 1) (Manning, 2002). Kinases catalytic properties are dependent on their structure that is highly conserved across the family.

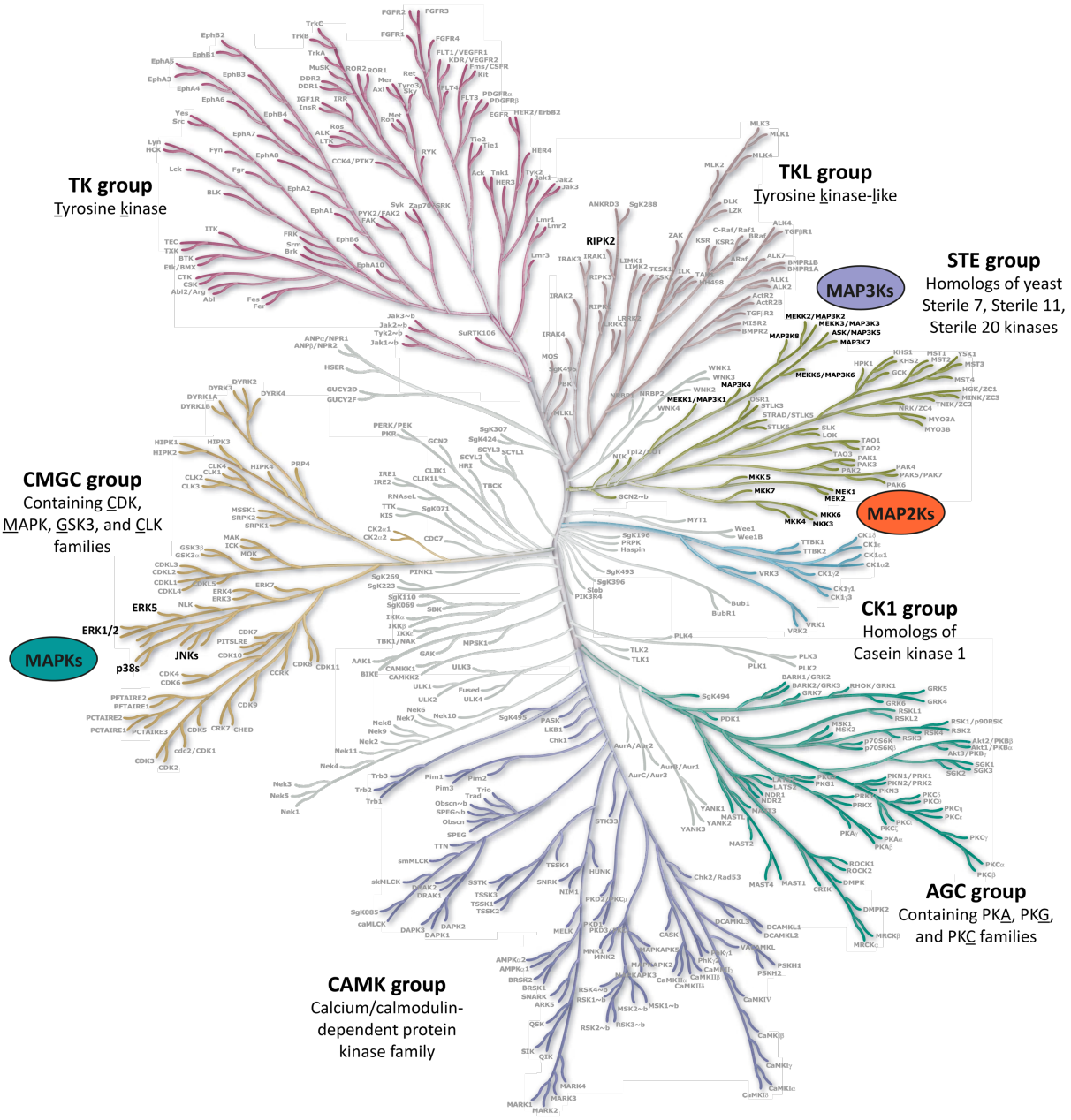


Figure 1: The human kinome phylogenetic tree (adapted from Cell Signaling Technology poster).

This thesis is focused on the signalling pathways involving the mitogen-activated protein kinases (MAPKs) and their upstream activators.

The canonical MAPK pathways

The mitogen-activated protein kinases (MAPKs) are a family of intracellular kinases that control fundamental mechanisms such as proliferation, differentiation, inflammation and cell death (Chen *et al.*, 2001; Kyriakis JM, 2001; Peti and Page, 2013). The canonical pathway is

constituted of three consecutive kinases: the MAPK kinase kinase (MAP3K) is phosphorylated in response to *intra-* and *extra*-cellular stimuli (growth factors, cytokines, cell stress and others) and in turn phosphorylates a MAPK kinase (MAP2K) which finally phosphorylates the effector MAPK (Figure 2). This activated MAPK can phosphorylate a number of substrates and triggers the activation of transcription factors, leading to gene activation and specific intracellular responses. Some MAP4Ks also serve as activators of the MAP3Ks. Although this linear pathway representation of the MAPK family is the most common one, the reality is more complex and includes cross-talk between these pathways, feedback loop regulation and potential scaffolding proteins (Zhang, Wu and Wang, 2011; Meister *et al.*, 2013; Ferrell and Ha, 2014).

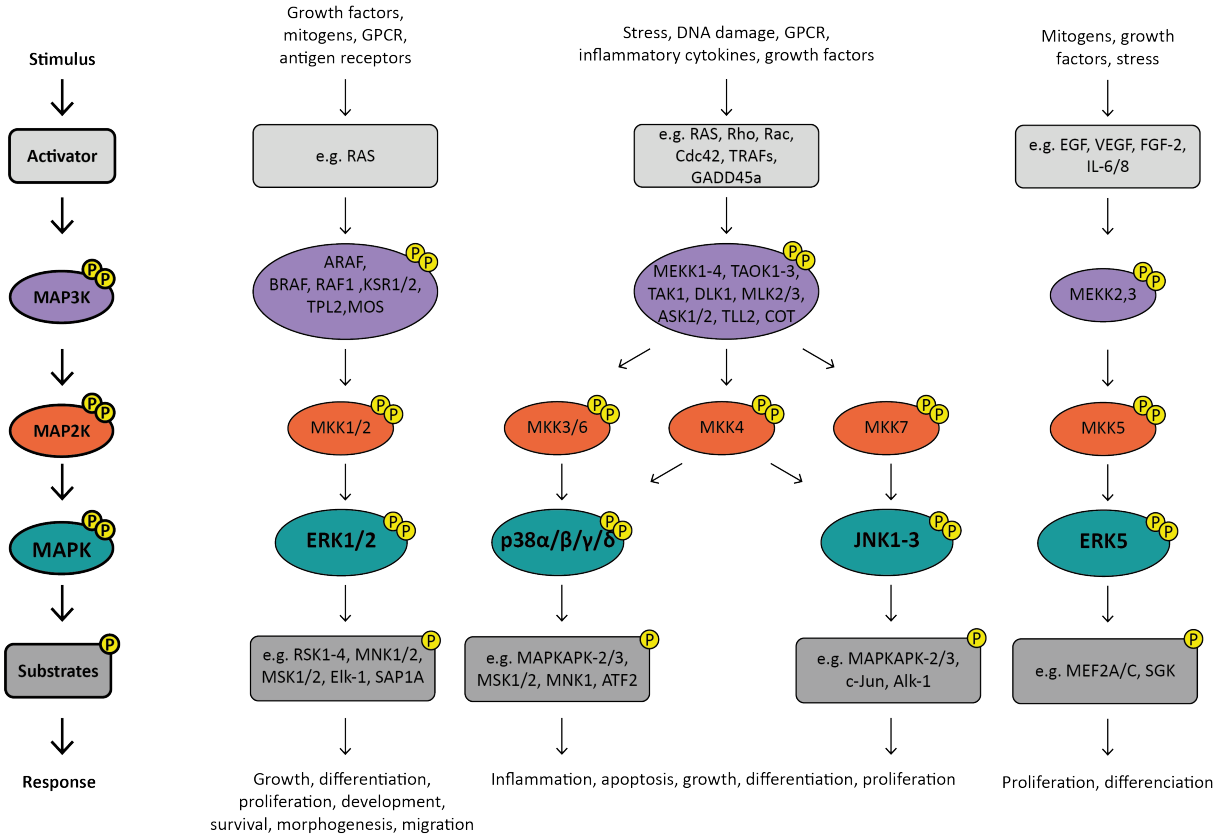


Figure 2: The MAPK pathways. The canonical pathway starts with the integration of specific stimuli by the MAP3K by activator proteins and triggers the MAP3K activation by (auto)phosphorylation. The MAP3K phosphorylates its downstream MAP2K, that in turn phosphorylates the MAPK. The activated MAPK phosphorylates a number of substrates to induce the appropriate response to the initial stimulus. The main MAPK families, ERK1/2, p38s, JNKs and ERK5, are represented with their upstream activating MAP2Ks and MAP3Ks, as well as with a selection of stimuli, pathway activators and substrates.

Mammalian MAPKs are comprised of four main families (Figure 2) (Cobb and Goldsmith, 1995; Nithianandarajah-Jones *et al.*, 2012; Peti and Page, 2013):

- the p38 MAPKs (isoforms α, β, δ and γ)
- the extracellular signal-regulated kinases 1/2 (ERK1/2),
- the c-Jun amino-terminal kinases (JNKs)
- the extracellular signal-regulated kinase 5 (ERK5)

The MAPKs are serine/threonine-specific and are part of the CMGC group (named after the initials of the key kinases CDK, MAPK, GSK3 and CLK) (Figure 1, p22). Their direct upstream activators, the MAP2Ks, are dual-specificity kinases coming from the STE branch (homologs of the yeast Sterile kinases). MAP3Ks and MAP4Ks are serine/threonine-specific and come from the STE group and both the STE group and the TKL group (Tyrosine kinase-like) respectively.

The MAPK cascades are evolutionary conserved amongst eukaryotes (Caffrey, O'Neill and Shields, 1999; Widmann *et al.*, 1999). The yeast *Saccharomyces cerevisiae* contains MAPK cascades including Hog1, related to the p38 and JNK families, and Fus3 and KSS1, related to the ERK1/2 family (Figure 3).

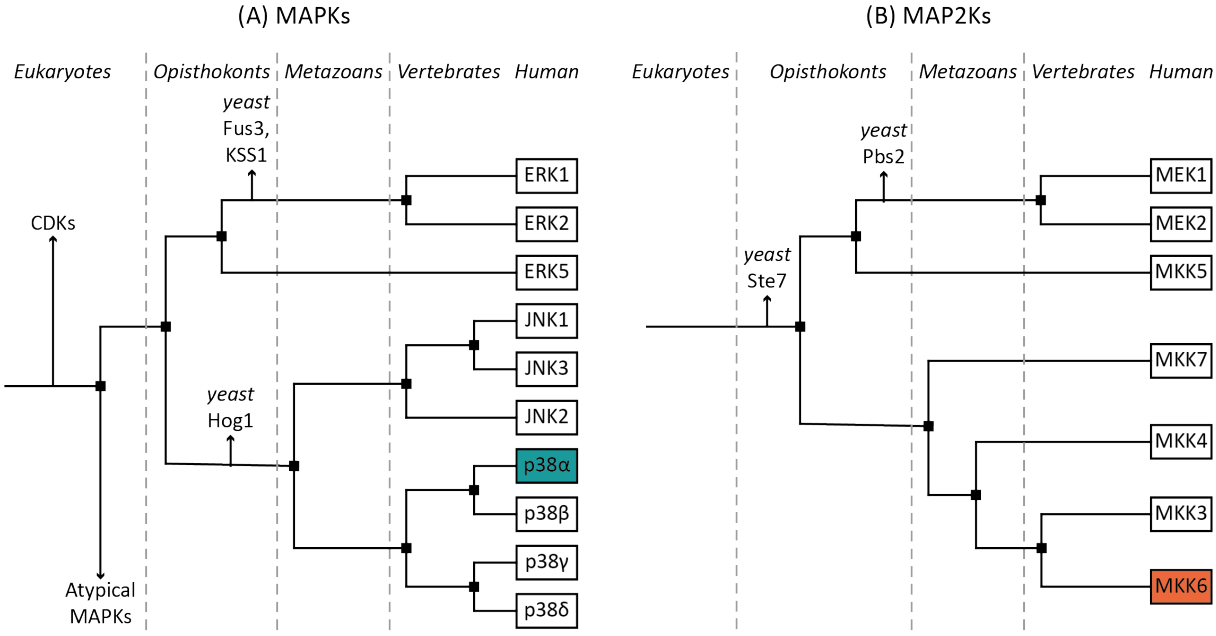


Figure 3: Evolutionary tree of human MAPKs (A) and MAP2Ks (B). Divergent branch corresponding to the main yeast homologues are indicated. The sequence alignment was performed using Clustal Ω (Sievers and Higgins, 2018). p38α and MKK6, the main objects of study of this work, are coloured.

The multiple roles of p38α in cell signalling

p38α cellular functions

p38α has been identified and described in 1994 in four independent studies:

- as p38, a 38kDa protein phosphorylated in response to LPS exposition and osmolarity changes (Han *et al.*, 1994)
- as RK, for MAPKAP kinase-2 reactivating kinase, a kinase that activates MAPKAP kinase-2 in response to sodium arsenite, heat shock and osmotic stress (Rouse *et al.*, 1994)
- as p40, a kinase activated by IL-1 (Freshney *et al.*, 1994)

- as CSBP2, for CSAID binding protein, the target of an inhibitor (CSAID = cytokine-suppressive anti-inflammatory drug) of IL-1 TNF-α production in response to LPS (Lee *et al.*, 1994)

These four proteins all correspond to p38α (MAPK14). Three other isoforms, named p38β, p38δ and p38γ were identified later (Cuadrado and Nebreda, 2010). p38α is essential for normal development, as p38α knock-out results in an embryonic lethal phenotype (Ihle, 2000; Mudgett *et al.*, 2000) whereas the other isoforms seem to be compensated as knock-outs, even in combination, and result in normal or near-normal adult phenotypes (Cuenda and Sanz-Ezquerro, 2017). The functions of p38β are mostly redundant with those of p38α. The expression of p38δ and p38γ is tissue-dependent and they are involved in similar functions, such as regeneration and the immune response (Cuenda and Sanz-Ezquerro, 2017).

The p38 MAPKs, and in particular p38α, are key actors in signal transduction during different physiological events. They mainly manage the response to environmental stresses and inflammatory signals. p38α is ubiquitously expressed and contributes to some general cellular functions but also takes part in some specific functions in specific cell types (e.g. in adipocytes, haematopoietic stem cells, neurons, hepatocytes) (Beardmore *et al.*, 2005; Canovas and Nebreda, 2021).

A wide variety of stressful conditions activate p38α. Several of its upstream MAP3Ks work as redox sensors and detect reactive oxygen species (Sakauchi *et al.*, 2017; Matsushita *et al.*, 2020). Similar to its yeast homologue HOG1 (High osmolarity glycerol 1), p38α also responds to hyperosmolarity (Han *et al.*, 1994). Additionally, bacterial and viral infections activate p38α through Toll-like receptors (TLRs) that recruit the MyD88 adaptor and IRAK kinases, TRAFs and TAK1. Pro-inflammatory cytokines also activate p38α: TNF, IL-6 and IL-1β are detected by cell receptors and activate p38α by intermediates such as TRAFs and TAK1. In addition to these specific stress responses, ionising radiation also induces the pathway (Cho *et al.*, 2016).

In response to damages, p38α can induce cell cycle arrest to allow reparation mechanisms of the induced damage, including DNA repair or telomere maintenance, to act. It can also trigger the production of pro-inflammatory cytokines (Lee *et al.*, 1994). Alternatively, p38α can also induce cell death or autophagy (Slobodnyuk *et al.*, 2019).

MKK6, the activator of p38α

Specific MAP2Ks activate p38: MKK3 and MKK6, but also MKK4, shared with the JNK pathway (Wang *et al.*, 2001). In this thesis we focus on the MAP2K MKK6 (MAP2K6).

MKK6 was first described by (Han *et al.*, 1996) as a p38-specific MAP2K. It is closely related to MKK3 with whom it shares 80% amino acid sequence homology. MKK6 exists in two isoforms: isoform 2 is missing the first 56 residues of isoform 1, and is restricted to skeletal muscles, whereas isoform 1 is widely expressed across tissues (Han *et al.*, 1996).

MKK6 is the main activator of p38α in case of osmotic stress (Moriguch *et al.*, 1996), as well as in response to LPS, TNF-α, or IL-1 (Cuenda *et al.*, 1996). It is also required for specific STAT4 activation through p38α activation in response to IL-12 (Visconti *et al.*, 2000).

p38α and diseases

Dysregulation of MAPK activity has been associated with a broad spectrum of pathologies, from cancer and inflammatory diseases to obesity and diabetes (Lawrence *et al.*, 2008; Burton *et al.*, 2021; Canovas and Nebreda, 2021). MAPKs are therefore key molecular targets for therapeutic intervention. For instance, p38α over-activation has been reported in rheumatoid arthritis, where it contributes to abnormal inflammation (López-Santalla *et al.*, 2011). The activation of p38α that accompanies myocardial ischemia aggravates the injury (Kumphune, Chattipakorn and Chattipakorn, 2012; Song *et al.*, 2020). Upregulation of the pathway has also been implicated in severe disease related to infection. A particularly relevant example is the severe effects of inflammation in patients suffering from COVID-19, caused by the SARS-CoV-2 virus, where the p38 pathway has been shown to be involved.

A recent review from Canovas and Nebreda (Canovas and Nebreda, 2021) offers a broad overview of the diversity of roles that p38α accomplishes in normal physiological conditions and in diseases (Figure 4).

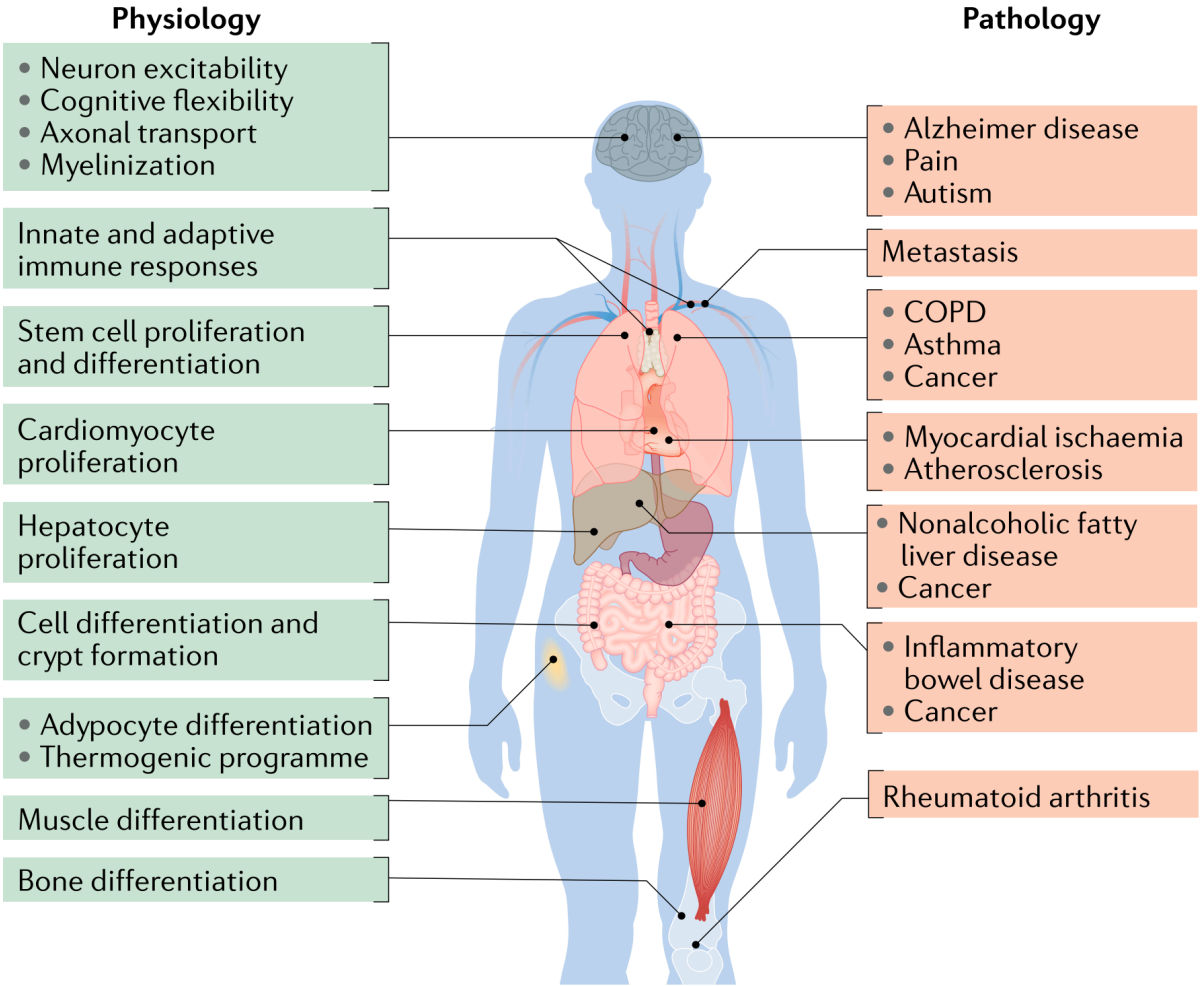


Figure 4: Diversity of p38α roles in health and disease, from (Canovas and Nebreda, 2021).

Structural details of the p38α and MKK6 kinases

Protein kinase shared structural characteristics

Protein kinases share key structural elements and conserved sequence elements linked to their function (Figure 5 and Figure 6) (Taylor and Radzio-Andzelm, 1994; Endicott, Noble and Johnson, 2012; Beenstock, Mooshayef and Engelberg, 2016; Modi and Dunbrack, 2019a). Some other characteristic structural features are specific to sub-families of kinases.

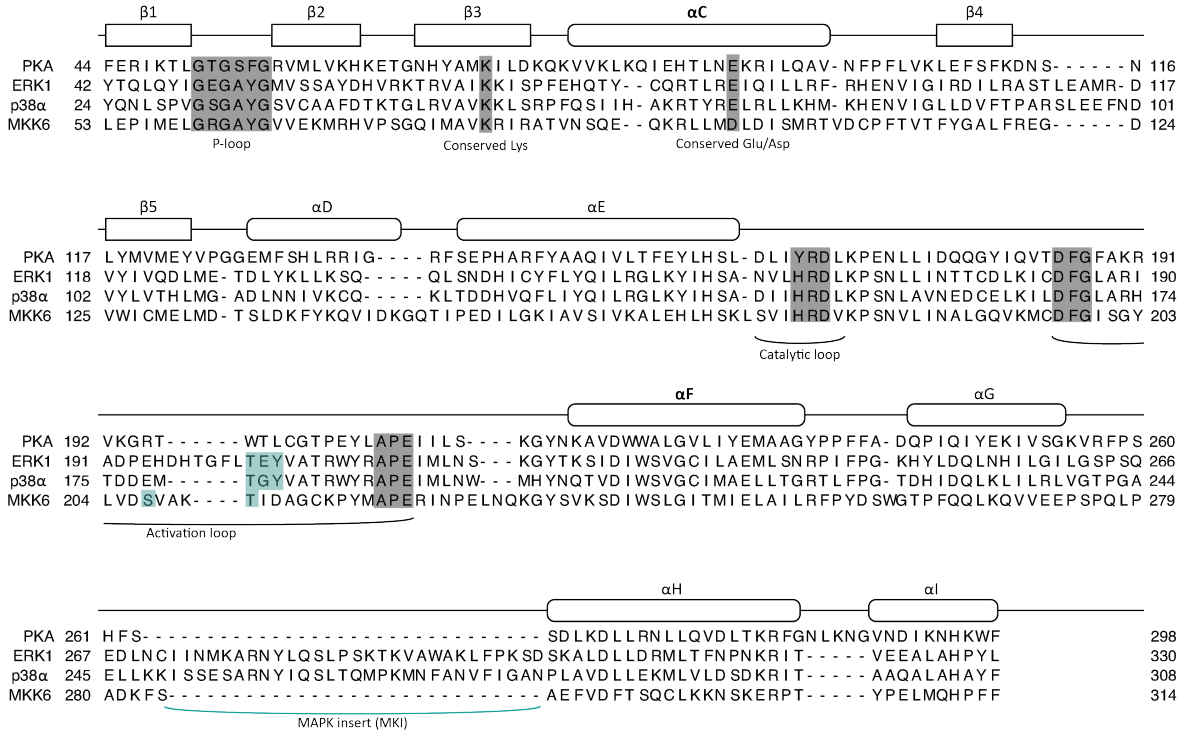


Figure 5: Structure-based sequence alignment of the kinases of interest in this work (ERK1, p38α and MKK6) relative to the reference kinase PKA. The alignment was published in Modi and Dunbrack, 2019a. Common kinase secondary structure with the kinase nomenclature are shown. Key conserved motifs and residues are highlighted in grey. Specific features of MAP2K and MAPK are indicated in teal.

Protein kinases have a bi-lobed structure with an N-terminal domain, constituted of a 5-stranded β-sheet and a single α-helix named the "**αC-helix**", and a C-terminal domain mainly composed of α-helices (Taylor and Radzio-Andzelm, 1994; Endicott, Noble and Johnson, 2012; Beenstock, Mooshayef and Engelberg, 2016). The two lobes are connected through a hinge that delineates the cleft of the ATP-binding site. The C-terminal lobe include the **activation**

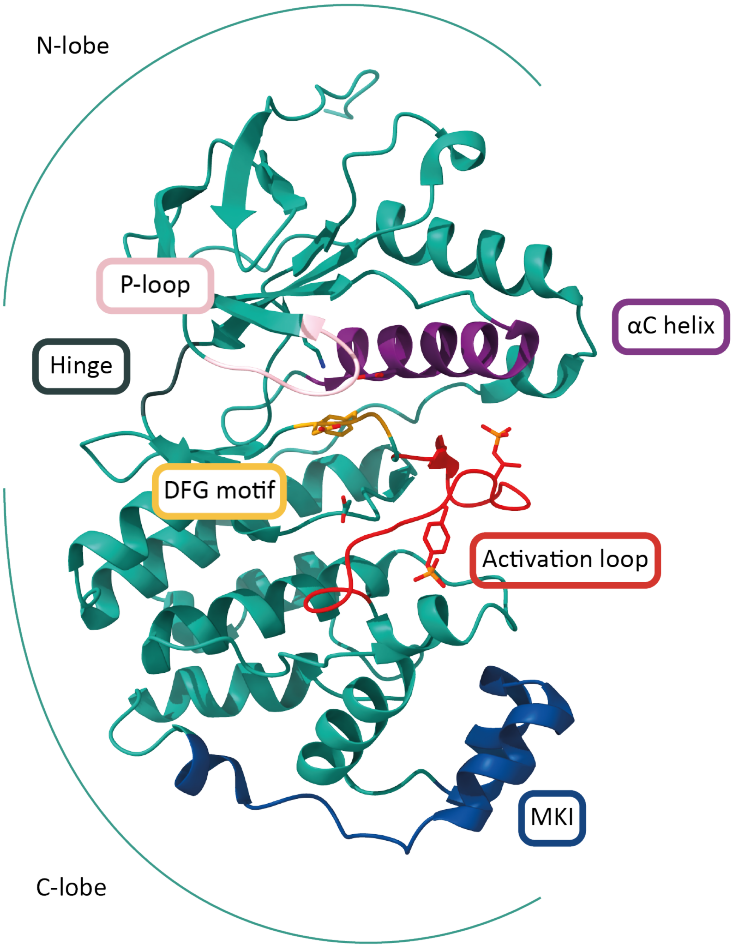


Figure 6: Kinase structural elements. The double phosphorylated p38γ structure (PDB 1cm8) is coloured according to conserved structural features. The β-sheet containing N-lobe and the α-helical C-lobe are connected through a hinge (dark grey). In the N-lobe, the P-loop (pink) and the αC-helix (purple) are essential for ATP binding and active kinase conformation. In the C-lobe, the activation loop (red) is phosphorylated on both residues of the conserved MAPK TxY motif. It is preceded by the DFG motif, important for ATP binding. The MAPK insert (MKI in blue) is specific to the MAPK family.

loop, a region composed of 25-30 residues, delimited by a conserved **DFG motif** and an **APE motif**. The **F-helix** is located at the core of the C-terminal domain, and is extremely hydrophobic.

The phosphorylation of a MAPK target can only occur when the kinase is in an active conformation: the catalysis of the transfer of the γ-phosphate group of ATP to the hydroxyl group of the phospho-acceptor residue of a substrate is made possible by the correct positioning of some conserved key features.

To form the ATP-binding site, the two lobes need to be in a closed conformation (Figure 7). A salt-bridge is formed between a **conserved glutamate** in the αC-helix (can be substituted by an aspartate in some kinases) and a **conserved lysine** in β3-strand. The ATP molecule is coordinated by several elements. The adenine ring forms van der Waals contacts with nonpolar groups from the pocket, as well as specific hydrogen bonds between N1 and N6 and the peptide backbone of the hinge region. The side chain of the **conserved lysine** of β3-strand forms hydrogen bonds with the α- and β-phosphate groups. The **aspartate of the DFG motif** and a **conserved asparagine** chelate two Mg^{2+} ions to orientate the ATP substrate and interact with the β-phosphate. A **glycine-rich loop** (containing a GxGxxG motif) also named **P-loop or ATP-binding loop**, is located between β1- and β2-strand, and stabilizes the β- and γ-phosphates by closing on top of the ATP.

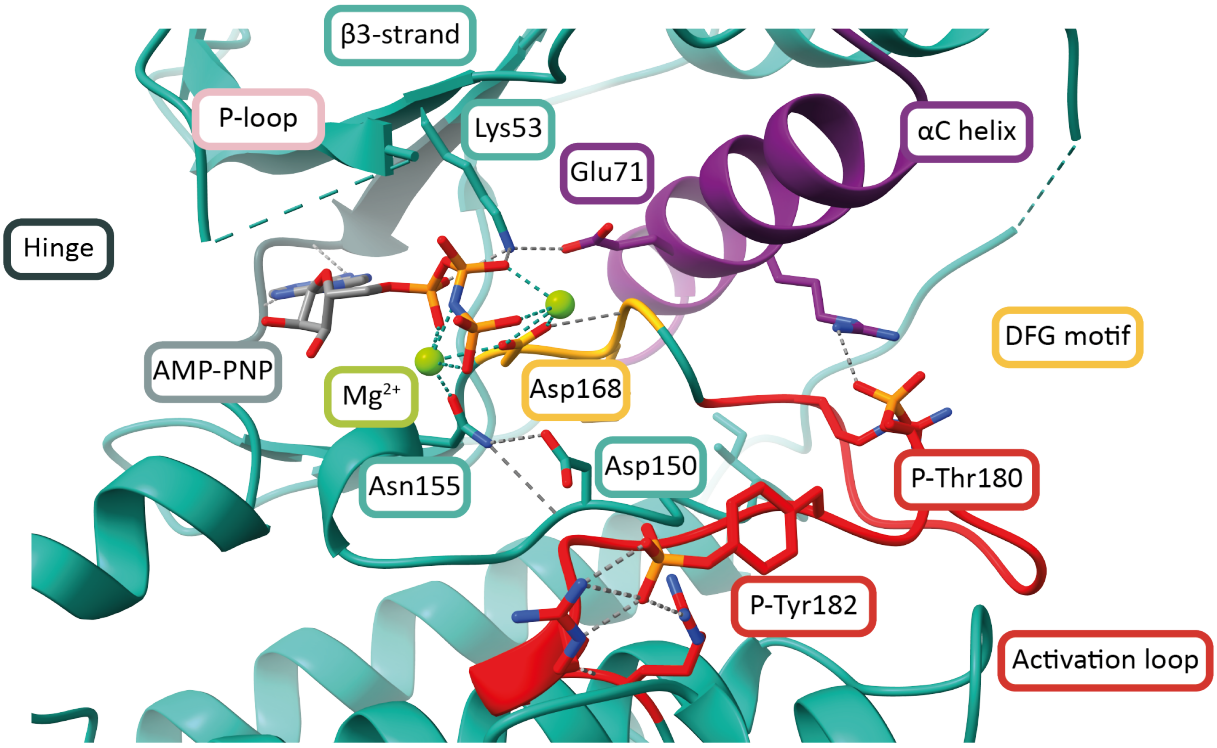


Figure 7: The ATP binding site. The biphosphorylated p38γ structure bound to the ATP analogue AMP-PNP (PDB 1cm8) is coloured according to conserved structural features. The numbering is based on p38α sequence. Lys53 from the β3-strand and Glu71 from the αC-helix (in purple) form a salt-bridge. The AMP-PNP nucleotide (in light grey) binds to the ATP binding site, forming H-bonds with the hinge (grey). The phosphates are coordinated by two Mg^{2+} ions and several key residues: Lys53, Asp168 from the DFG motif (yellow), Asn155. The catalytic aspartate Asp153 from the HRD motif and the Lys155 are important for the binding of the phospho-acceptor substrate and the catalysis of the reaction. The P-loop is disordered in this structure (pink). The two phosphorylated residues Thr180 and Tyr182 in the activation loop (in red) interact with positively charged residues in both lobes, maintaining the close conformation.

The phospho-acceptor substrate is oriented by the **conserved catalytical aspartate** residue from a **HRD motif** (or YRD in some kinases) in the **catalytic loop.**

The active fold is often triggered by the phosphorylation of the **activation loop**, where phosphate group(s) will stabilize the conformation. MAPKs have a specific **Thr-x-Tyr motif** in their activation loop, in which both the threonine and tyrosine are phosphorylated by a MAP2

kinase in order to activate them. The activation loop also contributes to the kinase specificity towards Ser/Thr or Tyr. In serine-threonine kinases, a **lysine residue** (two residues away from the catalytic aspartate) interacts with the γ-phosphate and serves to stabilize the local negative charges during catalysis. In tyrosine kinases, an **arginine residue** (4 residues away from the catalytic aspartate) stabilizes the same way the negative charges, while allowing the larger tyrosine residue in.

A so-called **P+1 loop**, downstream of the activation loop, accommodates the residue located one position down of the phosphorylated site of the substrate. This residue is important for substrate selectivity.

The phosphoryl transfer event is dependent on the proper alignment of the γ-phosphate of ATP, and the hydroxyl group of the phospho-acceptor residue of the substrate. The HRD-catalytic aspartate activates the hydroxyl group by accepting a proton. Phosphorylation results from the nucleophilic attack of this hydroxyl group on the ATP γ-phosphate, leading to an inversion of the phosphorus. The Mg^{2+} ions stabilize the negative charges that develop during the catalysis and aid the departure of the leaving group ADP. The transition state of the phosphoryl transfer is described to be either dissociative or associative. If dissociative, a metaphosphate intermediate is created, when the bond to the leaving group is broken before the new bond to the acceptor is formed. If associative, a phosphorane pentavalent intermediate is formed between the donor and acceptor. Phosphoryl transfer is a transient reaction, limited by the release of its two products, ADP and the phosphorylated protein. In the enzymatic reaction, the transition state scenario is probably an average between these two extremes.

Evaluating the state (active or inactive) of a kinase based on its structure is not straightforward. Several key elements are necessary for the activity of active kinases, but not sufficient on their own:

- The DFG motif should be in the **DFG-in** conformation, so the aspartate can coordinate the ATP molecule and a Mg^{2+} ion. On the contrary, the **DFG-out** conformation is characterised by its phenylalanine turned in toward the active site, blocking it.
- The salt-bridge between the glutamate of the αC-helix and the lysine from β3-strand must be formed
- A denominated **regulatory spine** is assembled by four hydrophobic residues: the HRD-His, the DFG-Phe, the 4th residue after the αC-helix-Glu and a residue in the loop before strand β4. It is considered to be present if the distances between side-chain pairs are less than 4.5Å and is spatially conserved in active kinases (Kornev and Taylor, 2010).

Inactive kinases are structurally diverse, sometimes presenting features from active kinases described above (Noble, Endicott and Johnson, 2004; Modi and Dunbrack, 2019b).

In some kinases, N- and C-terminal extensions can be involved in regulation. For example, the activation loop or other key features can be protected by regulatory domains.

Protein kinases are able to phosphorylate a number of residues and they are also able to discriminate between their targets. This specificity results from different mechanisms that are not completely elucidated. Frequently, certain residues are required for recognition around the phosphorylation site. For example, MAPKs are proline-directed, meaning that they

recognize their substrate phosphorylation sites at Ser and Thr residues that are followed by a proline residue at the P+1 position in most of the cases, they also favour substrates in which a proline is also present at the P-2 position.

The substrate phosphorylation sites are often located in less ordered regions, and/or exposed on the surface of the protein. Specificity for substrates is mostly determined by docking sites located away from the active site.

MAPK specificities

Docking site and KIM motifs

MAPKs share all the general kinase structural features presented so far, but also have some family specific characteristics.

Specificity between MAPK binding partners, such as upstream activating MAP2Ks, substrates, scaffolds and phosphatases, is mediated via a **docking site** for binding partners, located distal to the active site. It is composed of two parts: the **CD domain** (common docking domain, also referred as acidic patch), composed of negatively charged residues, and a **hydrophobic groove** (sometimes referred as the ED site, for the Glu160 and Asp161 residues in p38α) (Figure 8).

Binding partners, including the upstream MAP2Ks, but also substrates, MAPK-inactivating phosphatase (MKPs) and scaffold proteins, contain **Kinase Interacting Motifs** (**KIMs**, also referred as **D-motif** (for docking motif)) that interact with the MAPK docking site (Sharrocks, Yang and Galanis, 2000; Garai *et al.*, 2012; Zeke *et al.*, 2015). The KIM is a short linear motif (SLiM) of 7 to 18 amino acids, typically found in intrinsically disordered regions of the proteins, and adopts a stable conformation upon binding. It is composed of at least a positively charged amino acid (named θ), connected through a linker to a series of hydrophobic residues, usually three, named φ_L, φ_A and φ_B.

The general consensus motif sequence can be represented as follows:

$$(\varphi_U\text{-}x_{1\text{-}2})_{optional} - \theta_{1,2} - linker(x_{0\text{-}6}) - \varphi_L - spacing(x_{1,2}) - \varphi_A - x - \varphi_B$$

where φ_U, φ_L, φ_A and φ_B are hydrophobic amino acids (mostly Leu, Ile or Val, but also Met, Phe and Pro), θ denotes a positively charged amino acid (Lys or Arg) and x can be any amino acid.

Stricter motifs have been defined and partially explain the specificity between MAPK families and their binding partners (Table 1). The different classes of motif can be separated between the ones that target JNK, and those that target ERK and p38. The CD region of ERK and p38 is wider compared to JNK where it lies closer to the hydrophobic groove. The JNK CD surface is

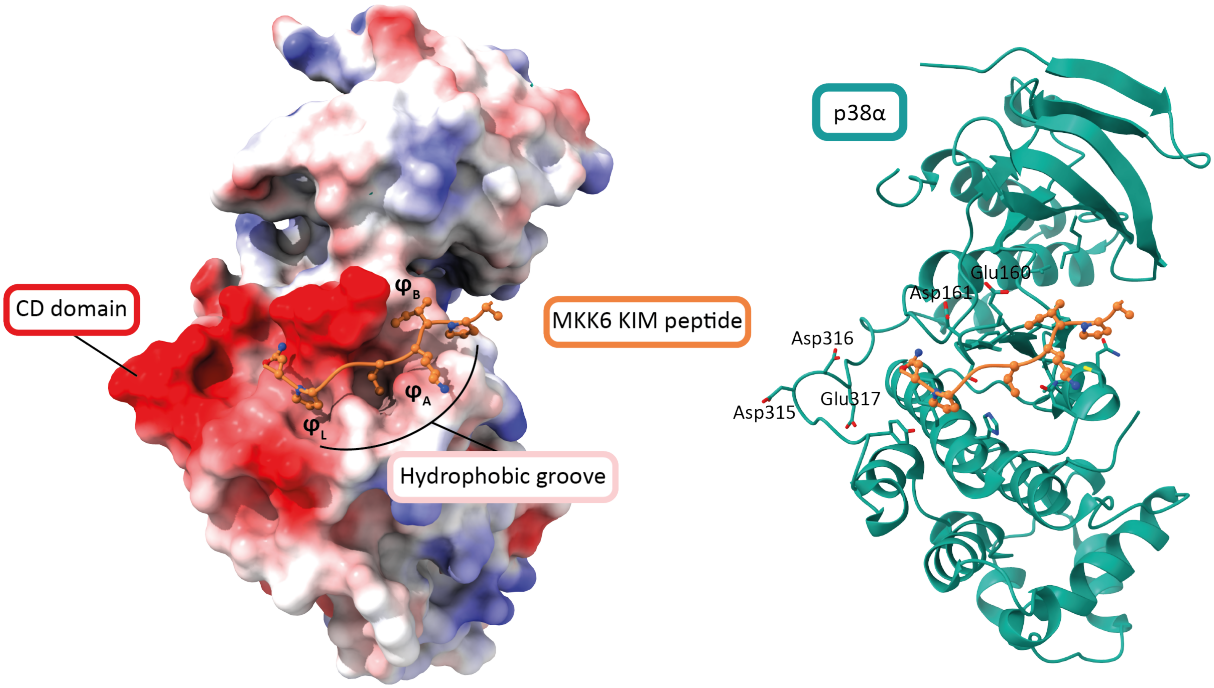


Figure 8: Structure of the MKK6 KIM peptide bound to the docking site of p38α (PDB 2y8o). MKK6 KIM peptide is represented as cartoon with balls and sticks side chains. Left: The surface of p38α is displayed coloured by its electrostatic potential (from -10 (red) to 10 kTe^{-1} (blue)). Right: p38α is represented as cartoon with key residues as sticks.

also less negatively charged compared to the other MAPKs, requiring only one basic residue θ for binding, when larger variability is observed at the N-terminus of KIMs of ERK and p38 binding partners. These differences are at the source of the specificity. However, given the homology of ERK and p38 docking sites, the KIMs are not sufficient to explain the discrimination between these two families.

Table 1: KIM classes

Class	Motif	Comments	Specificity		Examples
			JNK	ERK and p38	
JIP1 class	θ-φ_L-x-x-φ_A-x-φ_B	φ_L is usually a proline and the linker should not contain prolines	x		
NFAT4 class	θ-x-x-φ_L-x-φ_A-x-φ_B	The basic part of the motif adopts a small alpha-helical conformation binding to the CD domain typical to this class, and Pro is strongly disfavoured in this region.	x		MKK4 MKK7
DCC class	*DCC subtype* $\theta_{(1-2)}$-$x_{(2-4)}$-φ_L-P-x-φ_A-x-φ_B *Far1 subtype* $\theta_{(1-2)}$-$x_{(2-4)}$-φ_L-x-P-φ_A-x-φ_B	DCC motifs partly adopt a type II polyproline helix-like geometry, that require one or two prolines at the φ_L and spacer position.		x	MEK1
MEF2A class	*MEF2A subtype* θ-θ-$x_{(2)}$-φ_L-x-φ_A-x-φ_B *MKK6 subtype* θ-θ-$x_{(3)}$-φ_L-x-φ_A-x-φ_B	The C-terminal part is usually locked in a single conformation binding to the hydrophobic groove, while the basic N-terminus is usually invisible to X-ray crystallography probably because of multiple conformations. Most of the known motifs of the MKK6 subtype have a conserved Pro at the φ_B+1 position that can form an additional hydrophobic interaction.		x	MKK6 MKK3 MKK4 MEK2
HePTP class	*Ste7 subtype* φ_u-x-θ-θ-$x_{(4)}$-φ_L-x-φ_A-x-φ_B *HePTP subtype* φ_u-x-θ-θ-$x_{(5)}$-φ_L-x-φ_A-x-φ_B	The HePTP class present a longer helical N-terminus with an additional hydrophobic aa φ_U. The basic region is α-helical (Pro is disfavoured) and a conserved glycine is required to terminate the α-helix in the HePTP subtype. The internal linkers are longer.		x	
RevD class	φ_B-x-φ_A-$x_{(2)}$-φ_L-$x_{(4-6)}$-φ_u-$x_{(2)}$-$\theta_{(2)}$	Reverse motif, with an helical end. Present in RSK and MAPAPKs		x	
GRA24 class	φ_u-$x_{(2)}$-θ-θ-G-$x_{(3-6)}$-φ_L-(P-x/x-P)-φ_A-x-φ_B	Combination of DCC-type (partly polyproline II helix-like) geometry for the hydrophobic segment with HePTP-like helical terminus.		x	*T. gondii* GRA24

Hydrophobic pocket/DEF site/lipid-binding site

A second MAPK binding site is located in the C-lobe. Two helices connected by a short loop interrupt the canonical kinase C-terminal domain and constitute the **MAPK insert** (**MKI**) (Figure 5, p27 and Figure 6, p28). This insert is common to all MAPKs and to the CDK2 related kinases. It delimitates a **hydrophobic binding site** present in some conformations of ERK1/2 and p38 kinases, consisting in a lipophilic pocket mostly decorated with hydrophobic amino acid side chains. This binding site has been described in different terms.

In ERK1/2, it was originally named **DEF site** for "docking site for ERK, FXF" (Jacobs *et al.*, 1999). Binding partners interact with this docking site through an FxF(P) motif, located within 20 residues downstream of their phosphorylation site. It was believed to be relevant and bind to FxF(P) motifs in ERK1/2 and p38α, but not in p38β, δ, γ or JNK (Galanis, Yang and Sharrocks, 2001). The FxF(P) motif might act to stabilize the binding, either working cooperatively or subsequently to KIM motif binding to the active site to recruit specific substrate.

In p38 MAPKs, it has been characterised as a **lipid-binding site** where it was initially observed through the binding of n-octyl-β-D-glucoside (β-OG) in X-ray crystallography structures (Diskin, Engelberg and Livnah, 2008; Tzarum *et al.*, 2012; Bührmann, Hardick, *et al.*, 2017; Bührmann, Wiedemann, *et al.*, 2017). It can bind lipids, but it is also suggested to be an allosteric binding site that might be involved into fine-tuning of kinase activity.

MAPK extension

MAPKs also share a unique C-terminal region named the **MAPK extension**. It consists in two α-helices that extend from the C-lobe and interact with the N-lobe. The interaction between the αL16-helix of the MAPK extension and the αC-helix seems to have a role in the repression of autophosphorylation of p38α (Tesker *et al.*, 2016).

MAP2K specificities

MAP2Ks are dual specificity kinases. They are part of a different branch of the kinome tree, being the homologues of yeast sterile Ste kinases (Figure 1, p. 22).

Their activation loop contains two residues, a **conserved serine** and either another **serine** or a **threonine**, that are phosphorylated by their upstream MAP3K. A domain of versatile docking (**DVD site**) mediates the interaction with the MAP3K and this interaction is mandatory for the MAP2K activation (Takekawa, Tatebayashi and Saito, 2005). This DVD site is composed of 20 amino acids at the C-terminus of the MAP2K, and binds to the catalytic domain of their upstream MAP3K.

The **N-terminal extension** of MAP2Ks has a role in their specificity, as it contains the **KIM motif**, but also has a regulatory role. A regulatory α-helix, located between the KIM and the kinase domain, has been described for MEK1/2 (Mansour *et al.*, 1996; Fischmann *et al.*, 2009) and for MKK7 (Schröder *et al.*, 2020). In inactive conformations, this negative regulatory domain interferes with the αC-helix and prevents the formation of the conserved canonical salt bridge required for the active conformation.

p38α activation

Canonical p38α activation by MKK6

In order to activate p38α, biphosphorylated MKK6 binds through its KIM domain to the docking site of the MAPK. This interaction induces crucial conformational changes on both kinases. Together with the dual phosphorylation of MKK6's own activation loop and ATP binding to the MKK6 ATP binding site, this interaction with its substrate is believed to trigger the active conformation of MKK6. This docking interaction also induces an allosteric rearrangement of p38α resulting in a "prone to be activated" conformation: the two lobes slightly rotate towards a closed conformation, and the activation loop becomes disordered and accessible to the active site of MKK6 (Tanoue *et al.*, 2001; Tanoue and Nishida, 2002).

MKK6 is consequently able to phosphorylate the Thr180 and Tyr182 residues in the activation loop of p38α. The two phospho-residues then bind at basic sites, stabilising the closed conformation and forming a favourable conformation of the ATP-binding site for catalysis. ATP and substrate binding are probably necessary to achieve a fully active conformation in p38α.

Despite active research and interest in the field, the details of the mechanism of activation of p38α by MKK6, and more generally the activation of a MAPK by its upstream MAP2K, are poorly understood.

Alternative activation mechanisms of p38α

Under certain circumstances p38α can be activated outside of the MAPK pathway. The TGF-β-activated kinase 1 (TAK1)-binding protein (TAB1) is responsible for p38α *cis*-autophosphorylation on both Thr180 and Tyr182 residues (Ge *et al.*, 2002). TAB1 contains a KIM motif (MEF2A class, KTSVTLSLVM) that binds to the same docking site as other MAPK partners (De Nicola *et al.*, 2013). Another region binds to a non-canonical lower pocket in the C-terminal lobe. This mechanism occurs in cardiomyocytes under ischaemia, and have also been linked to T cell senescence, skin inflammation, triiodothyronine-mediated browning of white adipose tissue and endothelial inflammation triggered by G-protein-coupled receptor (GPCR) agonists (Canovas and Nebreda, 2021). However, the signals that activate the TAB1 pathway remain unclear. TAB1 can also trigger p38α activation through the canonical pathway by binding and activating the MAP3K TAK1.

Another alternative route towards p38α activation has been exclusively described in T-cells after T-cell receptor (TCR)-mediated stimulation (Salvador *et al.*, 2005). p38α (as well as p38β) is phosphorylated on Tyr323 by the TCR-proximal tyrosine kinase ZAP70, which leads to the MAPK autophosphorylation of Thr180 in *trans*. The monophosphorylated p38α is active, but demonstrates a different substrate specificity than the canonical biphosphorylated activate p38α (Mittelstadt *et al.*, 2009).

These alternative routes toward p38α activation participate in the fine-tuning of the response to different types of signal. It orientates p38α activity towards different targets by modelling substrate specificity, in a cell dependent and stress dependent manner, and seems to be operating in parallel to the canonical pathways.

Some pathogens have also adapted to create shortcuts and remodel signalling pathways to their advantage, such as *Toxoplasma gondii* triggering p38α autophosphorylation upon binding to its GRA24 effector protein (Pellegrini, Palencia, *et al.*, 2017).

PTMs that influence p38 activation and activity

Different post-translational modifications (PTMs) can affect p38α activation and/or activity. As reported above, the phosphorylation of Tyr323 by ZAP70 of p38α leads to the autophosphorylation and activation of the kinase (Salvador *et al.*, 2005).

Other PTMs enhance signal transduction. For example, arginine methylation on Arg49 and Arg149 of p38α by the protein arginine methyltransferase 1 (PRMT1) augment MKK3-dependent phosphorylation of the MAPK (Liu *et al.*, 2020). The acetylation of Lys53 in p38α increases the affinity for ATP and augments p38α kinase activity (Pillai *et al.*, 2011). Also, the isomerization of Pro224 enhances the accessibility of the Thr180 and Tyr182 for phosphorylation by MAP2Ks (Brichkina *et al.*, 2016).

On the contrary, some PTMs can also diminish the activation of the MAPK. The phosphorylation of p38α Thr123, located at the docking site, by GRK2 impairs the binding of both MKK6 and p38α substrates (Peregrin *et al.*, 2006).

Another example of a critical PTM is the acetylation of both phospho-sites Ser207 and Thr211 of MKK6 by *Yersinia* species, impeding their phosphorylation and therefore the MAP2K activation (Mukherjee *et al.*, 2006).

p38α regulation and signal termination

A number of feedback loops involving different components of the pathway allow a fine tuning of the cellular response to stress (Canovas and Nebreda, 2021). Several phosphatases, Tyr specific, Ser/Thr specific and dual specificity phosphatases, dephosphorylate the activation loop phospho-residues of p38α to deactivate it. In particular, DUSP1 (Dual-specificity phosphatase 1) is activated by p38α, and acts as negative feedback loop. The activation of p38α also represses MKK6 expression and reduces MKK6 mRNA stability, but not MKK3 mRNA (Ambrosino *et al.*, 2003). By phosphorylating TAB1, p38α impairs both the canonical (through the MAP3K TAK1 activation) and the TAB1 non-canonical route. The phosphorylation of ZAP70 reduces its binding to TCR and diminish the ZAP70-dependent p38α activation in T-cells. The activity of p38α could be regulated by controlling its cellular localization. p38α is also regulated by caspase-mediated degradation (Weaver *et al.*, 2020).

Kinase inhibitors and the “specificity” challenge

The MAPKs in general, and p38α specifically, are key molecular targets for the development of new therapeutic strategies. They are involved in such essential cellular processes that high target specificity is an indispensable condition for successful drug development. However, due to the high conservation of the protein kinase fold and catalytic elements, kinase inhibitors often have poor selectivity for their target, leading to off-site effects and toxicity. Most of them target the ATP binding site.

Kinase inhibitors are classified as:

- type I: bind to DFG-in conformation to prevent ATP binding
- type II: targeting the inactive DFG-out conformation to prevent ATP binding
- type III: non-ATP competitive inhibitors

In the past two decades, several p38α inhibitors have been developed to treat inflammation, autoimmunity, cardiovascular diseases, diabetes, neurodegenerative diseases, and cancer, but most of them failed in clinical phases (Canovas and Nebreda, 2021).

A few p38α inhibitors are currently still under clinical or preclinical trials (reviewed in (Canovas and Nebreda, 2021)). For instance, p38α inhibitors are currently under clinical trials against COVID-19 (Grimes and Grimes, 2020). The therapeutic hypothesis is that exaggerated acute inflammation mediated by p38α in response to SARS-CoV-2 infection is associated with increased mortality and severe disease (drug: Losmapimod, phase III, ClinicalTrials.gov Identifier: NCT04511819).

Targeting p38α in specific cell-types, or targeting specific p38α signalling branch instead of generalized inhibition could increase the chances of success of new inhibitors.

The idea of exploring binding sites distant from the ATP binding pocket and active site of MAPKs have been proposed. The docking sites of MAPKs are essential for the specificity and activity of MAPKKs and represents a potential alternative target for the development of a new category of probe molecules that are independent of nucleotide binding. The interaction between MAPKs and their upstream activating MAP2Ks probably extend beyond the KIM-docking site interaction, and some new druggable sites might be revealed by better structural understanding of the interaction.

Compounds that target p38α for degradation could be of clinical value, and might arise from the development of PROTACs inhibitors (Donoghue *et al.*, 2020). Structural studies are still required to understand the interactions and mechanisms leading to p38α activation, and of MAPKs in general. To this date, no complete structures or molecular models of the interactions between a MAPK and its activating MAP2K are available. New insights could lead to novel therapeutic approaches.

Host-pathogen interactions

SLiMs mimicry and interactions with host signalling pathways

Small linear motifs (SLiMs) are protein interaction sites composed of short amino acid sequences. They are present in many proteins, especially in intrinsically disordered regions (IDRs). These motifs are essentials in a number of mechanisms involving protein-protein interactions, which is particularly the case in signalling mechanisms (Van Roey *et al.*, 2014; Kumar *et al.*, 2020). An example of a SLiM is the kinase interacting motif (KIM), described above (Table 1, p33). KIMs have a crucial role in the interaction between MAPKs and their interacting partners, both in terms of specificity and activity.

A variety of pathogens, from viruses to parasites, comprising *Toxoplasma gondi*, use SLiMs to manipulate the host cell machinery (Davey, Travé and Gibson, 2011). They confer an evolutionary advantage as they are usually located in IDRs that tend to mutate more rapidly than structured regions. They allow the pathogens to take control of their host's cellular pathways and to remodel them to their advantage.

The mimicry of host SLiMs by pathogens represents an opportunity for drug development (Sámano-Sánchez and Gibson, 2020). Therapies can be designed to directly target the pathogen. For example, a peptidomimetic have been developed to target the *Plasmodium falciparum* PEXEL export motif as a potential antimalarial agent (Nguyen *et al.*, 2018). Another possibility is to explore the concept of host-directed therapy in which the host functions that are remodelled by the pathogens are targeted to inhibit the interaction between the pathogen SLiM and the host protein. For instance, the SARS-CoV-2 Spike protein contains a RGD motif that mediates the interaction with host integrins for cell attachment, opening the possibility to assess the effect of integrin inhibitors to modulate viral infection (Mészáros *et al.*, 2021).

Toxoplasma gondii

Toxoplasmosis

Toxoplasma gondii is an obligate intracellular parasite of the Apicomplexan family. *T. gondii* infection is the most predominant parasite infection in western countries. For instance, in France, it is estimated around 50% of the population is infected (Robinson *et al.*, 2021). Toxoplasmosis is the chronic zoonosis caused by *T. gondii* and is benign in the majority of cases, with sometimes the occurrence of mild flu-like symptoms. However, *T. gondii* reactivation can lead to complications for immunocompromised patients, such as people suffering from AIDS or undergoing cancer therapy. It can also cause severe developmental and neuronal damages to the foetus if a woman is primarily infected during pregnancy.

The *T. gondii* life cycle is summarized in Figure 9.

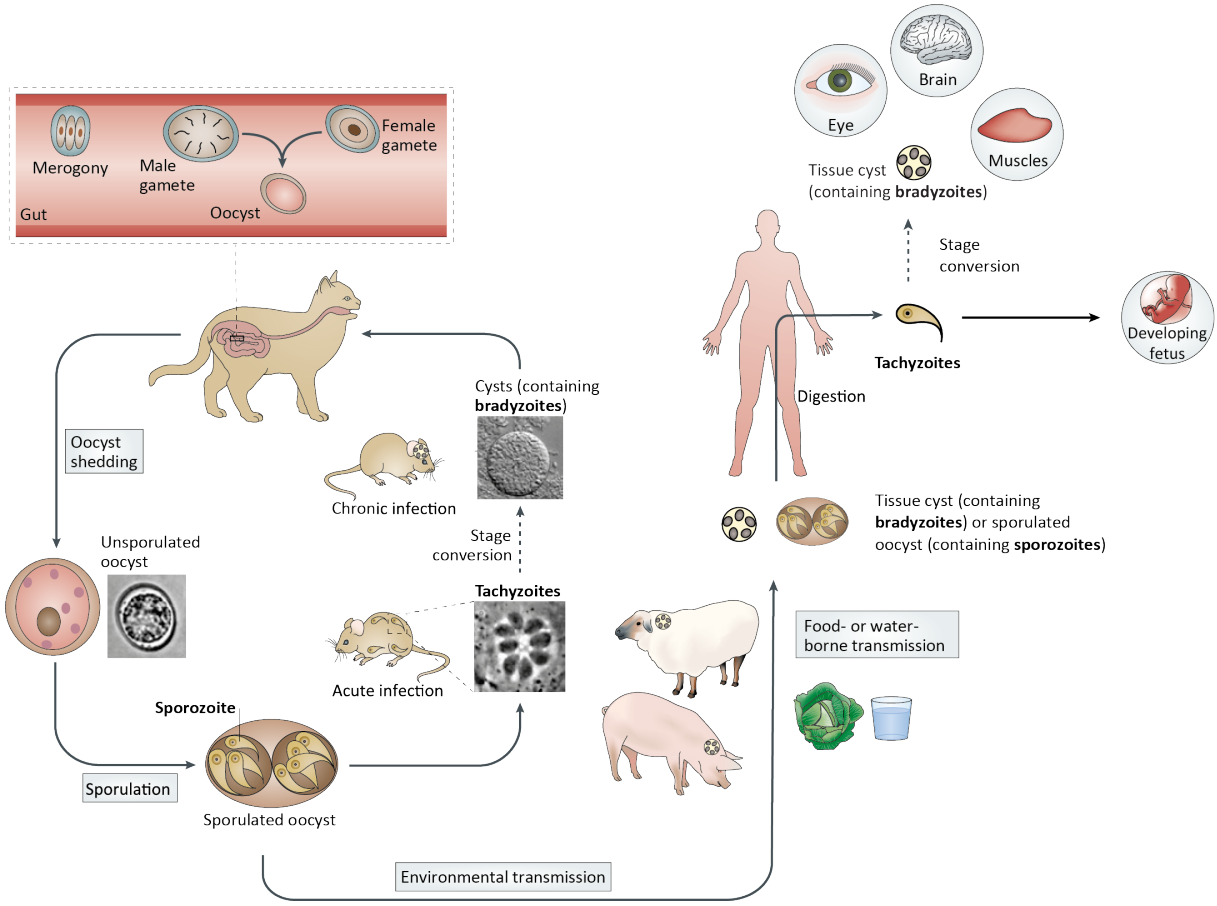


Figure 9: Life cycle of T. gondii adapted from (Hunter and Sibley, 2012). T. gondii presents different developmental stages along its lifecycle: non-sexual forms that can infect cells (sporozoite, tachyzoite, and bradyzoite), and sexual forms (gametes and oocyst). T. gondii definitive hosts are from the Felidae family (which includes the domestic cat): they are the only hosts in which the sexual stages of the parasite can occur. Most warm-blood animals can be infected though their environment and are considered as secondary or accidental hosts. Human infection usually occurs through the consumption of food or water contaminated by oocysts, or tissue cysts in undercooked meat of infected animals. The foetus can be contaminated by tachyzoites if the parent gets primarily infected during pregnancy.

Cellular invasion mechanism

This study concentrates on the primary infection of human cells by tachyzoites, as it is the most relevant stage in the study of toxoplasmosis related pathology. The internalization of the parasite can occur in most cells through the combined action of proteins secreted by different organelles (Figure 10) (Rastogi, Cygan and Boothroyd, 2019). Tachyzoites identify and dock to host cell receptors through an assembly of secreted microneme proteins (MIC). Together with rhoptry neck proteins (RON), they form the moving junction to mediate the parasite entry. The host cell membrane will surround the tachyzoite throughout the internalization process, forming the parasitophorous vacuole (PV) in which the tachyzoites will multiply and partly be protected from the host's defences. Then, a first wave of effectors, the rhoptry bulb proteins (ROP), are secreted by the rhoptry organelles. In a later phase, dense granule proteins (GRA)

are secreted, and unlike ROP proteins, they are secreted throughout the intracellular stage of the parasite.

After several division cycles inside the PV, the tachyzoites are released in the extracellular space, breaking the host cell membrane, and ready to infect new cells. After a few days of infection, the stage conversion to bradyzoites slowly occurs in muscle cells and neurons, where they remain semi-dormant within cysts.

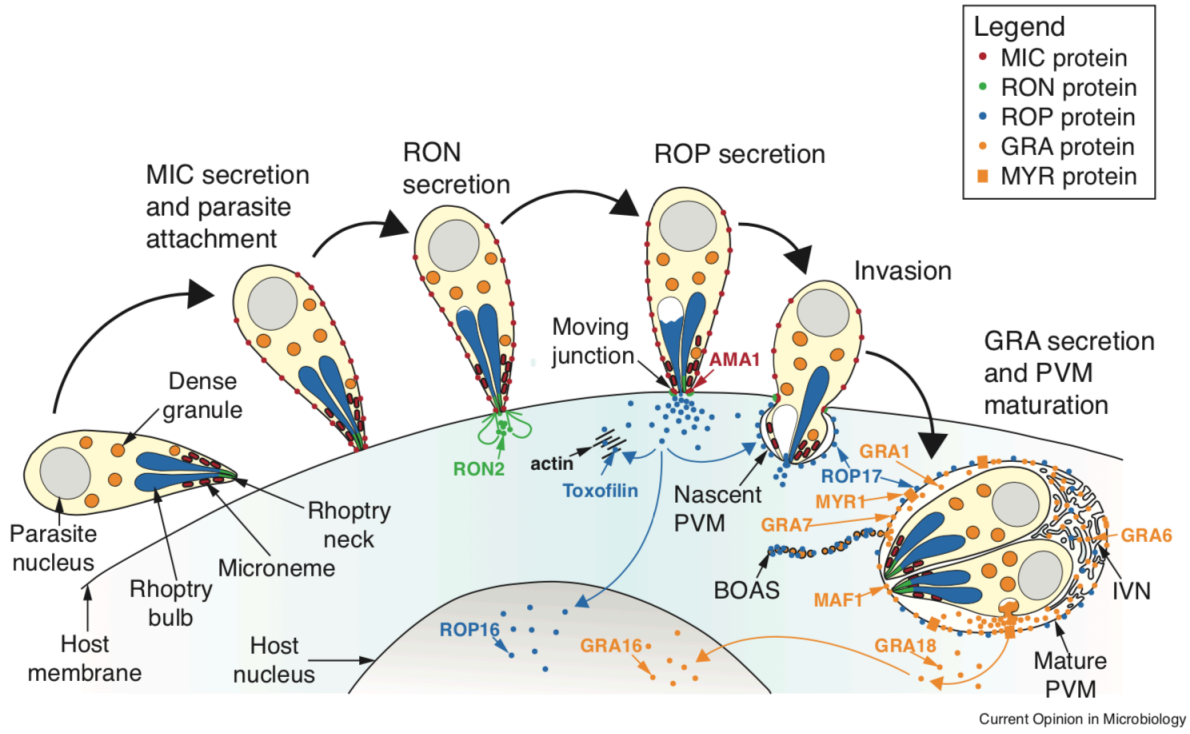


Figure 10: Model of T. gondii tachyzoites infecting a cell from (Rastogi, Cygan and Boothroyd, 2019). T. gondii tachyzoites contain three types of secretory organelles: micronemes, rhoptries and dense granules that secrete the effector proteins MIC, RONs, ROP and GRA to invade the host cell cytosol within a parasitophorous vacuole and remodel the host cell transcriptional machinery.

Host immune responses

T. gondii infection triggers host immune pathways. The immune response is mainly mediated by γ-interferons (IFN-γ), and pro-inflammatory cytokines. A strong and persistent Th1 proinflammatory response is necessary to control acute parasite multiplication and dissemination upon primary infection (Jensen *et al.*, 2011). The proinflammatory cytokine interleukin-12 (IL-12) is produced by immune cells and triggers the production of IFN-γ by activated Natural killers and T-cells. In turn, IFN-γ is responsible for a number of protective mechanisms. IFN-γ induces tryptophan degradation and therefore inhibits parasite growth. The nitrogen oxide pathway is activated to degrade L-arginine, a necessary amino acid for *T. gondii* survival. IL-12 also enhances the macrophage phagocytic activity. IFN-γ induces the activation of IFN-regulated GTPases (IRG) and guanylate-binding proteins (GBP) that accumulate on the PV, destroy the vacuole, allowing the elimination of the parasites by lysosomal degradation (Zhao *et al.*, 2009). In mice lacking IL-12 or IFN-γ, the *T. gondii* infection is not controlled and leads to the death of the host. Tachyzoites infection also has an effect

on the host cell cycle: it advances the cell into S phase and arrests it in the G2 phase, prior to mitosis (Brunet *et al.*, 2008).

T. gondii effectors

During host cell invasion, *T. gondii* tachyzoites secrete different effector proteins sequentially (Hakimi, Olias and Sibley, 2017): first the MIC and RON proteins for parasite internalization, then ROP proteins and finally GRA proteins. Together, *T. gondii* ROP and GRA effectors collaborate to remodel the host cell transcription machinery and achieve an equilibrium between:

- anti-parasitic immune response to control the parasite proliferation and assure host survival
- immune evasion mechanisms to allow the parasite survival and dissemination, to achieve the transition to the long-lasting chronic infection in skeletal muscles and brain tissues.

GRA proteins share common features: they usually have a signal peptide at their N-termini and they are mainly intrinsically disordered. Some of them are exported to the host cell cytoplasm through transporters such as Myr1 (Marino *et al.*, 2018). More than 2000 host genes are impacted by GRA proteins (Naor *et al.*, 2018).

A selection of *T. gondii* effectors, their host targets, and effects are presented in Table 2. They are reviewed in more detail in (Hakimi, Olias and Sibley, 2017; Panas and Boothroyd, 2021). This thesis focuses on two GRA proteins: GRA24 and GRA16.

GRA24 is an intrinsically disordered protein that contains two KIM motifs (Braun *et al.*, 2013; Pellegrini, Palencia, *et al.*, 2017). It is exported from dense granules to the host cytoplasm and nucleus where it binds to the p38α docking site through the interaction with its KIM motif. Through this binding interaction, it triggers p38α auto-phosphorylation in *trans*. The resulting activated p38α induces the production of pro-inflammatory IL-12 cytokine.

GRA16 is also exported to the host cell nucleus. It is found in a complex together with the host USP7 (Ubiquitin-specific protease 7, also known as the herpesvirus associated ubiquitin-specific protease HAUSP) and the PP2A-B55 phosphatase (Bougdour *et al.*, 2013). The exact mechanism of action of GRA16 remains unclear, but it is associated with c-Myc upregulation and perturbations of p53 levels (Kim *et al.*, 2019; Panas and Boothroyd, 2020).

Table 2: T. gondii effectors, their targets and effects in host cells

	T. gondii effector	Host targets	Effects	References
Activation of the immune response	**GRA24**	p38α	Promotion of p38α auto-phosphorylation and activation Induction of IL-12 production	(Braun *et al.*, 2013; Pellegrini, Palencia, *et al.*, 2017)
	GRA16	USP7 PP2A-B55	Upregulation of c-Myc Perturbations in p53 levels	(Bougdour *et al.*, 2013; Kim *et al.*, 2019; Panas and Boothroyd, 2020)
	GRA15		Induction of IL-12 production mediated by NF-κB Activation of caspase 1 and IL-1β production	(Rosowski *et al.*, 2011)
Escape of the immune response	**ROP16 kinase**	STAT3/STAT6	Activation of STAT3/STAT6 Transcription of anti-inflammatory factors Down-regulation of IL-12	
	TEEGR	E2F3/4 transcription factors	Upregulation EZH2 (major component of the Polycomb repressive complex 2) Induction of the epigenetic silencing of NF-κB-regulated cytokines	(Braun *et al.*, 2019)
	TgIST	STAT1 NuRD complex	Blocking of the IFN-γ pathway Silencing of IGNs responses	
	ROP17/18 **ROP5**	IRGs	Inhibition of IRG oligomerization and accumulation at the PV	
	GRA7	IRGs	Enhancement of ROP18 activity Promotes rapid disassembly and turnover of IRGs	
	TgNSM		Prevention of necrotic death.	
	GRA18	β-catenin degradation complex (including GSK3 and PP2A-B56)	Promotion of β-catenin accumulation by preventing its degradation Activation of the WNT pathway Upregulation of the anti-inflammatory CCL17 and CCL12 cytokines	(He *et al.*, 2018)
	GRA28		Upregulation of CCL22 chemokines	
	GRA6		Activation of NFAT4 transcription factor Production of Cxcl2 and Ccl2 chemokines	(Ma *et al.*, 2014)
	TgWIP	WAVE regulatory complex SHP2 phosphatase	Promotion of the motility in dendritic cells	

Therapeutic outcomes

The study of the interactions between *T. gondii* effectors and host proteins reveals multiple layers of action exploited by the parasite to promote long-term parasitism with the right balance of parasite clearance and immune escape. The effectors are released sequentially, with a first burst of ROP proteins followed by a prolonged secretion of GRA proteins. These effectors affect antagonistically different signalling pathways, and the success of long-lived establishment of the parasite resides in the tight equilibrium of these opposite effects.

The characterization of the interaction between *T. gondii* effectors and host cell proteins is essential for the understanding of parasite infection. It may allow the identification of therapeutic targets to treat toxoplasmosis in patients at risk. Available treatments are associated with adverse side effects and toxicity, including teratogenic effect in foetuses, which limit their use (Smith *et al.*, 2021). They also fail to eliminate tissue cysts.

To protect against primary infection, the development of a safe and effective vaccine is still ongoing. A GRA24 DNA vaccine gave promising results in mice (Zheng *et al.*, 2019). Novel therapeutic strategies could also aim to block the transition from tachyzoites to bradyzoites-containing cysts to prevent a long-term infection that could reactivate later in the patient life. Treatments to counter the reactivation of toxoplasmosis in immunosuppressed patients would be of great value.

T. gondii is a member of the apicomplexan family, that also include *Plasmodium falciparum,* the causative agent of malaria. Acquired knowledge on the first micro-organism may open leads in the crucial research field of malaria.

Finally, the subversion of host molecular pathways by pathogenic effectors brings new light on the human signalling network, and provide us with important tools for further studies.

Aims of this thesis

The work of this thesis has been structured around four projects, presented in the four following chapters. The objectives fall into two main areas: to define the molecular basis of activation of a MAPK by its MAPKK (Chapters I and II) and to define the binding modes and mechanism of actions of *T. gondii* GRA proteins with human proteins involved in signalling networks (Chapters III and IV).

The interaction leading to the activation of a MAPK by its upstream MAPKK is difficult to study due to its transient nature. The exact mechanism of interaction is not completely defined and, so far, there are no data on the molecular details of this event.

- The first chapter describes the architecture of MAP2K phosphorylating its target MAPK, focusing on the MKK6/p38α interaction, two important members of the MAPK family involved in the inflammatory response. Molecular details of the activation should provide essential details on how the proteins interact and catalytic details of phosphorylation of the p38α activation loop.
- The second chapter explores a novel non-canonical mechanism of activation of p38α by MKK6 using ADP in place of ATP.

The second half of this thesis focuses on the study of two *T. gondii* effectors, GRA24 and GRA16, and how they interact with host proteins:

- The third chapter presents the characterization of the interaction between GRA24 and the MAPK ERK1/2.
- Finally, the fourth chapter exposes the characterization of the interaction between GRA16 and human USP7.

Chapter I: Characterization of the activation of p38α by its upstream MAP2K MKK6

Introduction

MKK6 activates p38α by dual-phosphorylating the TxY motif of its activation loop. Many X-ray structures of the individual kinases and of complexes with isolated KIMs exist; however, these structures do not provide molecular details about selectivity and activation of a MAPK by MAPKK.

In order to activate p38, MKK6 needs to itself adopt an active conformation. In order to do so, MKK6 is phosphorylated on two sites of its activation loop, Ser207 and Thr211, by its upstream MAP3Ks. Only three crystal structures of MKK6 are available and none of them display a catalytically active conformation (Figure 11), even when phospho-mimetic mutations are inserted in the activation loop (Figure 11B and C).

We can speculate that to achieve a fully active conformation, MKK6 requires:

- the dual phosphorylation of its activation loop residues Ser207 and Thr211
- the binding of ATP to its active site
- the interaction with its substrate p38α, through its N-terminal extension and potentially other regions.

There are many structures of p38α (and a few of the other isoforms) that offer an overview of the different possible conformations of the MAPK (Figure 12). The apo, inactive and unphosphorylated p38α (PDB 5etc) adopts a conformation close to the canonical kinase fold (Figure 12A). The mouse biphosphorylated p38α structure PDB 3py3 is the only structure of biphosphorylated p38α available (Figure 12E). The phosphates of Thr180 and Tyr182 bind to the N-lobe and the C-lobe of the kinase, respectively, bringing the two lobes together towards an active conformation. However, the transition to the canonical active kinase fold seems to be achieved only in presence of nucleotide binding to the ATP site, as it can be observed in the structure of biphosphorylated p38γ in PDB 1cm8 (Figure 12F).

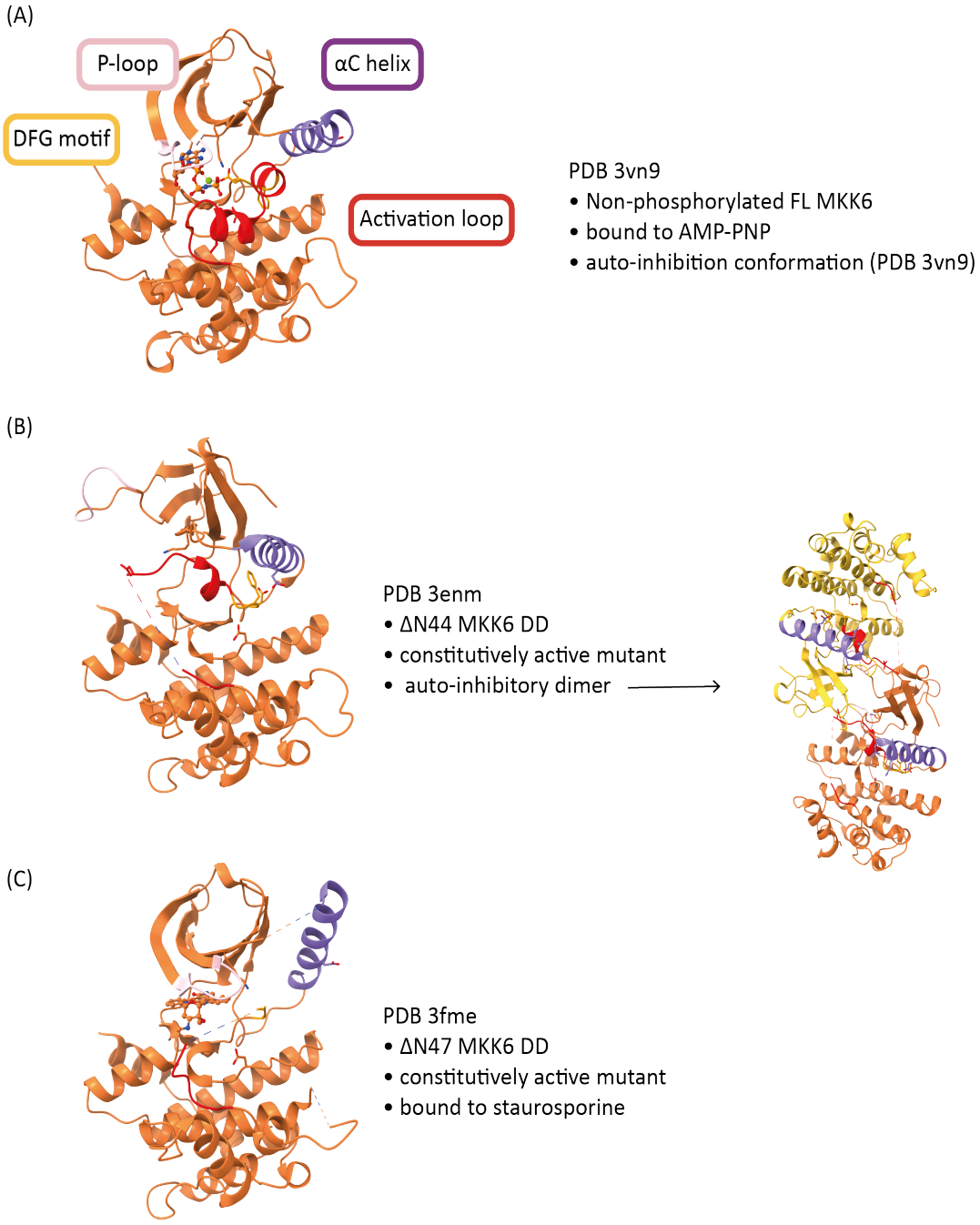


Figure 11: MKK6 crystal structures. (A) The crystal structure of non-phosphorylated FL MKK6 (PDB 3vn9) reveals of a potential auto-inhibition conformation (Matsumoto et al., 2012). The activation loop is structured with three short α-helices referred to as activation helices 1-3 (AH1-3) and contribute to lock the γ-phosphate of the ATP analogue inside its active site. AH1 displace the αC-helix and position it away from the active site, preventing the formation of the salt bridge with the β3-strand that is essential for the active conformation of the kinase. (B) The crystal structure of ΔN44 MKK6DD (PDB 3enm), a constitutively active mutant (Ser2017 and Thr211 of the activation loop are mutated to aspartate to mimic phosphorylation) proposes another auto-inhibition mechanism (Min et al., 2009). Two MKK6 molecules form an elongated dimer bringing the two N-termini together. At the interface, the Glycine rich loop forms a distorted Ω loop that interacts with the other MKK6 monomer in its active site. (C) In the crystal structure of ΔN47 MKK6DD with the ATP-competitive inhibitor staurosporine (PDB 3fme) MKK6DD does not form a dimer. It is not in an active conformation either. The DFG motif seems to be in DFG-in conformation (the Gly is not visible in the model and the side-chains of the Asp and Phe are not modelled, but they adopt a similar conformation to the DFG-in in PDB 3enm).

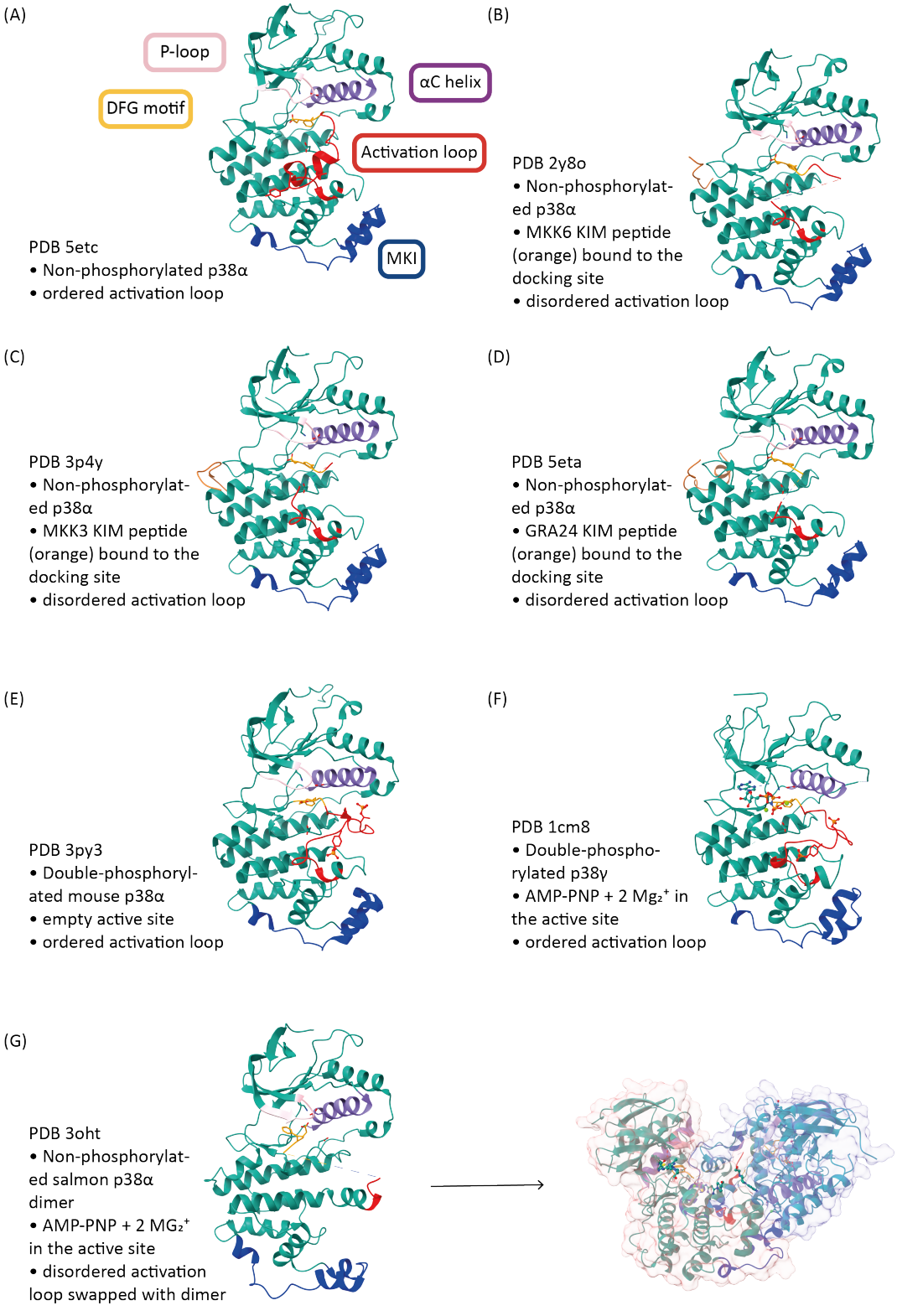


Figure 12: Selection of p38 crystal structures (A) Apo, inactive and unphosphorylated p38α (PDB 5etc). (B) Unphosphorylated p38α bound to MKK6 KIM peptide (PDB 2y8o). (C) Unphosphorylated p38α bound to MKK3 KIM peptide (PDB 3p4y). (D) Unphosphorylated p38α bound to T. gondii GRA24 KIM peptide (PDB 5eta). (E) Biphosphorylated mouse p38α (PDB 3py3). (F) Biphosphorylated p38γ bound to AMP-PNP (PDB 1cm8). (G) Unphosphorylated salmon p38α dimer (PDB 3oht).

The main structural information available on the interaction between MAPKs and MAP2Ks comes from crystal structures of MAP2K KIM peptides bound to MAPK docking site (Figure 12C and D). In these structures, the binding event induces allosteric conformational changes and the activation loop is disordered (Chang *et al.*, 2002; Zhou *et al.*, 2006; Akella *et al.*, 2010).The docking interaction between the MKK6 KIM motif and p38α is also required for MKK6 activity (Wang *et al.*, 2019). This interaction probably induces a conformational rearrangement in MKK6 upon binding its substrate, which promotes the MKK6 active conformation. Besides, the differences of affinity of KIM motifs for MAPKs are not sufficient to explain the specificity of the interaction. Other interaction interfaces between MKK6 and p38α likely exist.

In some specific circumstances, p38α can also auto-phosphorylate. The salmon p38α dimer structure PDB 3oht offers an interesting insight into this mechanism and the associated possible conformations. (Figure 12G). The activation loops are swapped (Rothweiler *et al.*, 2011). This structure is an example of the high adjustability and flexibility of the p38α activation loop.

However structural and molecular information of the p38α conformation leading to its phosphorylation by MAP2Ks is still lacking.

The aim of this project is to gain structural and functional information on the recognition and phosphorylation of p38α by MKK6.

The strategy to achieve this goal was to engineer an MKK6 construct with high affinity for p38α in order to stabilize the transient interaction between the two kinases. We substituted the MKK6 KIM with the one of the *Toxoplasma gondii* GRA24 protein. This protein contains two KIM motifs that mimic the interaction of MAP2Ks to hijack the signalling pathway (Figure 12E). The GRA24 KIM has similar allosteric effects as the MKK6 KIM but has an affinity 100 fold higher (Pellegrini, Palencia, *et al.*, 2017). We used an integrative structural approach using a combination of MS, SAXS, X-ray crystallography, negative stain and cryoEM techniques to optimize the preparation of the sample of the kinase heterodimer trapped in transition state. To trap the transition state, we explored the use of metallofluorides (Jin, Molt and Blackburn, 2017). We successfully obtained the first model of the interaction between a MAPK and its activating MAP2K.

Materials and methods

Protein expression and purification

Recombinant protein constructs

p38α refers to the human wild-type Mitogen-activated protein kinase 14 (MAPK14) canonical isoform CSBP2 (Uniprot MK14_HUMAN, Q16539-1). p38α K53R, T180V, and Y182F refers to the p38α mutants. In the first part of the chapter p38α WT and mutants correspond to an N-terminal truncated form (10-360). This construct is supposed to facilitate crystallization. For cryoEM applications, we decided to work with the full length (FL) version (1-360).

MKK6 refers to the human wild-type Dual specificity mitogen-activated protein kinase kinase 6 (MAP2K6) isoform 1 (Uniprot MP2K6_HUMAN, P52564-1). MKK6DDGRA refers to the activated mutant S207D T211D of MKK6 containing the GRA24 KIM1 motif GRA KIM1 (LLERRGVSELPPLYI) in place of the natural sequence. We initially worked with the mouse gene of MKK6 for a few experiments, therefore when specified, mMKK6DDGRA refers to the activated mutant S207D T211D of mouse MKK6 (Uniprot gene MP2K6_MOUSE, P70236-1) containing the GRA24 KIM1 motif.

Plasmids encoding p38α and MKK6DDGRA constructs were ordered from Genscript (gene synthesis, cloning in expression vector and mutagenesis).

Lambda Phosphatase plasmid was ordered from Addgene (gift from John Chodera & Nicholas Levinson & Markus Seeliger ; Addgene plasmid # 79748 ; http://n2t.net/addgene:79748 ; RRID:Addgene_79748) (Albanese *et al.*, 2018).

Recombinant protein constructs used in this chapter are presented in Table 3.

Table 3: Recombinant protein constructs

Protein	**Construct**	**Protein sequence** **(tag and protease recognition cleavage site in bold and cleavage site indicated by ↕)**	**MW (cleaved protein MW)**
p38α (10-360)	N-ter His6 tag + TEV cleavage site pETM-11 (Kan)	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMARQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKKL SRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQILR GLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCIM AELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLDS DKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,482 kDa (40,484)
p38α (1-360)	N-ter His6 tag + 3C cleavage site pET-28b+ (Kan)	MGSS**HHHHHH**SSG**LEVLFQ↕GP**MSQERPTFYRQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKK LSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQIL RGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCI MAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLD SDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,594 kDa (41,448)
p38α T180V (10-360)	N-ter His6 tag + TEV cleavage site pETM-11 (Kan)	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMARQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKKL SRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQILR GLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMVGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCIM AELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLDS DKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,480 kDa (40,480)
p38α T180V (1-360)	N-ter His6 tag + 3C cleavage site pET-28b+ (Kan)	MGSS**HHHHHH**SSG**LEVLFQ↕GP**MSQERPTFYRQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKK LSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQIL RGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMVGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCI MAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLD SDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,592 kDa (41,446)
p38α Y182F (10-360)	N-ter His6 tag + TEV cleavage site	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMARQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKKL SRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQILR GLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGFVATRWYRAPEIMLNWMHYNQTVDIWSVGCIM	43,466 kDa (40,468)

	pETM-11 (Kan)	AELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLDS DKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	
p38α Y182F (1-360)	N-ter His6 tag + 3C cleavage site pET-28b+ (Kan)	MGSS**HHHHHH**SSG**LEVLFQ↕GP**MSQERPTFYRQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKK LSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQIL RGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGFVATRWYRAPEIMLNWMHYNQTVDIWSVGCI MAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLD SDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,578 kDa (41,432)
mouse MKK6 S207D T211D GRA24KIM1 **(referred to as mMKK6DDGRA)**	N-ter Twin strep tag + 3C cleavage site pFastBac1	MGSQLLERRGVSELPPLYIPKEAFEQPQTSSTPPRDLDSKACISIGNQNFEVKADDLEPIVELGRGAYGVVEKMRHVPS GQIMAVKRIRATVNSQEQKRLLMDLDVSMRTVDCPFTVTFYGALFREGDVWICMELMDTSLDKFYKQVIDKGQTIPE DILGKIAVSIVKALEHLHSKLSVIHRDVKPSNVLINTLGQVKMCDFGISGYLVDDVAKDIDAGCKPYMAPERINPELNQK GYSVKSDIWSLGITMIELAILRFPYDSWGTPFQQLKQVVEEPSPQLPADKFSADFVDFTSQCLKKNSKERPTYPELMQH PFFTVHESKAADVASFVKLILGD**LEVLFQ↕GPWSHPQFEKGGGSGGGSGGSAWSHPQFEK***	41,714 kDa (38,691)
human MKK6 S207D T211D GRA24KIM1 **(referred to as MKK6DDGRA)**	N-ter Twin strep tag + 3C cleavage site pFastBac1	MGSQLLERRGVSELPPLYIPKEAFEQPQTSSTPPRDLDSKACISIGNQNFEVKADDLEPIMELGRGAYGVVEKMRHVPS GQIMAVKRIRATVNSQEQKRLLMDLDISMRTVDCPFTVTFYGALFREGDVWICMELMDTSLDKFYKQVIDKGQTIPE DILGKIAVSIVKALEHLHSKLSVIHRDVKPSNVLINALGQVKMCDFGISGYLVDDVAKDIDAGCKPYMAPERINPELNQK GYSVKSDIWSLGITMIELAILRFPYDSWGTPFQQLKQVVEEPSPQLPADKFSAEFVDFTSQCLKKNSKERPTYPELMQH PFFTLHESKGTDVASFVKLILGD**LEVLFQ↕GPWSHPQFEKGGGSGGGSGGSAWSHPQFEK***	41,774 kDa (38,751)
Lambda phosphatase	none pET13S-A (Spec)	MRYYEKIDGSKYRNIWVVGDLHGCYTNLMNKLDTIGFDNKKDLLISVGDLVDRGAENVECLELITFPWFRAVRGNHE QMMIDGLSERGNVNHWLLNGGGWFFNLDYDKEILAKALAHKADELPLIIELVSKDKKYVICHADYPFDEYEFGKPVDH QQVIWNRERISNSQNGIVKEIKGADTFIFGHTPAVKPLKFANQMYIDTGAVFCGNLTLIQVQGEGA	25,219 kDa

Plasmid amplification

The plasmids were transformed into ultra-competent XL10 cells (Agilent Technologies) following the supplier's protocol. The cells were plated on LB agar plates with appropriate antibiotic and grown overnight at 37°C. Individual colonies were amplified in 5ml LB + antibiotic overnight. Plasmids were purified using a MiniPrep kit (Macherey-Nagel). The purified plasmids were sent to Genewiz for control sequencing.

Protein expression in *E. coli*

Each pET plasmid was co-transformed with lambda phosphatase plasmid into 50 µl Rosetta™(DE3)pLysS competent cells (Novagen). Cells were incubated on ice for 30 minutes before being heat shocked at 42°C for 30 seconds. Cells were then placed on ice for 2 minutes and 450 µl of LB was added before incubating at 37 °C for one hour. The cells were then plated on LB agar plates + kanamycin + chloramphenicol + spectinomycin and incubated at 37 °C overnight.

Glycerol stocks of these cultures were prepared as follows. LB (5ml) + kanamycin + chloramphenicol + spectinomycin was inoculated with several colonies and grown overnight at 37°C. Cells were pelleted by centrifugation at 3200 g for 10 minutes, resuspended into 500 µl of LB and transferred to a 2 ml cryotube. Cold glycerol solution (500 µl at 50% v/v) was added while homogenizing smoothly and the tubes were immediately frozen into liquid N_2 and stored at -80°C.

A preculture of 50 to 100 ml LB + kanamycin + chloramphenicol + spectinomycin was prepared inoculating with the glycerol stock and incubating at 37 °C overnight. 8 x 500 ml LB + kanamycin + chloramphenicol were inoculated with 1:20 to 1:50 of preculture and incubated at 37°C until the OD_{600} reached 0,6 to 0,8. Protein expression was induced with 0,5mM IPTG and cultures were incubated at 16°C overnight.

Cells were collected by centrifugation at 4000g for 15 minutes and stored at -80°C until further purification.

Protein expression in insect cells

MKK6DDGRA constructs were expressed in *Spodoptera frugiperda* Sf21 cells using the baculovirus expression system (Nie *et al.*, 2014). Bacmid transfection, virus production, virus amplification and protein expression have been performed at the EMBL Eukaryotic Expression Facility, following guidelines provided by the facility.

Cells were harvested 48h after proliferation arrest by centrifugation at 1000 g for 15 minutes. Cells pellets were flash-frozen in liquid nitrogen and stored at -80°C until further purification.

Protein purification

Cells pellets were thawed and resuspended into 100 ml lysis buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP, with a Pierce protease inhibitor EDTA-free tablet (Thermo Scientific) and a trace of DNaseI (Sigma)) for every litre of culture. The cells were lysed by sonication on ice. The lysate was centrifuged at 40 000 g for one hour.

The p38α supernatant was loaded onto a pre-packed 5 ml HisTrap column (GE Healthcare), equilibrated according to the supplier's protocols with wash buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP) with 1% of elution buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP, 500mM Imidazole, at a flow rate of 1 ml/min and washed until the UV signal returned to zero. Tagged-protein was eluted with a gradient of elution buffer over 60ml from 1% to 100% of elution buffer. Fractions containing protein were identified using SDS PAGE gels, these were pooled and quantified using a NanoDrop ND1000 to measure absorption at 280 nm. His-tagged 3C protease was added to the sample (1 mg for every 100 mg of protein) in the pooled fractions. The sample was dialysed overnight at 4°C against wash buffer. After dialysis, the sample was run through the HisTrap column again to eliminate uncut protein and protease, and flow-through fractions were collected. SDS PAGE was run, and protein containing fractions pooled and quantified. Samples were stored at -80°C in 1 mg aliquots until further purification.

The MKK6DDGRA supernatant was loaded onto a pre-packed 5 ml StrepTactin XT column (IBA), equilibrated according to the supplier's protocols with wash buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP), at a flow rate of 1 ml/min and washed until UV signal returned to zero. Tagged-protein was eluted with 50 ml elution buffer (wash buffer with 50 mM Biotin). SDS PAGE gel was run to identify fractions containing protein, which were pooled and quantified using a NanoDrop ND1000 to measure absorption at 280 nm. His-tagged 3C protease was added to the sample (1 mg for every 100 mg of protein) in the pooled fractions. The sample was dialysed overnight at 4 °C in wash buffer. After dialysis, the sample was run through the StrepTactin XT column again and flow-through fractions were collected. SDS PAGE was run, and protein containing fractions pooled and quantified. Samples were stored at -80°C in 1 mg aliquots until further purification.

In order to prepare the kinase complex, MKK6DDGRA and p38α aliquots were thawed, mixed in 1:1 mass ratio and incubated on ice for 1h. The sample was then concentrated to 10 mg/ml with a 10 kDa Amicon membrane filters concentrators (Merck) and loaded onto a size-exclusion Superdex 75 10/300 GL column (GE Healthcare) equilibrated in wash buffer beforehand (with a reduced glycerol concentration of 2% when for cryoEM studies) and running at 0.3 ml/min. SDS PAGE was used to assess purity and identify protein containing fractions, which were pooled and quantified as before.

Pure protein samples were either used fresh or stored at -80°C.

To prepare the MKK6DDGRA/p38α samples with the AlF_4^- transition state analogue, the purified complexes were complemented with 10 mM ADP, 10 mM NH_4F and 1 mM $AlCl_3$ and incubated on ice for 30 minutes before proceeding.

SDS-PAGE gels

Pre-cast 4-20% gradient Tris-Glycine gels (ThermoFisher Scientific) were used. Protein samples were mixed in a 1:4 ratio with SDS sample buffer (0.4% bromophenol blue, 0.4 M DTT, 0.2 M Tris pH 6.8, 8% SDS, 40% glycerol) and heated at 95°C for 5 minutes. 20 µl of each sample was loaded and the gel run in Tris-Glycine running buffer (2.5 mM Tris Base, 19.2mM glycine pH 8.3, 1% SDS) at 220 V for 40 minutes. Precision Plus Protein Dual Xtra standards were used as molecular weight markers (BioRad). Gels were stained with InstantBlue (Expedeon) and

imaged with a BioRad GelDoc system. Images were analysed using Image Lab (Bio-Rad) software.

Buffer optimization (Thermofluor assay)

To improve the thermal stability of the sample, buffer conditions were screened with a thermal shift assay using (Thermofluor assay) (Boivin, Kozak and Meijers, 2013). Two 96-conditions screens were used to explore different parameters: buffer type, pH, buffer concentration and salt concentration (RUBIC screen), and 76 additives under one constant buffer condition.

Mass spectrometry

Protocols kindly provided by the PEP-core facility of EMBL Heidelberg.

Intact mass by Q-TOF MS (Quadrupole time-of-flight Mass Spectrometry)

Protein samples were acidified using 1% TFA prior to injection onto the Acquity UPLC System (Waters Corporation). Approximately 5 µg of each sample were loaded onto a protein separation column (Acquity UPLC Protein BEH C4 column, 2.1 mm x 150 mm, 1.7 um). The outlet of the analytical column was coupled directly to a quadrupole time of flight (Q-TOF) Premier mass spectrometer (Waters/Micromass) using the standard ESI source in positive ion mode. Solvent A was water, 0.1% formic acid, and solvent B was acetonitrile, 0.1% formic acid. The samples were loaded in 96% A, 4% B at a constant flow of 0.2 mL min-1. The column was held at 4% B for 5 min before ramping to 25% B by 6 min. A linear gradient to 80% B was then applied until 17 min. For the Q-TOF, a spray voltage of 3.5 kV was applied with a cone voltage of 35 V and extraction cone at 5 V. The desolvation temperature was set at 350 °C, with source temperature 120 °C. Desolvation gas was nitrogen at a flow rate of 500 L min-1. Collision energy was set at 5 eV with argon in the collision cell at a pressure of 4.5·10-5 mbar. Data was acquired in continuum mode over the mass range 500–3500 m/z with a scan time of 0.5 s and an interscan delay of 0.1 s. Data were externally calibrated against a reference standard of intact myoglobin, acquired immediately prior to sample data acquisition. Spectra from the chromatogram protein peak were then summed and intact mass was calculated using the MaxEnt1 maximum entropy algorithm (Waters/Micromass) to give the zero charge deconvoluted molecular weight.

Digestion and PTMs analysis by LC-MS/MS (Liquid chromatography coupled to tandem Mass Spectrometry)

Sample preparation

SP3 protocol (PMID: 25358341) was used for sample preparation. In brief, cysteines were reduced using dithiothreitol (56 °C, 30 minutes, 10 mM) and alkylated with 2-chloroacetamide (room temperature, in the dark, 30 minutes, 10 mM). All reagents were prepared in 50 mM HEPES (pH 8.5). Subsequently, protein digestion was carried out with trypsin (enzyme to

protein ratio, 1:50) at 37 °C overnight. OASIS® HLB µElution Plate (Waters) was used for peptide clean up prior to LC-MS/MS. Samples were resuspended after clean-up in 20 ul of 1% formic acid, 4% acetonitrile and 1.5ul were injected.

LC-MS/MS

An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters). Trapping was carried out with a constant flow of solvent A (3% DMSO, 0.1% formic acid in water) at 30 µL/min onto the trapping column for 6 minutes. Subsequently, peptides were eluted via the analytical column with a constant flow of 0.3 µL/min with increasing percentage of solvent B (3% DMSO, 0.1% formic acid in acetonitrile) from 2% to 4% in 6 min, from 4% to 8% in 1 min, then 8% to 25% for a further 41 min, and finally from 25% to 40% in another 5 min. The outlet of the analytical column was coupled directly to a Fusion Lumos (Thermo) mass spectrometer using the proxeon nanoflow source in positive ion mode.

The peptides were introduced into the Fusion Lumos via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and an applied spray voltage of 2.4 kV. The capillary temperature was set at 275°C. Full mass scan was acquired with mass range 375-1200 m/z in profile mode in the orbitrap with resolution of 120000. The filling time was set at maximum of 50 ms with a limitation of $4x10^5$ ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 30000, with a fill time of 86 ms and a limitation of $2x10^5$ ions. A normalized collision energy of 34 was applied. MS^2 data was acquired in profile mode.

Data analysis

Acquired data were processed by IsobarQuant (PMID: 26379230), as search engine Mascot (v2.2.07) was used. Data were searched against Uniprot *E.coli* proteome database containing common contaminants, reversed sequences and the sequences of the proteins of interest (*Homo sapiens*). The data were searched with the following modifications: Carbamidomethyl (C; fixed modification), Acetyl (N-term), Oxidation (M) and Phospho (STY) (variable modifications). The mass error tolerance for the full scan MS spectra was set to 10 ppm and for the MS/MS spectra to 0.02 Da. A maximum of 2 missed cleavages was allowed. For protein identification, a minimum of 2 unique peptides with a peptide length of at least seven amino acids and a false discovery rate below 0.01 were required on the peptide and protein level.

SAXS

Small Angle X-ray Scattering data were collected at the bioSAXS beamline BM29 (Pernot *et al.*, 2013) at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, with a Pilatus 1M detector (Dectris Ltd, Baden, Switzerland) at a wavelength of 0.931 Å and a camera length of 2.42 m. Scattering curves were measured from solutions of the mMKK6DD/p38α complexes in SAXS buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP). Measurements were performed at protein concentrations of between 0.4 and 4.5 mg/ml to verify that any inter-particle effects that may have been present could be accounted for and rule out their influence on the analysis. Data were processed using the ATSAS program package (Manalastas-Cantos *et al.*, 2021). Radii of gyration Rg were evaluated from Guinier

plots using PRIMUS and pair distance distribution functions p(r) were computed with GNOM (Svergun, 1992). The solution shapes of the complex were reconstructed from the experimental data (GNOM functions) using the *ab initio* method (DAMMIN). For each sample twelve independent DAMMIN reconstructions were aligned, averaged and filtered using the program package DAMAVER.

Additional SAXS data were collected at the bioSAXS beamline B21 at the Diamond Light Source (DLS), UK. Buffer subtractions and all other subsequent analysis were performed with the program Scatter (Förster, Apostol and Bras, 2010).

X-ray crystallography

Crystallization

Initial crystallization experiments were carried out at the High-Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble (Dimasi *et al.*, 2007; Dupeux *et al.*, 2011). Crystals were harvested using CrystalDirect technology (Zander *et al.*, 2016).

Negative stain electron microscopy

Grafix crosslinking

Grafix crosslinking (Kastner *et al.*, 2007) was performed on purified mMKK6DDGRA/p38α complex as follow. Sample (150 µl) was loaded on top of a 10-30% galactose and glutaraldehyde gradient in sample buffer. The tube was centrifuged at 50 000 rpm for 18 hours in a SW60i rotor in a Beckmann ultra-centrifuge. 150 µl fractions were collected manually from the top and crosslinking was quenched with 10 mM glycine. Fractions containing the cross-linked complex were detected by dot-blotting on PVDF membrane and silver stain. Buffer exchange was performed using Zebra spin desalting columns (Thermo Scientific).

Grid preparation

Sample (5 µl) was applied to glow-discharged carbon coated copper grid (300 mesh, Electron Microscopy Science) and left to absorb for 30 seconds. After removing the excess of liquid, the grid was stained by incubating for 30 seconds on top of a 30 µl drop of 2% uranyl acetate solution; the grid was then blotted on wet paper and air-dried.

Data collection and processing

Data were manually collected on a FEI Tecnai T12 electron microscope at 49 kx operating at 120 kV equipped with a Ceta 16 M camera.

The collected 76 micrographs were processed using the Relion 2.1 package. Approximately 2000 particles were manually picked. Two of the best classes from 2D classification were used as template for auto picking, which resulted in ~18 600 particles. A subset of 2600 particles

was selected from the 13 most populated 2D classes and a *de novo* 3D initial model was then generated

CryoEM

Grid preparation

UltraAufoil 1.2/1.3 grids were glow-discharged for 30 seconds at 25 mA (PELCO easy glow). During the vitrification procedure on a Vitrobot Mark IV (FEI), 4 µl of sample at 7 µM concentration was applied to the grid. The blotting force was set to 0 for a total time of 3.5 seconds.

Screening and data collection

Grids were then clipped and screened on a FEI Talos Glacios electron microscope (EMBL Grenoble) operating at 200 kV. Micrographs were recorded on a Falcon3 direct electron detector operated in counting mode using the EPU software (FEI), at a nominal magnification of 150 kx (corresponding to 0.96 Å/pixel at the specimen level), with a defocus range of -0.7 to -2.5 µm. Data collections were monitored on-the-fly, using WARP software (Tegunov and Cramer, 2019).

Selected grids were then sent for data collection on a FEI Titan Krios (EMBL Heidelberg) operating at 300 kV. Micrographs were automatically collected using SerialEM software (Schorb *et al.*, 2019) from a K2 Quantum detector (Gatan) and a GIF Quantum energy filter (Gatan), at a nominal magnification of 215 kx (corresponding to 0.638 Å/pixel at the specimen level). A total of 9132 movies of 40 frames were collected in electron counting mode with a defocus range of -1.5 to -3 µm, 0.1 step, with a total dose of 62.77 $e^-/Å^2$.

Data processing

All datasets were processed using WARP for particle picking and CryoSPARC v3.2.0 (Punjani *et al.*, 2017) for all the other steps.

The details of the data processing for the Krios dataset are described here. For this particular dataset, it was necessary to train the picking tool to pick and select good particles.

Movies were imported in CryoSPARC. All movie frames were aligned and motion-corrected with Patch Motion correction and CTF estimated with Patch CTF. Particles were initially picked in WARP and extracted in CryoSPARC. Data were manually curated to eliminate micrographs with large motions, poor resolution, and high ice thickness, resulting into the selection of 6 900 movies containing a total of 1 901 213 particles. The first round of 2D classification was performed (100 classes) using a mask of 100 Å in diameter and limiting the resolution to 8 Å for better particle alignment.

Junk classes were discarded and the resulting 1 222 665 selected particles were submitted to *ab initio* modelling job requesting 3 models. The classes appearing to contain the hetero-dimer (400 681 particles) was further processed with heterogeneous and homogeneous refinement, followed by a second cycle of new *ab initio* reconstruction, heterogeneous and

homogenous refinement. The heterogeneous refinement job combines 3D classification with refinement, whilst homogenous refinement job refines the map using the gold standard FSC procedure. Multiple rounds of 2D classification were then applied to particles corresponding to the hetero dimer and 98 247 particles were finally selected (Appendix 1A, p. 206).

Topaz particle picker (Bepler, Noble and Berger, 2019) was trained on a subset of 500 micrographs using the best obtained 2D classes. A total of 259 447 particles were picked and 2D classified. Selected 2D classes were used for *ab initio* reconstruction followed by heterogeneous and homogeneous refinement (Appendix 1B, p. 206). One class containing 49 665 particles corresponded to the complex of interest and was used to retrain the Topaz picker, and a new dataset of 3 710 684 particles were picked on 6900 micrographs. From this step, an iterative process of 2D-classifications, *ab initio* reconstructions, heterogeneous and homogeneous refinement was performed to further eliminate particles not corresponding to the desired kinases hetero-dimer (junk, poorly aligning particles, particles of dissociating complex or open conformation). Due to an expected close proximity of particles, we also eliminated particles closer than 60 Å. The presented 3D reconstruction is the product of the homogeneous refinement of a sub-selection of 9 108 particles (Appendix 1C, p. 206).

Results

How to express and stabilize a functional MKK6/p38α complex in the transition state

Engineering of an MKK6/GRA24 chimera to stabilize the interaction with p38α

The interaction between MKK6 and p38α is transient and difficult to stabilize. The co-purification of these two kinases was attempted before but remained unsuccessful (Pellegrini, 2013). Kumar and co-workers recently published a study in which they managed to isolate an MKK6/p38α complex on gel filtration (Kumar, Page and Peti, 2021), but too few details in the materials and methods do not allow the reason for their success to be determined.

To study the complex of these two proteins, we engineered a construct of an activated mutant (S207D T211D) of MKK6 where its N-terminal KIM motif was replaced by the KIM sequence of the *T. gondii* GRA24 protein (MKK6DDGRA). GRA24 KIM peptide has a 100x higher affinity (~1.6 µM K_D for GRA24 KIM vs ~160 µM for MKK6 KIM) for p38α than the MKK6 KIM peptide, but induces a similar allosteric conformational change on p38α upon binding (Pellegrini, Palencia, *et al.*, 2017). The activation loop residues Ser207 and Thr211 of MKK6 were mutated to aspartate residues to mimic phosphorylation and confer constitutive activation to the MAP2K (Figure 13).

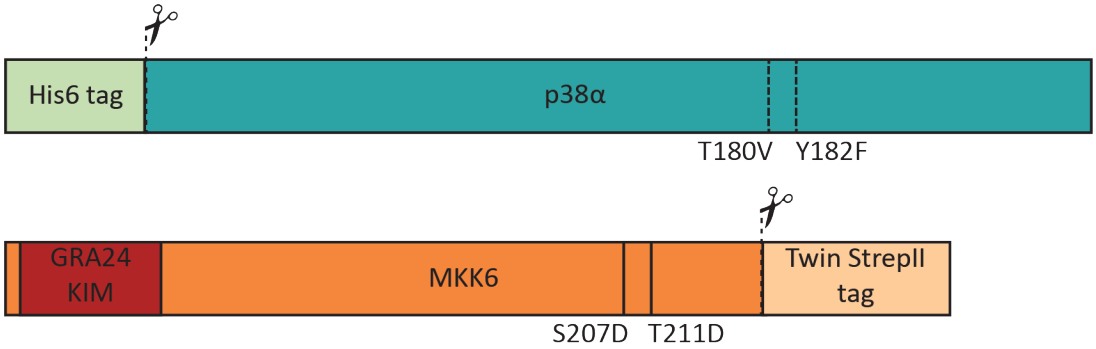


Figure 13: Final recombinant protein constructs designed for biochemical and structural studies. p38α constructs are tagged with an N-terminal His6 tag fused to a TEV or 3C protease cleavage site. Single point mutations (T180V and Y182F) were introduced. MKK6DDGRA constructs are composed of the GRA24 KIM1 sequence fused to the MKK6 sequence with phosphomimetic mutation on the activation loop (S207D T211D). The constructs are tagged with a C-terminal Twin StrepII tag linked attached to a 3C protease cleavage site.

Initially, p38α constructs were designed with a 9 residue N-terminal truncation. The N-terminal extension of p38α is quite flexible and we first decided to truncate it for better crystallization potential as the N-terminus can make contact with the docking site of another p38 molecule potentially inhibiting complex and/or crystal formation. However, it is not known whether the p38α N-terminus is involved in the interaction with MKK6. We therefore decided to work with the FL protein for cryoEM studies. In this chapter, we used the WT p38α

sequence as well as mutants of the activation loop residues Thr180 and Tyr182. p38α T180V and p38α Y182F were designed to only have one phosphorylation site on the activation loop.

Expression and purification optimization

Strategy 1: separate expression in *E. coli* and co-purification

We initially produced the mMKK6DDGRA/p38α complex as follows: separate expression of untagged mMKK6DDGRA and His6 tagged p38α in *E. coli* with IPTG induction, followed by co-sonication and co-purification of the complex by nickel affinity, ion exchange and size exclusion chromatography. This protocol led to the production of a pure and stable mMKK6DDGRA/p38α complex (Figure 14).

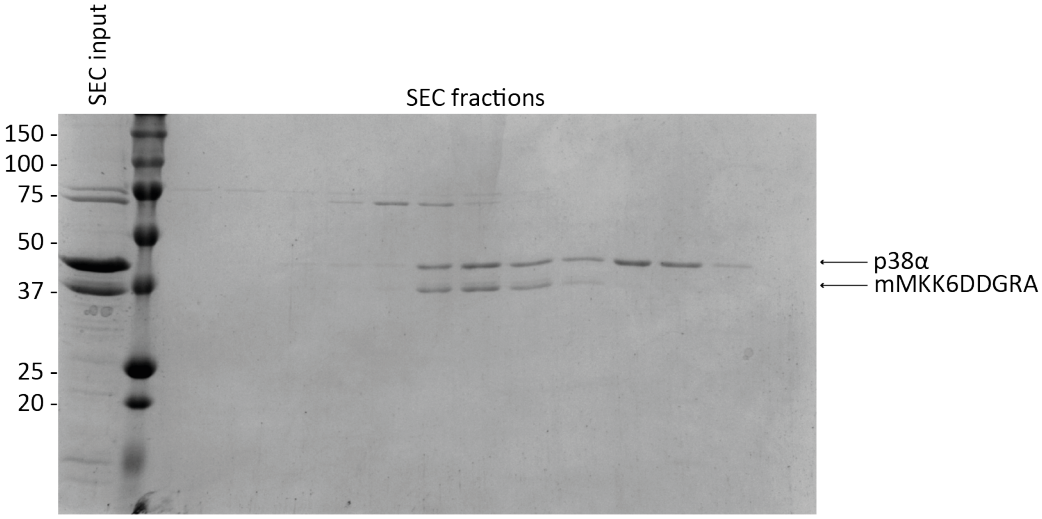


Figure 14: SDS-PAGE gel of the mMKK6DDGRA/p38α complex obtained with the initial expression and purification strategy.

However, the production yield was low and with some inconsistencies between batches (~0.07 mg of pure MKK6DDGRA/p38α sample per litre of culture). This limited the possibilities of structural studies of the sample. The expression of p38α WT and mutants resulted in milligrams of protein per litre of culture. The main limiting element was the expression and stability of the mMKK6DDGRA construct in *E. coli*. This can partially be explained by the high activity of the activated mutant that could induce toxicity to the cells. Several strategies were explored to improve the sample production and quality.

Co-expression with lambda phosphatase was previously described to enhance the expression of kinases in *E. coli* (Parton *et al.*, 2016). Unfortunately, it was not the case for mMKK6DDGRA expression.

The mMKK6DDGRA/p38α complex was sensitive to NaCl concentration, and was not stable during ion exchange chromatography. Nickel affinity chromatography (based on the His6-tagged p38α) and size exclusion chromatography were not sufficient to achieve desired purity for structural studies.

The purification buffer was optimized using the Thermofluor assay proposed by the HTX platform (LaCava *et al.*, 2016). This assay gave us valuable information to choose the best buffer and purification strategy for the mMKK6DDGA/p38α complex. The sample was

unstable at acidic pH but tolerated well basic ones. The presence of NaCl was necessary for stability, but the sample was destabilised at salt concentrations higher than 250 mM (confirming the impracticality of an ion-exchange purification step). HEPES buffer slightly increased the stability of the sample in comparison to Tris buffer, and has the advantage to be compatible with most technics such as crosslinking and mass spectrometry. We therefore opted for the following buffer: 50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP.

We evaluated the possibilities for tagging mMKK6DDGRA both for higher solubility and better purification. However, MKK6 interacts with p38α through its N-terminal extension, therefore, the possibilities of expressing it with an N-terminal tag such as MBP or SUMO to enhance its solubility and expression was not explored in order to avoid interference with the formation of the complex. We decided to add a C-terminal 3C cleavable twin StrepII tag, to the mMKK6DDGRA construct.

To address the low yield limitation, we expressed this new construct using auto-induction media (Studier, 2005) instead of IPTG induction. Combined with the additional step of StrepTactin affinity chromatography, the production of the sample was achieved with a higher but still limiting yield (~0.25 mg of pure MKK6DDGRA/p38α sample per litre of culture).

The resulting sample was used for characterization by negative stain EM and SAXS.

Strategy 2: Separate expression in insect cells (MKK6DDGRA) and *E. coli*, separate initial purification and co-purification by gel filtration

The MKK6DDGRA construct with a C-terminal 3C cleavable twin StrepII tag was cloned into a BAC plasmid for insect cells expression.

The MKK6DDGRA/p38α complex was obtained using the following protocol. The His6 tagged p38α was co-expressed with lambda phosphatase in *E. coli* by IPTG induction and purified by nickel affinity chromatography. The tag was then cleaved by TEV/3C protease and separated from the cleaved protein by a second step of nickel affinity chromatography (Figure 15B). The MKK6DDGRA construct was expressed in Sf21 insect cells and purified by StrepTactin affinity chromatography. The tag was cleaved by 3C protease and separated from the cleaved protein by a second step of StrepTactin affinity chromatography (Figure 15A).

The two proteins were then mixed in equimolar ratio, incubating on ice for 30 minutes to 1 h. The sample was then concentrated up to ~10mg/ml, and finally purified by size-exclusion chromatography (~2mg of pure MKK6DDGRA/p38α sample per litre of insect cell culture expressing MKK6DDGRA), aliquoted and kept at -80°C for further studies (Figure 15C).

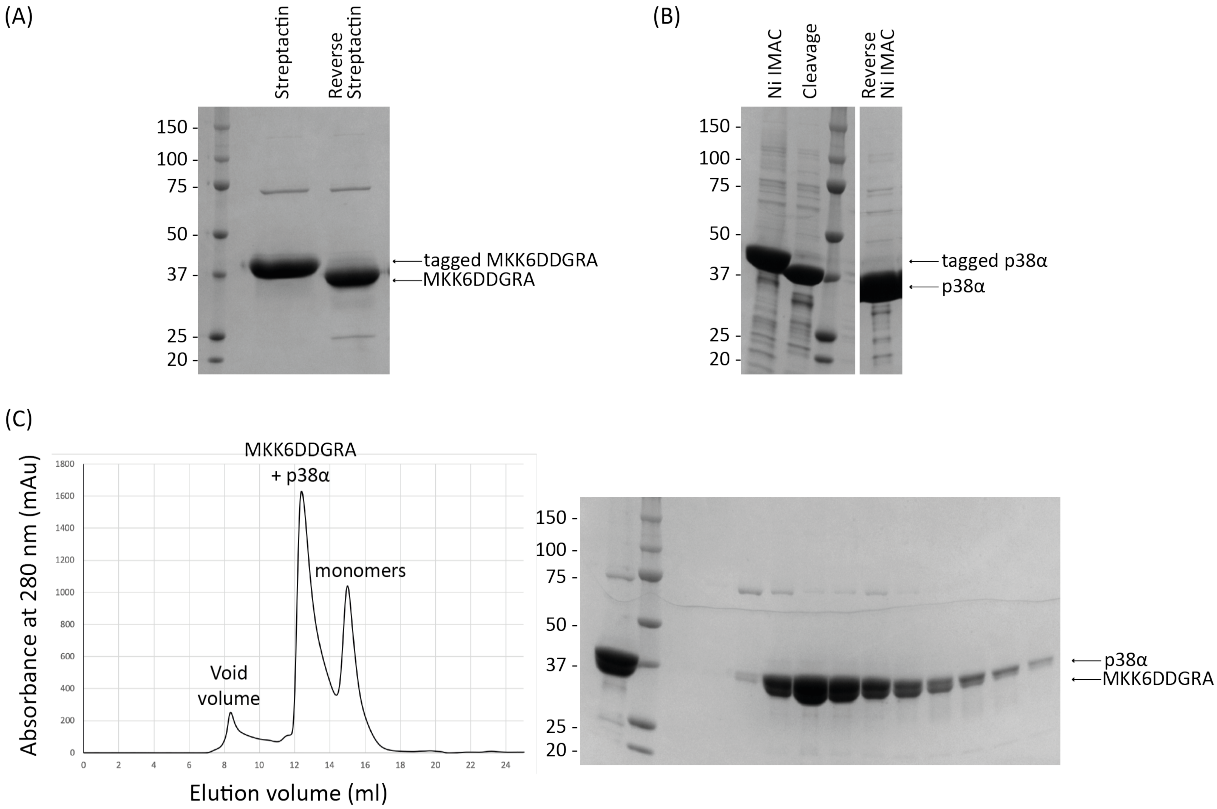


Figure 15: MKK6DDGRA/p38α complex purification. (A) MKK6DDGRA affinity purification with Streptactin. (B) p38α affinity purification with Ni IMAC. (C) MKK6DDGRA/p38α purification by size exclusion chromatography.

Control of the MKK6DDGRA/p38α complex phosphorylation state and transition states analogues

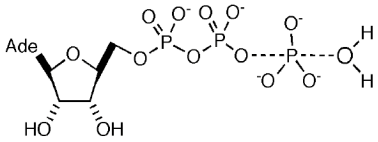
We are interested in understanding the specificity and the activation mechanism of p38α by MKK6. To structurally characterize this activation process, we aimed to capture the complex in an intermediate phosphoryl transfer conformation by using either non-hydrolysable nucleotide analogues or transition state analogues. With this in mind, we decided to prioritise expression and purification strategies where we could keep p38α unphosphorylated and where we could control the formation of the complex in the presence of the desired nucleotide analogues. This therefore excluded co-expression strategies for p38α and MKK6.

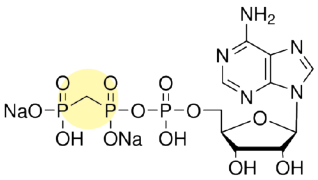
Despite co-expression of lambda phosphatase with mMKK6DDGRA increased expression of the construct failed but the co-expression of p38α with lambda phosphatase was revealed useful for our study: it led to the production of homogeneous and non-phosphorylated p38α as confirmed by mass spectrometry. This allowed us to test and control the phosphorylation events.

We tested both p38α WT and T180V and Y182F single mutants of the activation loop residues to try to help in the stabilization of the complex in an intermediate phosphoryl transfer

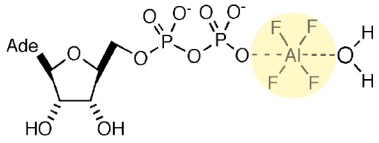
conformation. To capture the complex in either pre-catalytic state or transition state, we evaluated the nucleotide analogue AMP-PCP, as well as the transition state analogues (AlF_4^- combined with ADP) (Cliff *et al.*, 2010; Jin *et al.*, 2017).

Phosphoryl transfer transition state





AMP-PCP

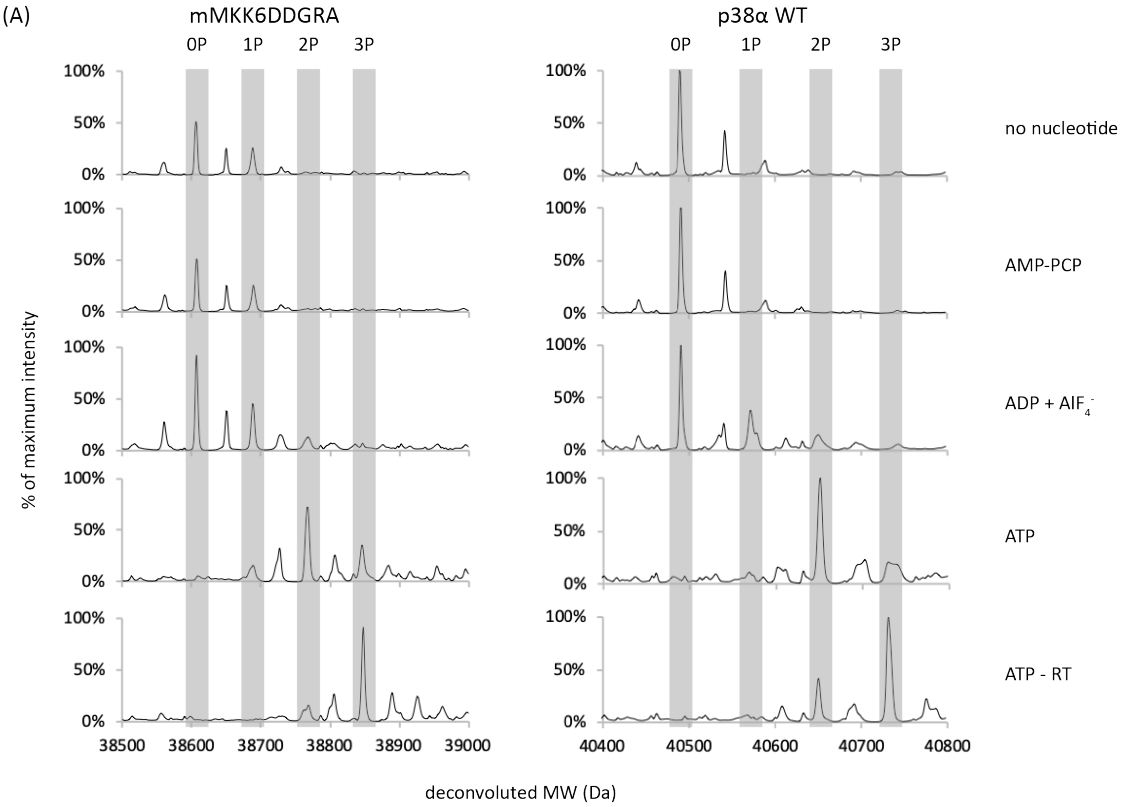


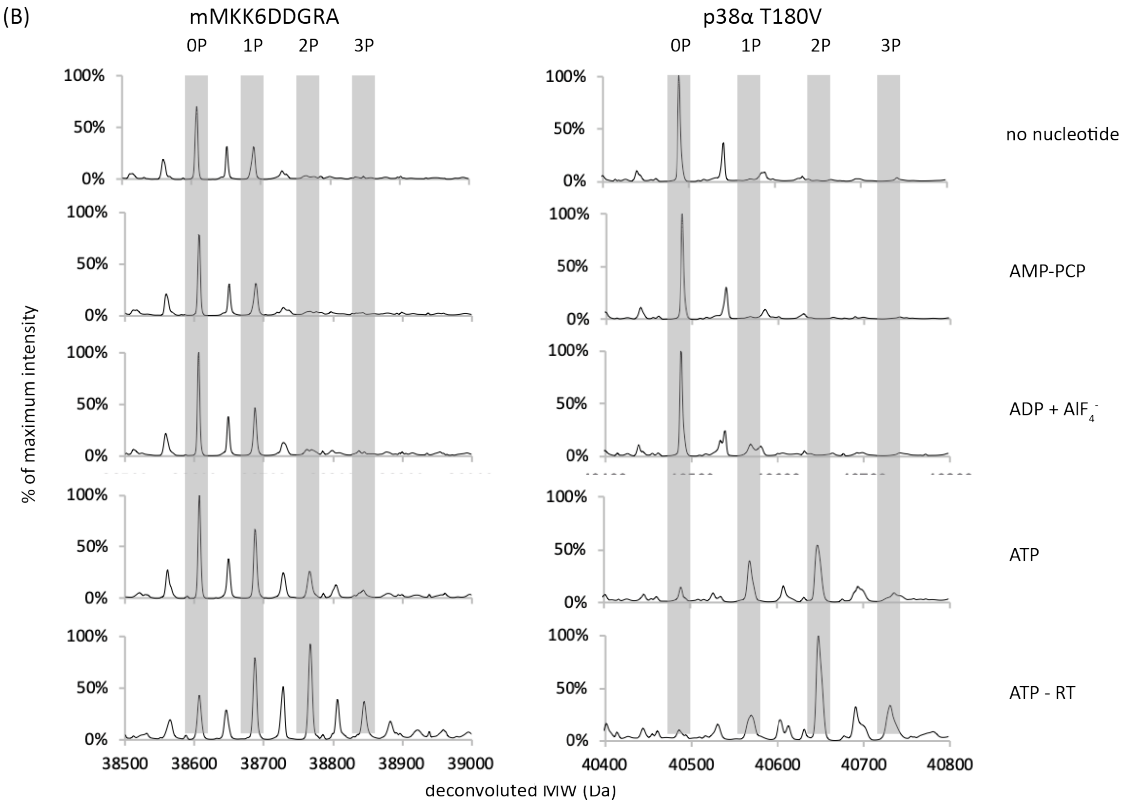
AlF_4^- +ADP

Figure 16: To trap the sample in an intermediate phosphoryl transfer conformation, we evaluated AMP-PCP non-cleavable ATP analogue, and ADP in combination with AlF_4^-.

An active and functional complex

We performed MS experiments to control the quality and homogeneity of the samples and to assess the ability of mMKK6DDGRA to phosphorylate p38α. We analysed SDS-PAGE bands by





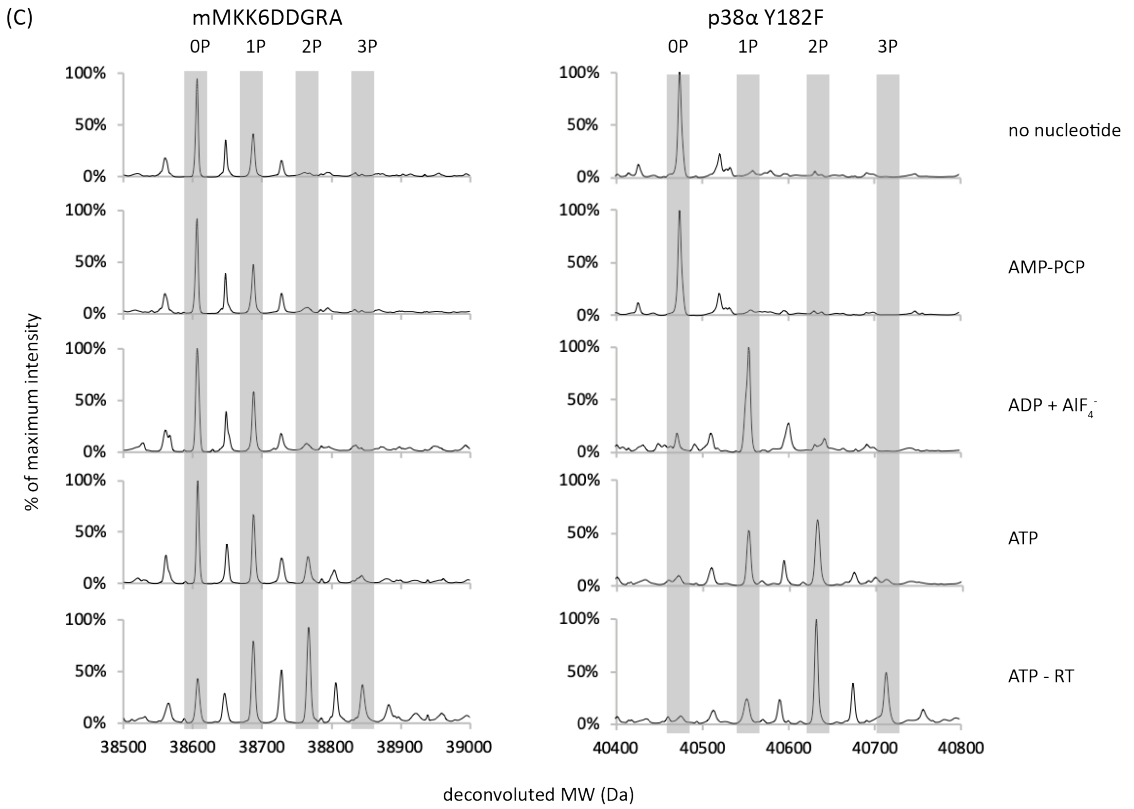


Figure 17: native ESI-QTOF MS spectra of mMKK6DDGRA with p38α (A) WT, (B) T180V, and (C) Y182F samples incubated on ice with 10 mM AMP-PCP, ADP + AlF_4^-, ATP or without nucleotide, or incubated at RT with 10 mM ATP. For each sample, peak for each phosphorylation form (increased mass of 80 Da for each phosphorylation) is highlighted.

LC-MS/MS to confirm their MW and phosphorylation state (not shown). We complemented this analysis with native ESI-QTOF MS experiment (Figure 17) as well as LC-MS/MS detection of phosphorylated sites on sample in solution (Figure 18 and Appendix 2, p. 210).

The samples were composed of the mMKK6DDGRA/p38α complex incubated with different nucleotides. We tested p38α WT, T180V and Y182F single mutants.

p38α

Without nucleotide, p38α recombinant proteins were homogeneous and mainly non-phosphorylated. A mass deviation of ~5 Da compared to the expected mass was consistently detected for all p38α independently of the p38α construct (WT or mutant) and of the expression and purification batch, and was not explained by measurement error (the sensibility of ESI-QTOF MS measurement is +/- 2 Da). All plasmids were sequenced and were confirmed to have the expected sequence.

In the presence of AMP-PCP, a non-cleavable ATP analogue, p38α constructs were non-phosphorylated, as without nucleotide.

Surprisingly, when incubated with ADP + AlF_4^-, some phosphorylation was detected on p38α, WT and mutants, especially in the LC-MS/MS analysis of phosphorylated sites (Figure 18 and Appendix 2, p. 210). This novel phosphorylation mechanism in presence of ADP will be

explored in Chapter II: Novel mechanism of activation of p38α by MKK6 using ADP as a source of phosphate (p. 99). Looking at the intact mass results obtained by ESI-QTOF MS experiments (Figure 17), the phosphorylated population of p38α does not seem abundant compared to the non-phosphorylated population. Samples for structural studies were usually prepared and analysed at 4°C, and at this temperature the ADP consumption is even less efficient. We decided to pursue the characterization of the complex with ADP + AlF_4^-.

In the presence of ATP, p38α WT and mutants are phosphorylated. When incubated on ice, p38α WT is mainly biphosphorylated on both sites of its activation loop Thr180 and Tyr182. When incubated at RT, Thr180 and Tyr182 are also phosphorylated as well as some other sites. The T180V and Y182F single mutants can also be phosphorylated by mMKK6DDGRA, with one phosphorylated site less compared to the WT.

An alternative p38α minor population was also present in the samples, with a mass deviation of ~50 Da compared to the expected mass, for the WT and mutant constructs. It behaves similarly to the main p38α population, with increasing phosphorylated site number when adding ADP and ATP and was considered negligible.

These results confirmed that the mMKK6DDGRA construct is able to phosphorylate on both sites of the activation loop of p38α.

mMKK6DDGRA

Regarding the mMKK6DDGRA construct, as for p38α constructs, two populations were detected. The less abundant one corresponded to the expected mass of mMKK6DDGRA with a truncation of its N-terminal methionine (-130 Da). The main mMKK6DDGRA population presented a mass deviation of ~84 Da less compared to the expected mass, equivalent to a mass deviation of 45 Da extra compared to the truncated N-terminal methionine form.

When incubated with p38α WT or mutants in the absence of nucleotide, the Thr28 of mMKK6DDGRA was partially phosphorylated (Figure 17 and Appendix 2, p. 210). This phosphorylation could result from the recombinant expression in insect cells.

The incubation with AMP-PCP did not affect the phosphorylation profile of mMKK6DDGRA compared to the samples without nucleotide (Figure 17).

When ATP was added to the complex, either incubated on ice or at RT, several amino acids present in the disordered MKK6 N-terminal extension were phosphorylated, along with a few minor sites over the protein sequence (Figure 17 and Appendix 2, p. 210). These phosphorylations could result from the activity of biphosphorylated p38α or mMKK6DDGRA auto-phosphorylation, and indicate a highly activated complex.

Interestingly, the same sites were partially phosphorylated when the mMKK6DDGRA/p38α complexes were incubated with ADP + AlF_4^-, in reduced proportion compared with ATP incubation.

Further analysis of the origin of the phosphorylation with ADP and ATP will be presented in Chapter II: Novel mechanism of activation of p38α by MKK6 using ADP as a source of phosphate (p. 99).

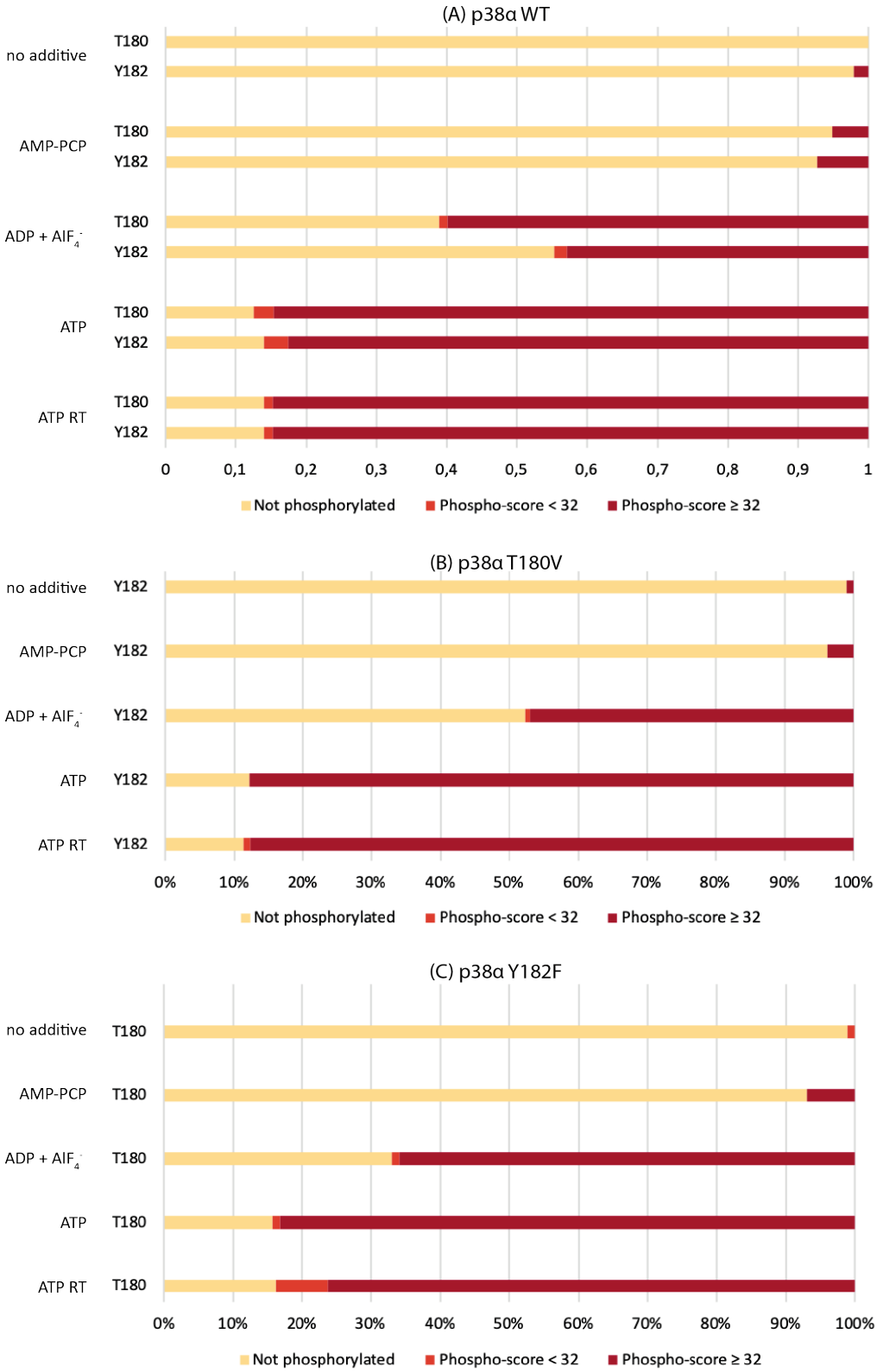


Figure 18: Phosphorylation state of p38α Thr180 and Tyr182 residues from the LC-MS/MS analysis of mMKK6DDGRA with p38α WT (A), T180V (B), and Y182F (C) samples incubated on ice with 10 mM AMP-PCP, ADP + AlF_4^-, ATP or without nucleotide, or incubated at RT with 10 mM ATP. For each analysed sample, the results of phosphorylation state of Thr180 (top) and Tyr182 (bottom) residues of p38α are represented in % of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (MASCOT phospho-score bellow 32) (orange) and phosphorylated (MASCOT phospho-score higher than 32) (red) by MASCOT analysis.

This extensive analysis was performed using the mMKK6DDGRA and ΔN9 p38α constructs. It was partially repeated with the hMKK6DDGRA and FL p38α constructs that displayed similar behaviour.

Overall, the recombinant complexes of MKK6DDGRA (mouse or human) with p38α (ΔN9 or FL; WT, T180V or Y182F) are homogeneously expressed and mainly unphosphorylated when purified without nucleotide. They are highly active in presence of ATP (and in a lesser extent in presence of ADP). We concluded that they constitute good candidates for structural studies in combination with nucleotide analogues to trap the complex in transition state.

Towards a better structural understanding of the MKK6/p38α interaction

Attempts to solve the complex structure by X-ray crystallography

We first attempted to crystallize the MKK6DDGA/p38α complex. Different parameters for sample preparation were tested and sent for crystallization to the EMBL Grenoble HTX platforms (Table 4). Promising conditions were selected to design refinement crystallization screens and to test additive screens (Additives (Hampton) and Opti-Salt (Qiagen)).

Despite the appearance of crystalline like forms in several conditions, that almost all combinations of the parameters presented in Table 4 were tested, that several homemade refinement screens were set up, the ~120 harvested pins with crystals or crystal-like forms measured were either salt crystals or not diffracting.

Table 4: Parameters tested for sample preparation and crystallisation screens

Parameter	Options explored
MKK6 construct	• mouse MKK6DDGRA • human MKK6DDGRA
p38α construct	• WT • T180V • Y182F • FL • N-terminal truncation (10-360)
Nucleotides	• AMP-PCP • ADP + AlF_4^-
Buffer composition	• with 0.5 mM TCEP • without TCEP
Crystallisation setup temperature	• 4°C • 20°C
Sitting drop sample:reservoir ratio	• 1:1 • 2:1 • 1:3 • 3:1
Commercial crystallisation screens	• The Classics Suite (Qiagen) • The JCSG + (MD) • The PACT (MD) • The PEGs I (Qiagen) • Salt-Grid_hampton • Wizard I & II (Rigaku) In some cases: • Index Screen (Hampton) • Midas plus (MD) • Morpheus (MD) • ProPlex (MD)

The apo mMKK6DDGRA/p38α complex is present and stable in solution, but flexible

We then proceeded to characterise the complex with SAXS and negative stain EM. SAXS was first used to determine the average structure of the apo mMKK6DDGRA/p38α complex in solution (data collected on BM29, ESRF). It confirmed that the samples were homogeneous and without aggregation. The molecular weight estimation and dimensions were coherent

with the presence of the complex (Table 5). Results were not affected by sample concentration.

Table 5: SAXS data analysis

Sample	**Concentration (mg/ml)**	**0.40**	**0.95**	**4.41**
Guinier	Rg (nm)	3.33	3.10	3.44
	I(0)	49.0	45.8	52.0
Gnom	Rg (nm)	3.32	3.21	3.45
	D_{max} (nm)	11.7	10.8	12.1
Porod	Volume (nm3)	113	111	120
	MM Vol est. (kDa)	56-75	56-74	60-80

Ab initio modelling showed an elongated shape that may indicate flexibility in the complex (Figure 19) as the D_{max} is incompatible with a compact dimer. Our hypothesis is that the main population of the complex is interacting in a closed conformation and that a smaller population is in a more extended conformation, only connected through the interaction between the KIM domain and docking site.

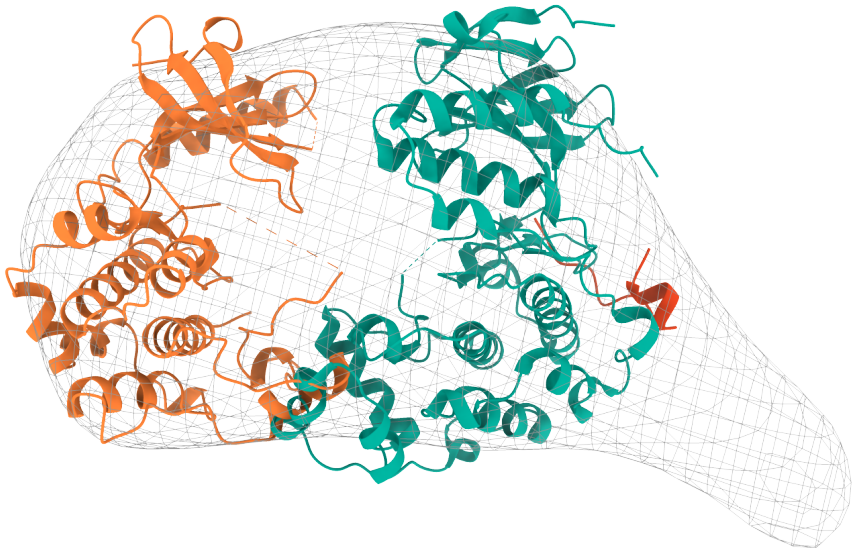


Figure 19: Ab initio model of the mMKK6DDGRA/p38α WT complex obtained from SAXS. Ab initio model (grey mesh) is represented with p38α (PDB 5eta, teal), MKK6 activated mutant (PDB 3fme, orange) and GRA24 KIM1 peptide (from PDB 5eta, red).

Crystal structures of p38α (PDB 5eta), MKK6DD (PDB 3fme) and GRA24 KIM peptide (PDB 5eta) were placed inside the *ab initio* model. These three structures are not shown as a model of interaction but to illustrate the size of a potential complex. The GRA24 KIM domain would bind the docking site of p38α and is connected to the MKK6 structure through its N-terminus (the first 46 amino-acids of the N-terminus of MKK6, including the KIM domain and the connecting extension, are missing from the 3fme structure).

Low resolution negative stain EM map of the crosslinked mMKK6DDGRA/p38α complex

The mMKK6DDGRA/p38α sample was then used to prepare negative stain grids for electron microscopy. However, we could not distinguish particles on micrographs, supposedly because of the small size of the complex combined with the flexibility observed with SAXS measurements and potential complex dissociation.

To address the issue of flexibility of the complex and stability during the specimen preparation, we evaluated the effect of crosslinking. Grafix crosslinking (Kastner *et al.*, 2007) was performed on the purified mMKK6DDGRA/p38α complex. From the grids prepared with the Grafix crosslinked sample we could generate micrographs in which we could distinguish homogeneous repartition of small particles for the first time (example of a representative micrograph in Figure 20).

Using this grid, 76 micrographs were manually collected on a FEI Tecnai T12 electron microscope at 49 kx at 120 kV. Data were processed using the Relion 2.1 package. ~2000 particles were manually picked and submitted to 2D classification. The best 2 classes from 2D classification were used as references for auto picking, which resulted in a total of ~18600 particles. Following the guidelines described in the Relion tutorial, ~9200 particles were selected based on their Z-score to generate 2D classes (Figure 21). Despite the low resolution inherent to negative stain combined with the small size of complex and few particles dataset, we can distinguish in some of the classes a dimeric assembly of two bi-lobed kinase folds.

After further 2D classification, a 3D initial model at ~20 Å resolution was generated with a subset of 2600 particles from the 13 most populated classes (Figure 22). The volume clearly shows two bi-lobed kinase shapes next to each other, allowing the orientation and relative position of the kinases to be determined. The crystal structures of p38α (PDB 5eta), MKK6 (PDB 3fme) and GRA24 KIM peptide (PDB 5eta) were manually placed into the 3D model.

This low-resolution model gives us insights on probable undescribed interacting regions between the MKK6 and p38α lobes. The extra density between the two kinases could be due to the N-terminal extension of MKK6 interacting with p38α: 31 amino acids are missing between the MKK6 structure and the GRA24 KIM peptide structure. It could also be the interaction between the p38α activation loop (disordered in PDB 5eta) and the MKK6 active site, including the activation loop of MKK6 (disordered in PDB 3fme). There is also a possible interaction between the bottom of the two C-lobes.

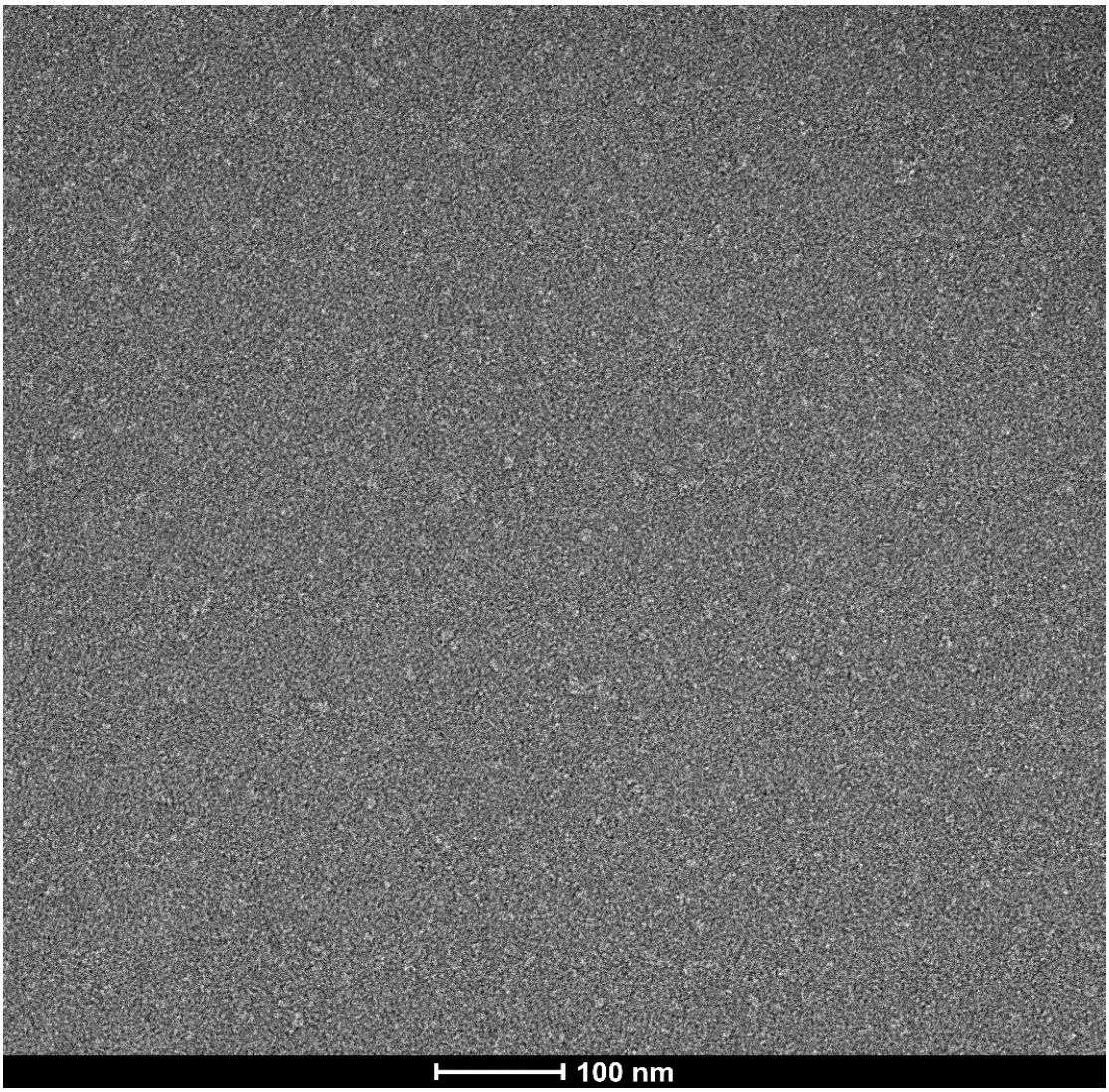


Figure 20: Example of one NS micrograph of the crosslinked mMKK6DDGRA/p38α complex sample taken on a FEI Tecnai T12 electron microscope operating at 120 kV at 49 kx magnification, equipped with a Ceta 16M camera.

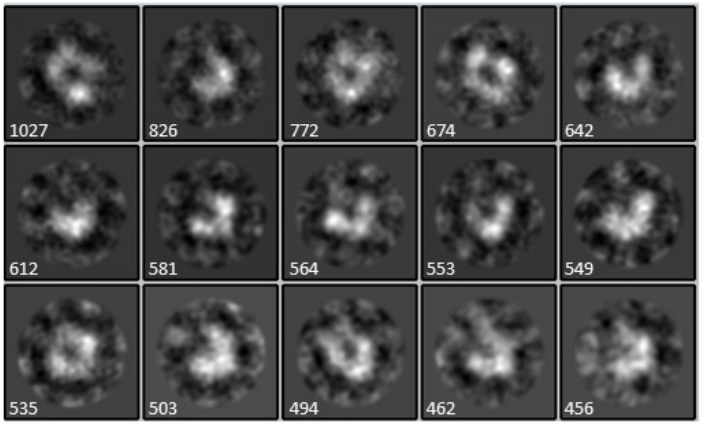


Figure 21: 2D classification of NS EM particles of crosslinked mMKK6DDGRA/p38α complex. The number of particles per class is indicated (box size = 20,6 nm).

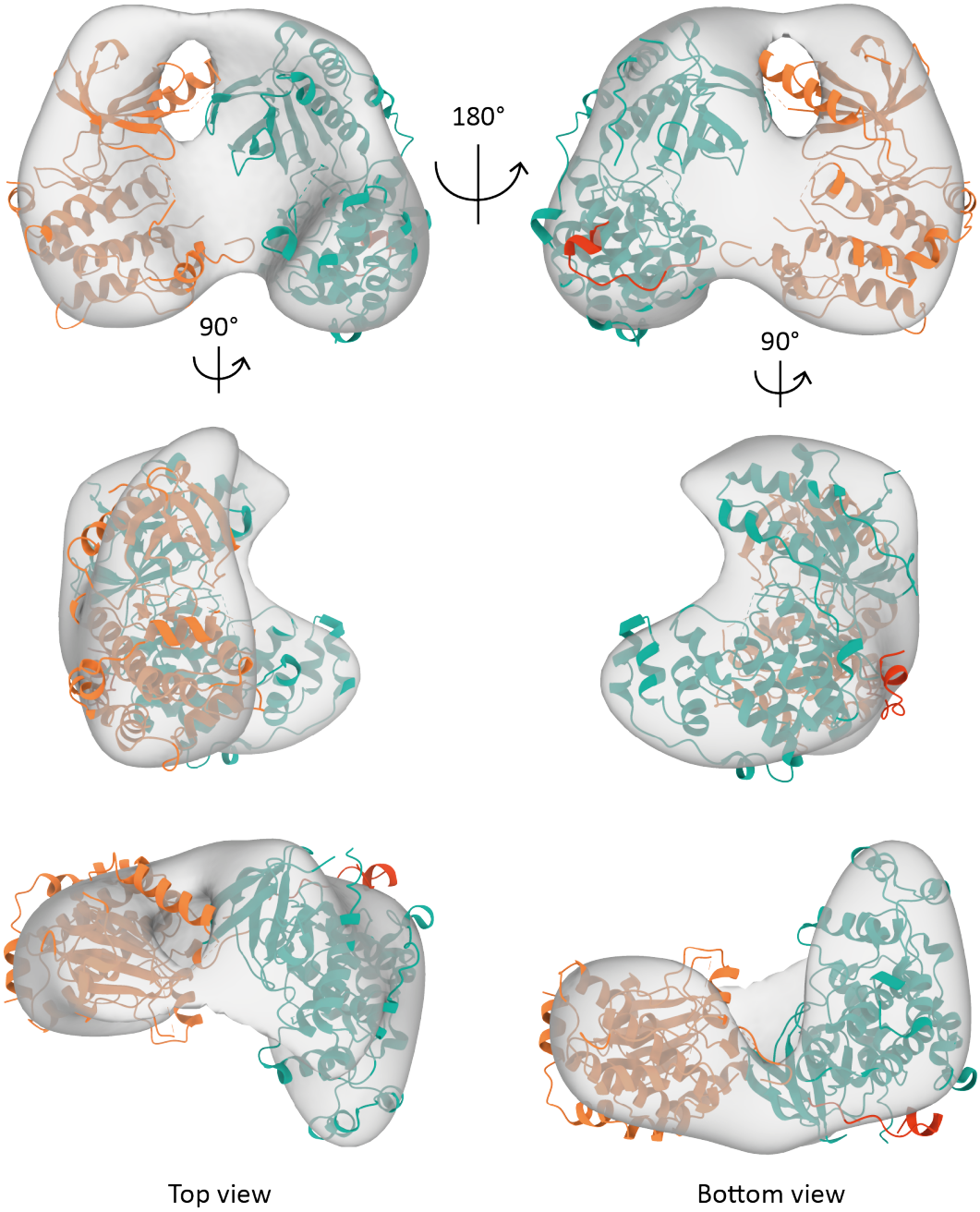


Figure 22: Different views of the 3D initial model of the crosslinked mMKK6DDGRA/p38α complex, fitting crystal structures of p38α (PDB 5ETA) in blue, GRA24 KIM peptide (PDB 5ETA) in red and MKK6 (PDB 3FME) in orange.

The mMKK6DDGRA/p38α T180V complex with ADP and AlF_4^- is in a more compact conformation

The observed flexibility of the mMKK6DDGRA/p38α complex confirmed the need for a strategy to stabilize it in a compact conformation. The crosslinking approach gave interesting results at low resolution with negative stain EM. However, Grafix crosslinking, as other gradient based crosslinking methods, offers less flexibility in terms of final sample volume and concentration. Therefore, in order to have a more reproducible sample to be used for further

structural studies, we decided to abandon for the moment this technique and to consider different complex assembly compositions, with different p38α constructs and nucleotides. We decided to evaluate these assemblies with SAXS before proceeding with cryoEM for the most appropriate complex.

We evaluated different p38α constructs: p38α WT, T180V and Y182F. These two mutants were designed to limit to only one phospho-acceptor site in the activation loop with the aim of stabilizing the complex in a single possible transition state. We also evaluated two different ATP analogues: AMP-PCP (a non-cleavable ATP analogue) and ADP in combination with AlF_4^-.

The different samples were characterised by SAXS to assess their homogeneity, size, and flexibility (data collected and processed by Mark Tully at B21 beamline (DLS)) (Table 6 and Figure 23).

Table 6: SAXS analysis of the mMKK6DDGRA/p38α complexes. Comparison of combination with p38α WT or mutants and with different nucleotides analogues (AMP-PCP and AlF_4^- + ADP).

	I(0)	**Rg**	**MW estimation (kDa)**	**dmax (Å)**
MKK6 + p38α WT + AMP-PCP	3,0E-01	40,99	68	133
MKK6 + p38α T180V + AMP-PCP	3,0E-01	39,45	70	132
MKK6 + p38α Y182F + AMP-PCP	2,3E-01	39,39	78	132
MKK6 + p38α WT + ADP + AlF_4^-	2,6E-01	32,37	66	114
MKK6 + p38α T180V + ADP + AlF_4^-	2,3E-01	29,58	60	109
MKK6 + p38α Y182F + ADP + AlF_4^-	2,8E-01	35,38	70	125

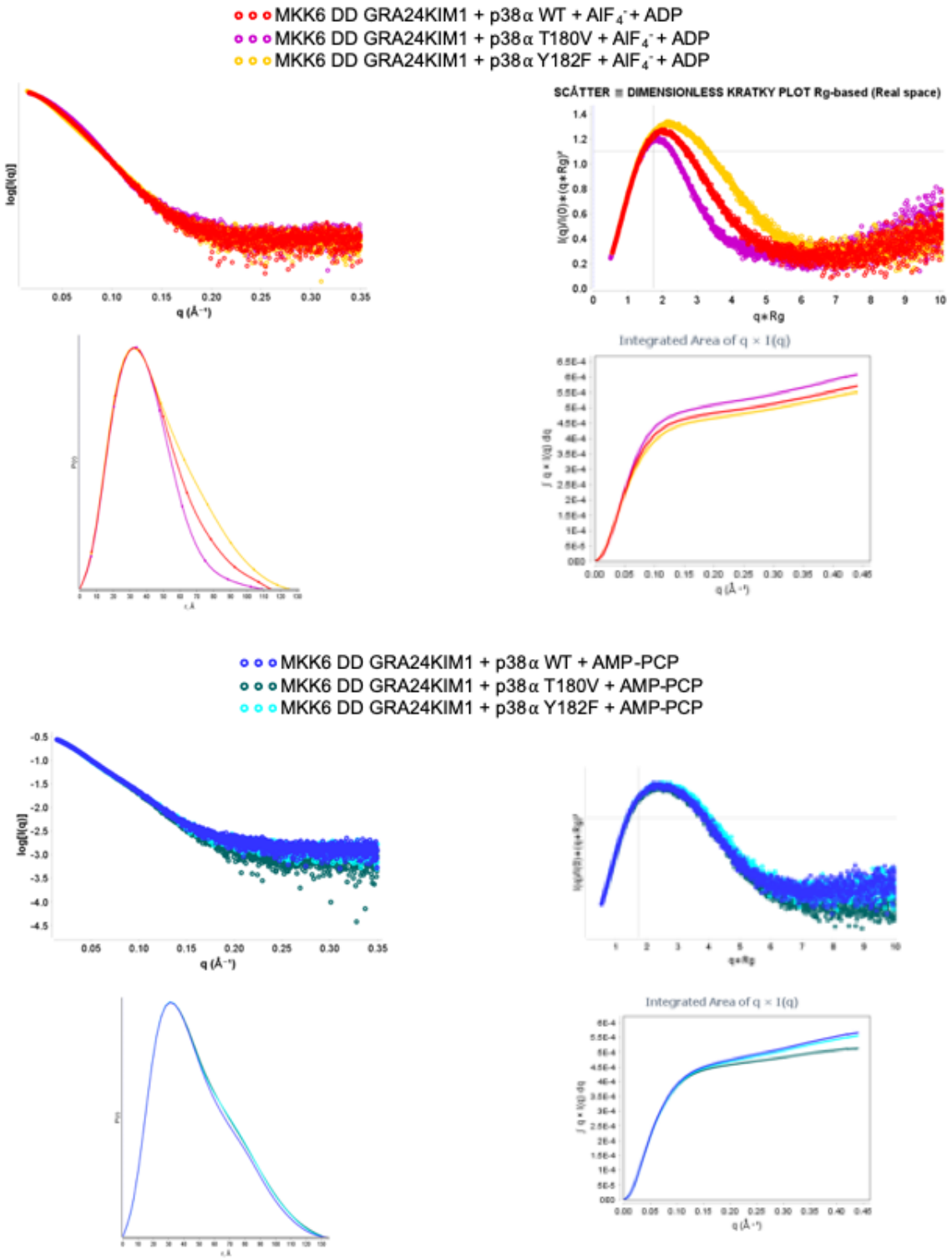


Figure 23: SAXS analysis of the MKK6/p38α complex. Comparison of combination with p38α WT or mutants and with different nucleotides analogues (AlF_4^- + ADP and AMP-PCP).

The results showed that the mMKK6DDGRA24/p38α T180V complex with ADP + AlF_4^- is in a significantly more compact conformation in solution than all the other complexes.

The quality of the collected data was not sufficient for 3D reconstruction.

Preliminary cryoEM map of the MKK6DDGRA/p38α T180V complex with ADP + AlF_4^-

Specimen optimization

Based on the SAXS results, we decided to apply cryoEM on a sample similar to the mMKK6DDGRA24/p38α T180V complex with ADP + AlF_4^-, changing for the hMKK6DDGRA and FL p38α T180V constructs.

We faced important challenges with this sample in cryoEM, mostly due to its small size (80 kDa) and heterogeneity. In particular, the evaluation of the quality of the micrographs was difficult by eye due to the low signal-to-noise ratio. Therefore, it was necessary to implement a protocol to properly screen grids and to investigate different cryoEM software for data processing.

We applied the sample to UltraAUfoil grids, as they are particularly recommended for samples smaller than 100 KDa (Herzik, Wu and Lander, 2018). We initially prepared grids with different sample concentrations and blotting times (2-5 seconds). In order to find the condition that corresponded to particles embedded in thin ice, we implemented a rigorous way to screen our grids at our in-house Glacios microscope operating at 200 kV and equipped with a Falcon3 camera. For each promising grid, we performed short data collections in counting mode on different squares (e.g. 10 holes per square, 4-6 squares per grid) and applied on-the-fly data processing using WARP software. We defined “good squares” as the ones for which WARP was returning CTF at high resolution (3.5-4.5 Å) combined with good picking (at least 150 particles at -0.7 µm defocus). The final optimized condition for grid preparation was using a sample concentration of 7 µM sample and 3.5 sec of blotting time on the Vitrobot.

Data processing was initially done in both Relion and CryoSPARC. Particle picking was performed in both packages and in WARP. 2D classes from particles picked by WARP were the only ones showing secondary structural features in a few classes and at least one projection where we could easily see two kinases in a face-to-face conformation. *Ab initio* modelling in CryoSPARC was then able to separate particles corresponding to a single kinase, hetero kinases complex and other complex conformations. After refinement of the hetero-kinases complex map, we obtained a density map where we could fit both kinases in a face-to-face conformation (Figure 24). Therefore, as a general processing strategy, we decided to combine CryoSPARC processing with picking in WARP.

We finally collected and combined two datasets on the Glacios electron microscope (~750 micrographs) that fulfilled our expectations: in the 2D classes we could identify both kinases (Figure 24A) and one of the 3D *ab initio* class corresponded to the expected hetero-kinase dimer (Figure 24B and C). This latest map at 12 Å resolution corresponds to a subset of ~33k particles from the ~139k particles picked by WARP.

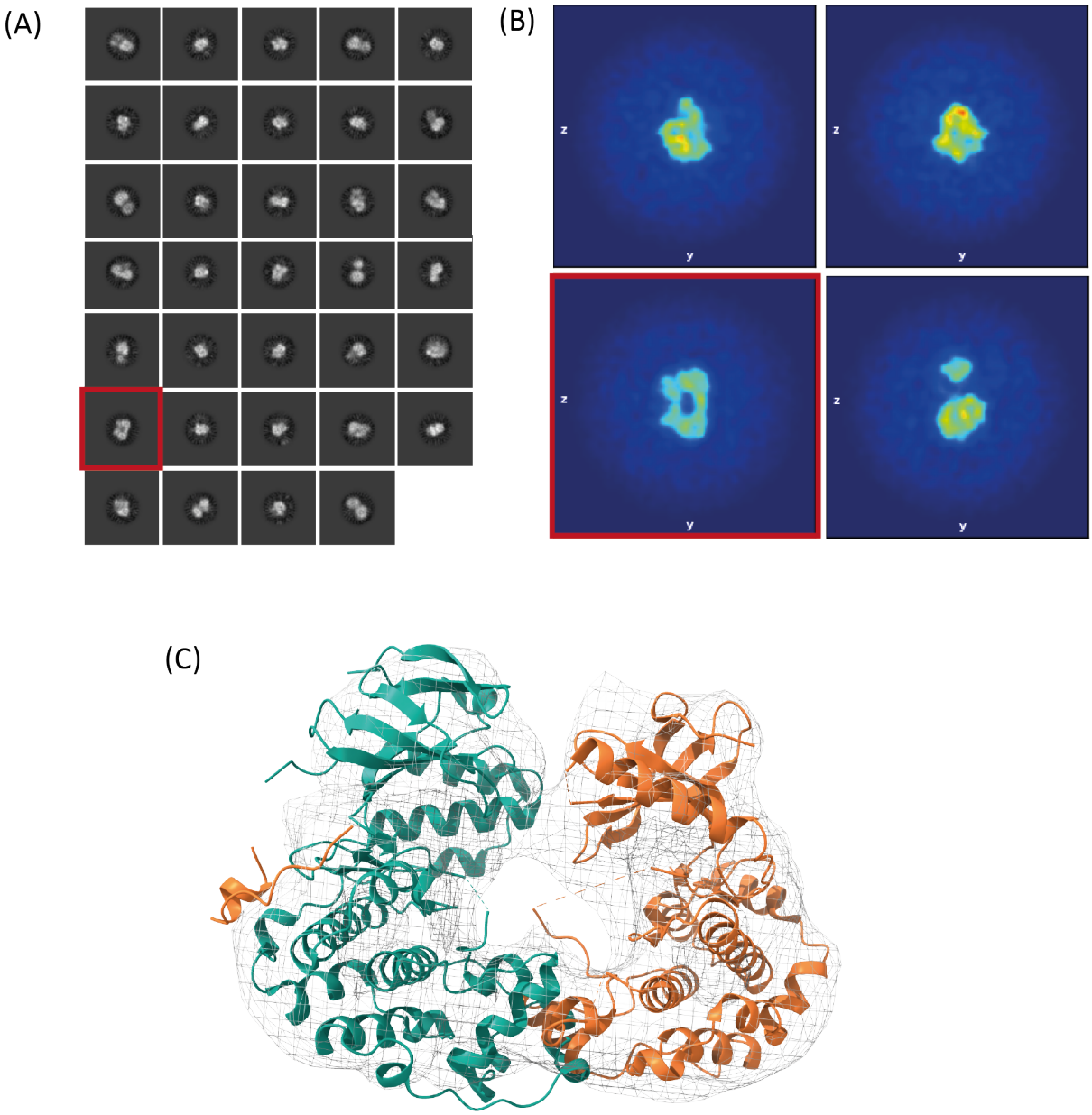


Figure 24: The MKK6/p38α T180 complex with ADP + AlF_4^- from the Glacios dataset (A) 2D classes. The two kinases in a face-to-face conformation can be observed in the class highlighted in red. The resolution was limited to 8 Å for alignment. (B) Projections through the four 3D classes generated by ab initio modelling from the selected 2D classes in (A). The orientation highlighted in red corresponds to the hetero-kinase dimer. (C) 3D density map obtained after the refinement of the class selected in (B) The crystal structures of p38α (teal) with GRA24 KIM peptide (orange) (PDB 5eta) and MKK6DD (orange) (PDB 3fme) were automatically fitted as rigid bodies into the density map using ChimeraX.

Data processing of the Krios dataset

We then proceeded with new grid preparation and screening on the Glacios electron microscope with the aim of performing a data collection on a Krios electron microscope operating at 300 kV. As data collection at Krios allows the collection of 6 images per hole (vs 1 image per hole on the Glacios), we expected to gain in number of particles and therefore in resolution. The selected grid from the screening session had homogeneous ice, homogeneous particle distribution with ~300-400 particles per micrograph picked by WARP at a defocus of -1 µm, and a good resolution estimation on the few micrographs collected (Figure 25).

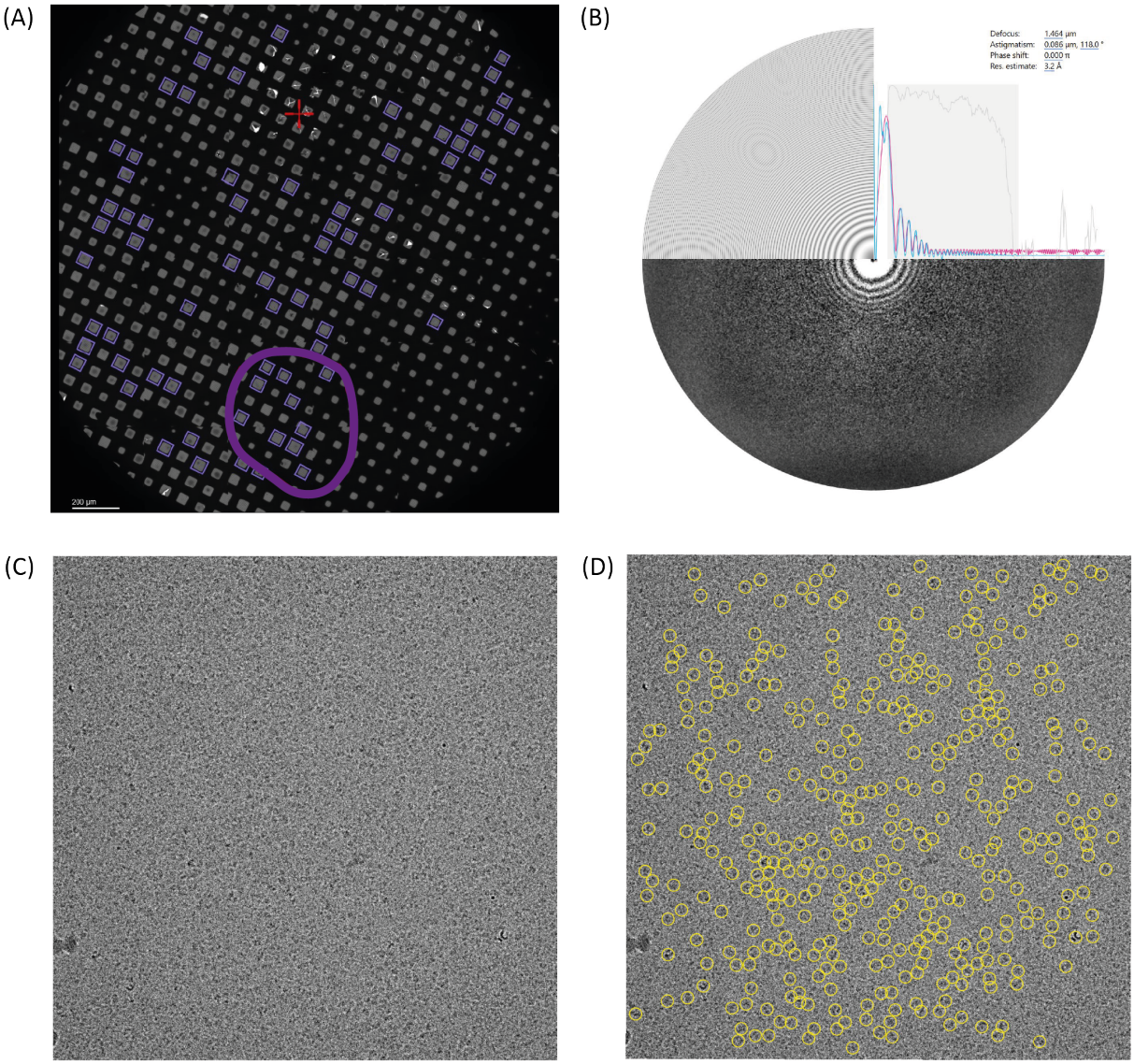


Figure 25: (A) Atlas of the grid indicating the region selected for data collection (B) CTF estimation (C) Micrograph with an estimated defocus of 1.5 µm (D) WARP picked 391 particles on the micrograph.

A dataset of ~9000 movies was collected in electron counting mode on a Krios Electron microscope at EMBL Heidelberg, with a defocus range of -1.5 to -3 µm, 0.1 step, and a total dose of 62.77 $e^-/Å^2$.

Unfortunately, the grids had been contaminated in between screening and data collection, resulting in a certain number of unusable micrographs, increased ice thickness and an even more challenging data processing (Figure 26).

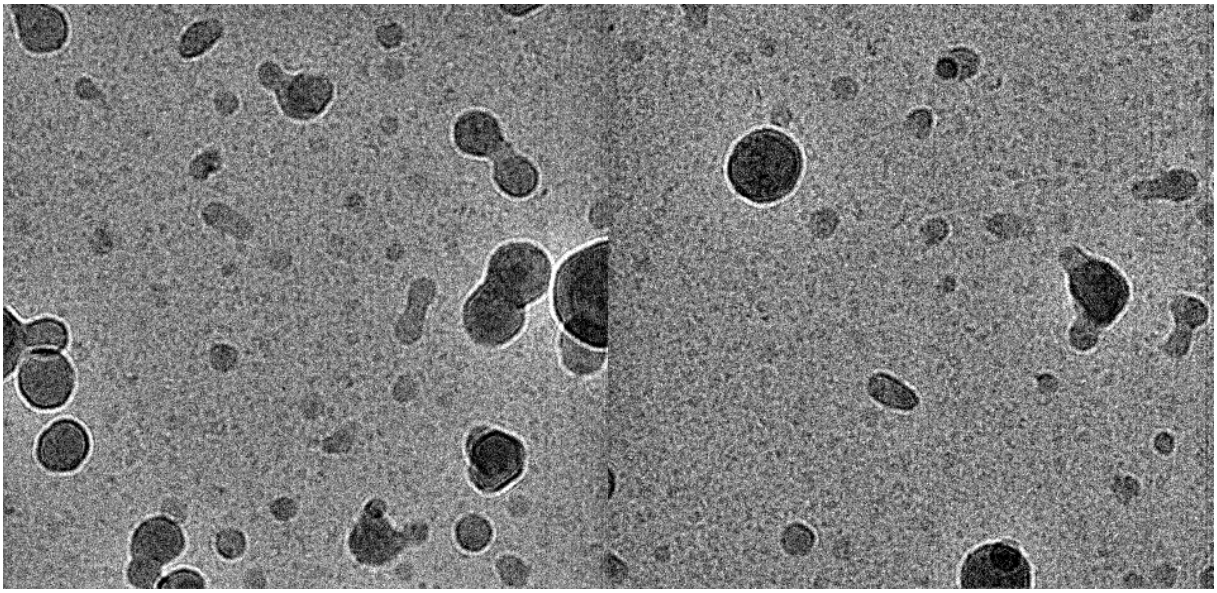


Figure 26: Examples of micrographs from the Krios data collection. Noticeable ice contamination can be observed in comparison to the micrographs from the same grid collected during the Glacios screening session (Figure 25).

Particle picking was not straightforward due to the size of the complex, the ice contamination of the grid and the poor signal-to-noise ratio. 2D classification of particles picked in WARP resulted in junk and poorly aligned classes. However, as the complex was present on the grid, we decided to proceed as follows (Appendix 1, p. 206).

We selected the few classes potentially containing particles corresponding to those observed in previous datasets and proceeded with several iterations of 2D classification, *ab initio* reconstruction, 3D classification and refinement in CryoSPARC, eliminating bad classes and junk at each step. We eventually obtained 2D classes that were similar to the expected complex (~ 90K particles, Appendix 1A). We then used these particles to train Topaz particle picker multiple times on a subset of 500 micrographs, until we obtained a 3D map corresponding to the expected complex (Appendix 1B, p. 206). We applied the optimised picking parameters on the full curated movie dataset, resulting in ~3 700 000 particles (Appendix 1C, p. 206). This new dataset of particles was processed with multiple iterations of 2D classification, *ab initio* reconstruction, 3D classification and refinement.

The final selected pool contained only 9 108 particles (Appendix 1C, p. 206). We could observe secondary structure features in the 2D classes, in particular the α-helices of C-lobes of the two kinases (Figure 27A). The final density map resolution was estimated to 7.93 Å based on gold standard FSC calculation (Figure 27B). All the complex orientations were present in the dataset, with a preference for side views (Figure 27C).

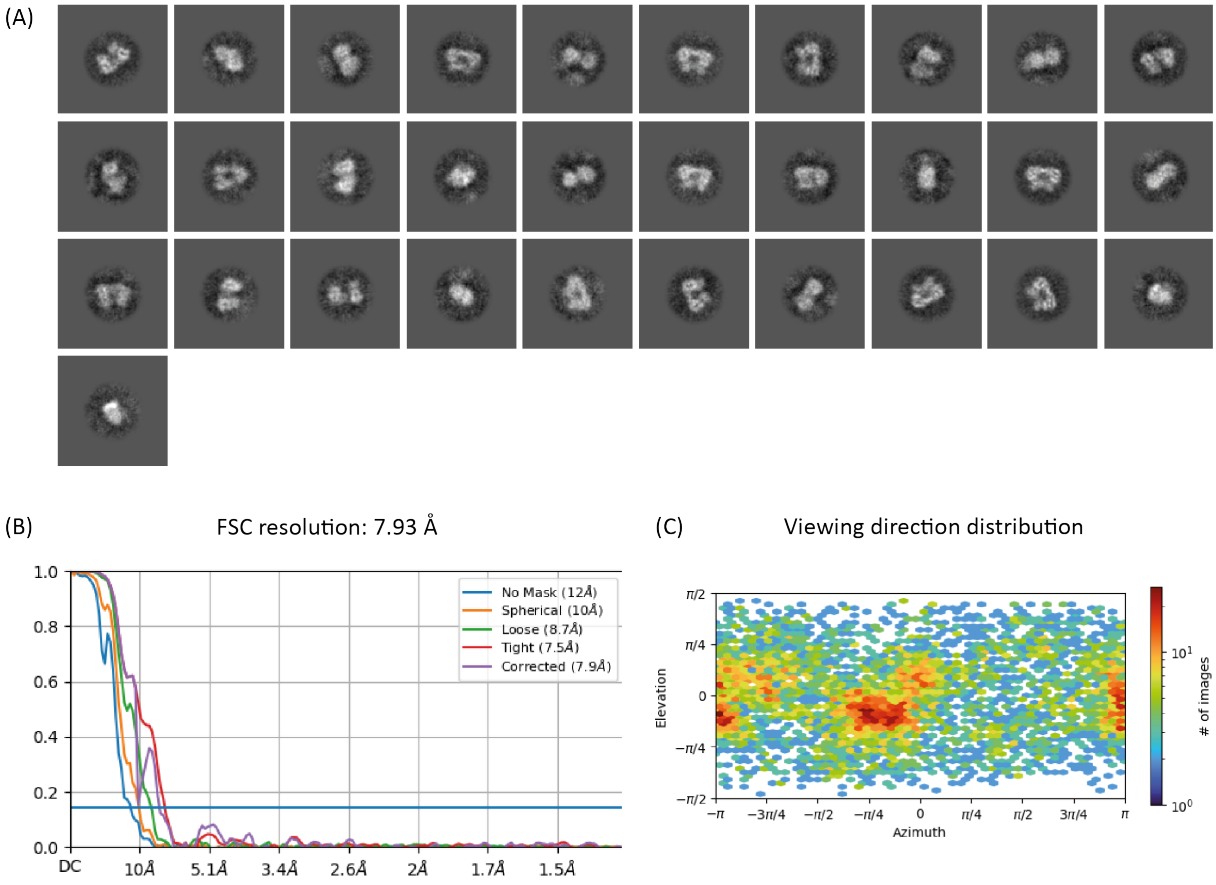


Figure 27: Statistics of the final 3D model of the MKK6DDGRA/p38α T180V complex with ADP + AlF_4^- (A) 2D classes (B) FSC corrected resolution estimation is 7.93 Å based on gold standard calculation (C) Viewing direction distribution.

Architecture of the MKK6DDGRA/p38α T180V complex

The final 3D density map of the MKK6DDGRA/p38α T180V complex revealed the relative positioning of the two kinases based on their "two-lobed" morphology. It was also possible to discriminate between MKK6 and p38α based on their secondary structural features and additional density corresponding to the KIM peptide and the activation loop of p38α. We generated an initial model composed of the crystal structures of p38α with GRA24 KIM peptide (PDB 5eta) and MKK6DD (PDB 3fme), automatically fitted into the density map as rigid bodies using ChimeraX (Pettersen *et al.*, 2021). This map was used for post-processing using DeepEMhancer (Sanchez-Garcia *et al.*, 2020). The hetero-kinase dimer model was finally refined into the resulting sharpened density map with rigid-body real space refinement in Phenix (Headd *et al.*, 2012). The MKK6 and p38α kinase cores were positioned unambiguously into the density map (Figure 28A). Low contour levels indicated the position of secondary structure features, in particular α-helices.

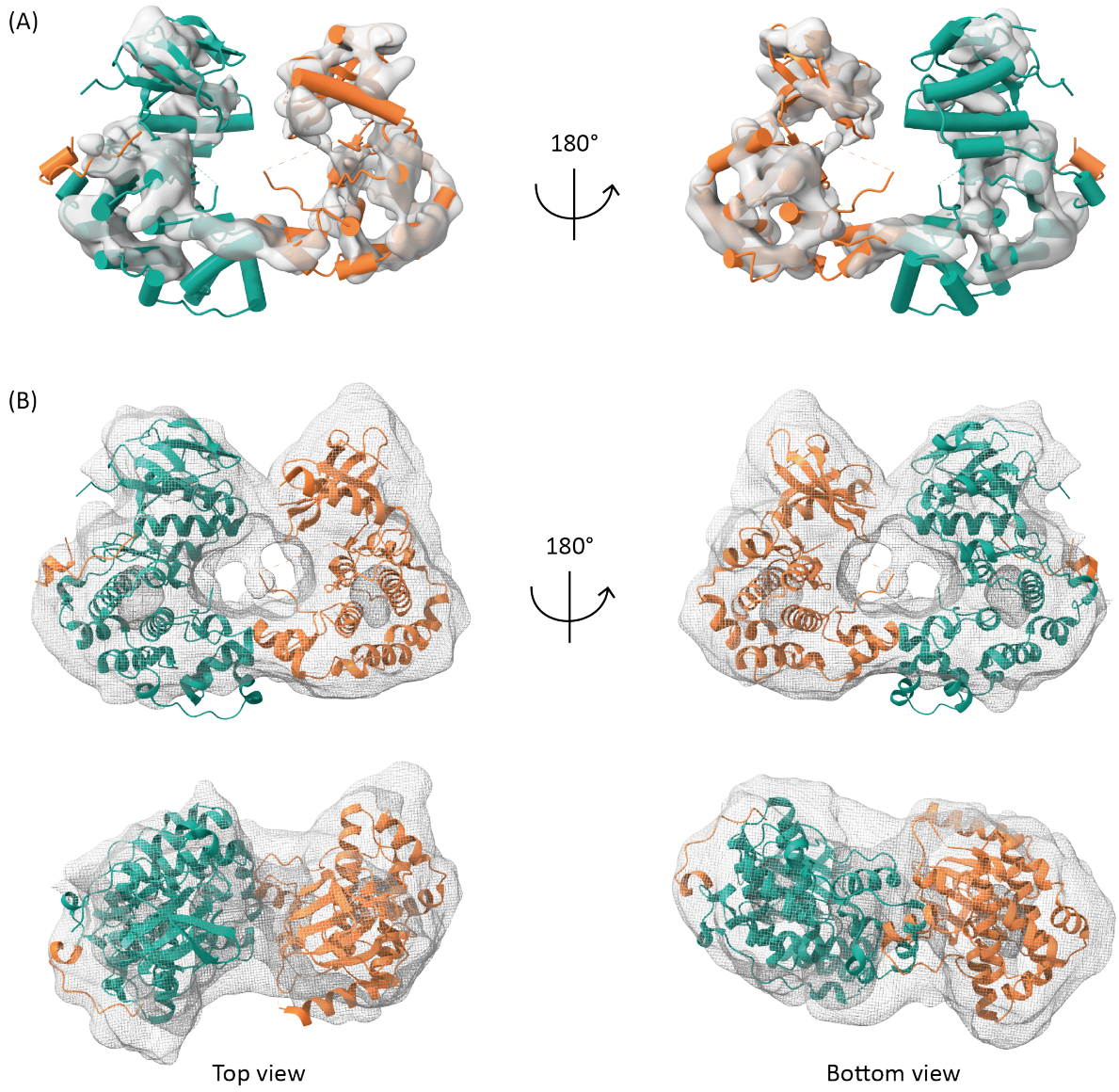


Figure 28: 3D model of the MKK6DDGRA/p38α T180V dimer. (A) The density map sharpened with DeepEMhancer reveals the position of secondary structure features. The crystal structures of p38α (in teal) with GRA24 KIM peptide (in orange) (PDB 5eta) and MKK6DD (in orange) (PDB 3fme) were refined in Phenix into the density map. The density map contour level is set to a low value and the α-helices are represented as cylinders for clarity. (B) The density map contour level is set to enclose 120 nm^3 (Porod volume from SAXS measurement). Two side views, top and bottom views are represented.

The MKK6DDGRA/p38α T180V heterodimer is in face-to-face conformation that would allow the activation loop of p38α to access to the MKK6 active site. The GRA24 KIM peptide fits into the density at the docking site of p38α, the same site that binds MKK6 KIM peptide (Figure 28B).

Novel interaction interface between p38α hydrophobic pocket and MKK6 αG-helix

The two kinases interact extensively between their C-lobes. The αG-helix (residues 262-273) of MKK6 is positioned in a hydrophobic pocket of p38α which has been described as a lipid binding site (Diskin, Engelberg and Livnah, 2008) (Figure 29A). The resolution of the cryoEM density map does not allow us to place residue side chains, their conformations originate from the crystal structures. However, when lowering the contour level to the minimum, we can see that the density around MKK6 Phe264 is persistent (Figure 29A), this residue is equivalent to the F in the FxF motif suggesting that the placement is correct.

The hydrophobic residues Phe264 and Leu267 of MKK6 point into p38α hydrophobic pocket (Figure 29B). The aromatic rings of MKK6 Phe264 and Tyr258 could form a π π-interaction.

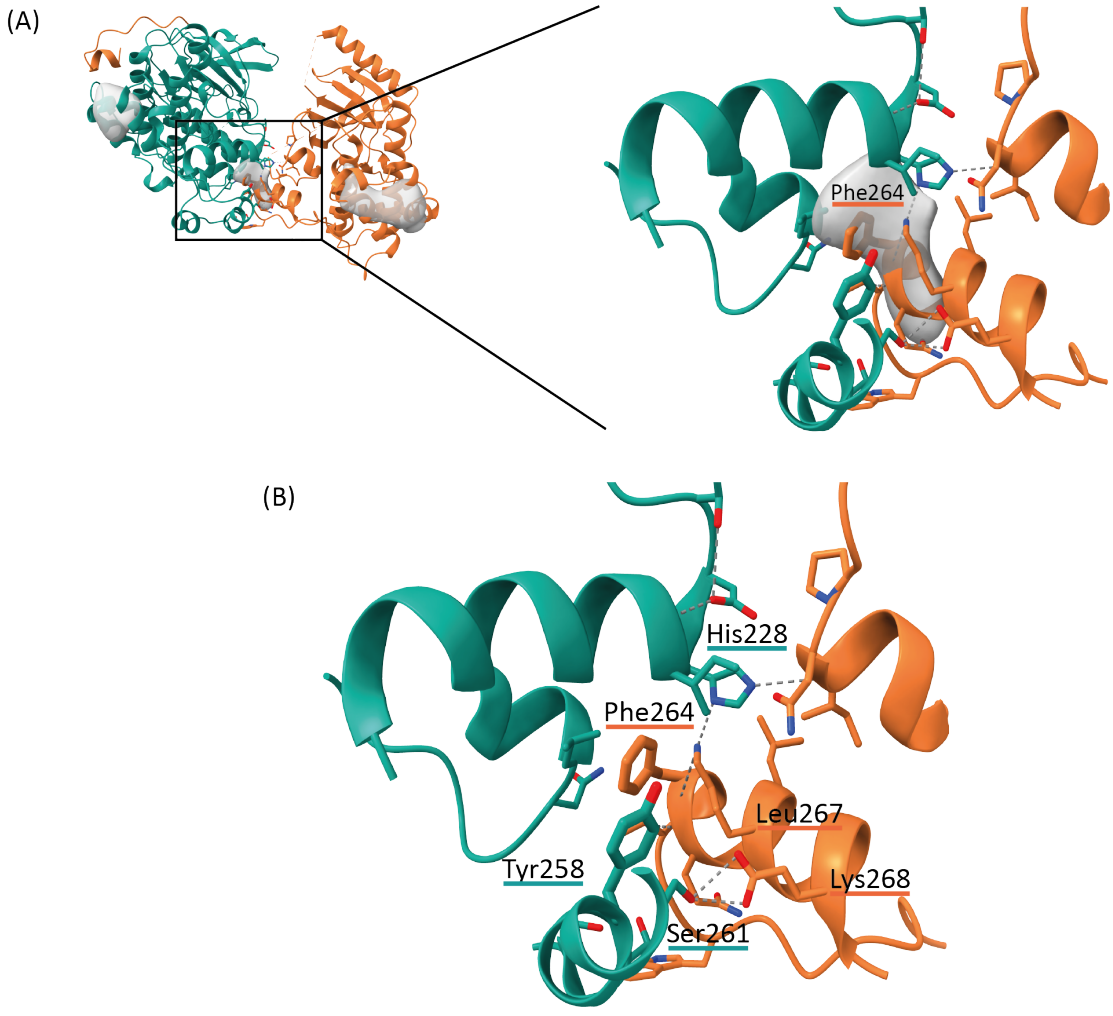


Figure 29: Novel interaction interface between MKK6 αG-helix and p38α hydrophobic pocket. (A) The density map contour level is lowered to a minimum, the density of the interaction interface between the two C-lobes, and more specifically around MKK6 Phe264 is persistent. (B) MKK6 αG-helix (in orange) forms a combination of hydrophobic interactions and hydrogen bonds with the p38α hydrophobic pocket (in teal). The main interactions are represented with thicker sticks.

Several hydrogen bonds could be formed between p38α His228 and Ser261, and MKK6 Leu267 and Lys268 respectively.

This interaction participates in the stability of the conformation of the heterokinase dimer, maintaining the two kinases in the peculiar face-to-face conformation. This novel interaction interface could have a role in the specificity of MKK6 for p38α, in addition to the KIM motif/docking site interaction.

MKK6 αC-helix conformation and N-terminal extension

The αC-helix of MKK6 can adopt different conformations as we can see in the alignment of the three available structures (PDB 3fme, 3vn9 and 3enm) (Figure 30A). In our sample, we expect MKK6DDGRA to adopt an active conformation. The αC-helix should therefore be close to the N-lobe, allowing the formation of the conserved salt bridge between the αC-helix residue Asp99 and the β3-strand residue Lys82. This observation excludes the most distant position of the αC-helix from the N-lobe, as observed in MKK6 structures in PDB 3fme and 3vn9 (Figure 30A in yellow and salmon). The extra empty density around the αC-helix could originate from the N-terminal extension extending towards the p38α docking site.

Some MAP2Ks, including MKK7 and MEK1/2, possess an extra α-helix at their N-terminus, before their KIM domain. This so-called regulatory (or activation) α-helix has a role in the inhibition of auto-phosphorylation in these MAP2Ks. In the available crystal structures of MKK6, the N-terminus is either removed from the recombinant protein construct or not visible in the electron density because it is disordered. So far, there is no experimental evidence for such a domain in MKK6. By superimposing the crystal structures of MKK6DD (PDB 3enm) and MKK7 (PDB 6yg1), we can see that their αC-helices adopt a similar conformation (Figure 30B). The regulatory α-helix of MKK7 fits into the cryoEM density, supporting the hypothesis that the N-terminal extension of its homologue MKK6 could also present an ordered conformation in our sample (Figure 30B).

Based on these observations, we generated a model of MKK6 built from the MKK6DDGRA sequence and the MKK7 crystal structure (pdb 6yg1) using the SWISS-MODEL tool (Waterhouse *et al.*, 2018). This new model (Figure 30C) contains MKK6 residues 31 to 329 (when MKK6 crystal structures contain only residues 47 to 333 (PDB 3fme), 46 to 332 (PDB 3enm) and 44 to 334 (PDB 3vn9)). A small β-sheet (res 38-47) preceded by an α-helix (res 33-35) are modelled (coloured in green in Figure 30C). The recombinant protein construct of MKK6DDGRA comprises the GRA24 KIM domain replacing the MKK6 one. The GRA24 KIM peptide structure available from PDB 5eta completes the model. There are then 15 residues (16-PKEAFEQPQTSSTPP-30) of MKK6 that remain unmodeled. The distance between the Cα of the last residue of GRA24 KIM peptide and of Lys31 (the first modelled residue in our homology model of MKK6) is 45 Å, corresponding to ~3 Å/residue, coherent with the average size of unstructured protein sequences (dashed line in green in Figure 30C). This conformation allows the formation of the conserved salt bridge necessary for the catalytic fold (Figure 30C).

The resolution of the density map does not allow us to build the MKK6 N-terminal extension *de novo*. However, low contour levels reveal the presence of density that could corresponds to the presence of an α-helix (Figure 30D). The small β-sheet and α-helix built in the MKK6 homology model based on MKK7 structure might adopt a different conformation, forming a regulatory α-helix interacting with the αC-helix, stabilizing the MKK6 catalytically active conformation.

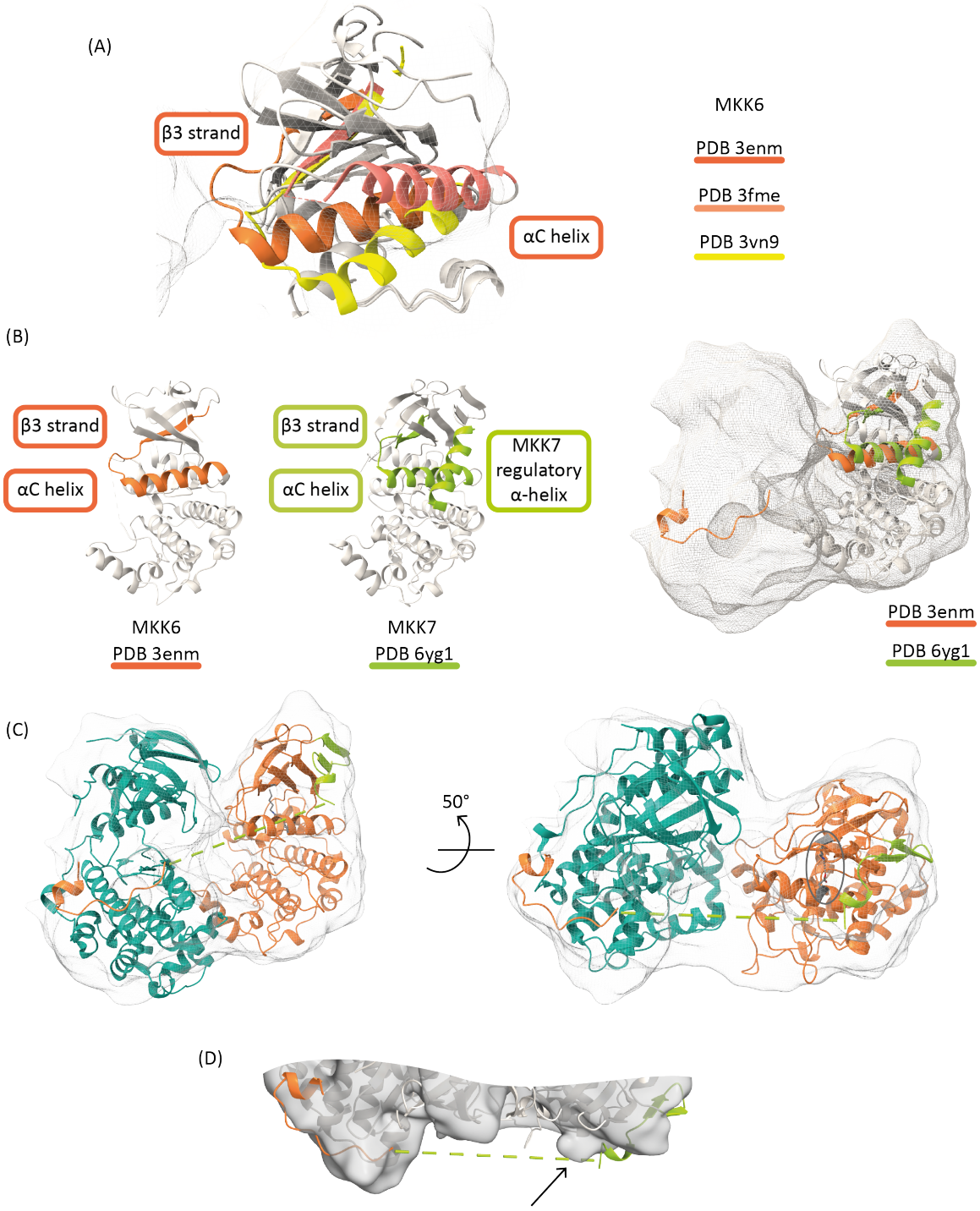


Figure 30: MKK6 N-terminus potential conformations. The density map contour level is set to enclose 120 nm^3 (Porod volume from SAXS measurement). (A) Different possible conformations of the αC-helix of MKK6 observed in crystal structures are represented. Only the αC-helices and β3-strands are represented in colour (PDB 3enm in orange, PDB 3fme in salmon and PDB 3vn9 in yellow). (B) Superimposition between MKK6 (PDB 3enm) and MKK7 structures (PDB 6yg1). The GRA24KIM peptide (PDB 5eta) is represented. Only the αC-helices, the β3-strands, the regulatory α-helix and the KIM peptide are represented in colour (MKK6 and GRA24KIM in orange, MKK7 in green) The αC helices are overlapping, and the MKK7 regulatory helix fits into the density, weighting in favour for a structured N-terminal extension in our sample. (C) Generated model of MKK6 (res 31-338) based on MKK7 structure (PDB 6yg1) (in orange the kinase core and in green the newly modelled N-terminus). The green dashed line represent the 15 residues missing from the model. The conserved salt bridge between the αC-helix and β3-strand is displayed in grey. (D) Close view with low contour level of the cryoEM density map. The arrow points to the density supporting the presence of a N-terminal regulatory α-helix in MKK6.

Focus on the active site and activation loops conformations

In addition to the clear positioning of the kinases relative to one another, the C-lobe interaction and the MKK6 N-terminus, there is also density observed between the two kinases. The p38α activation loop appears to be ordered and placed towards the active site of MKK6, as shown in Figure 31. The crystal structures of p38α with ordered activation loop either non-phosphorylated (Figure 31A) or biphosphorylated (Figure 31B) do not fit well into the cryoEM density map, and predict a more extended novel conformation of the activation loop. The non-phosphorylated p38α structure was refined into the cryoEM density map using real-space refinement in COOT (Emsley and Cowtan, 2004). We can see that the activation loop can extend away from the p38α core toward the MKK6 active site (Figure 31C), placing both the Thr180 and the Tyr182 in the centre of the two kinases. Our sample contains the p38α T180V mutant, limiting the possibilities to only one phosphorylation site in its activation loop: Tyr182.

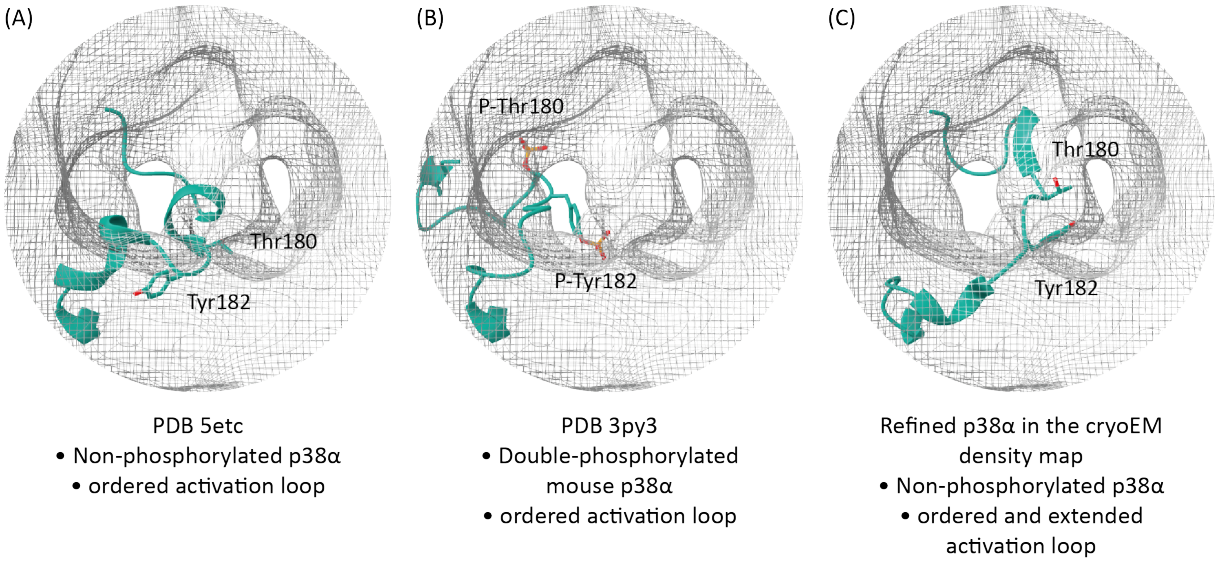


Figure 31: p38α activation loop is ordered and oriented towards the active site of MKK6. The contour level of the cryoEM density map are adjusted to visualise the activation loop (A) Activation loop of the non-phosphorylated p38α (PDB 5etc). (B)) Activation loop of the biphosphorylated mouse p38α (PDB 3py3). (C) Non-phosphorylated p38α activation loop refined into the cryoEM density map. The valine is shown as a threonine to illustrate the WT situation.

The first possible scenario is that the complex is actually in transition state. The Thr182 would be oriented into the active site of MKK6. The sample contains ADP and AlF^{4-}, the ADP could sit into the ATP binding site of MKK6, and the AlF_4^- could behave as the leaving phosphate group of ATP, forming a transition state between the ADP and the receptor hydroxyl group of Tyr182.

Another possible scenario is that the Tyr182 of p38α is already phosphorylated. As we will discuss in Chapter II, MKK6 can use ADP to phosphorylate p38α, with less efficiency than with ATP. We chose to use ADP and AlF_4^- in our sample with the aim to form a transition state analogue complex, based on the SAXS analysis that showed that together with the p38α T180V mutant this assembly was in a more compact than other p38α construct/nucleotide combination, and supposing that the phosphorylation resulting from ADP use for

phosphoryaltion would be minor while keeping the sample on ice at all time. However, for future experiments, we will use AMP-CP, a non-cleavable ADP analogue.

The activation loop of MKK6 is highly mobile and can adopt a wide variety of conformations, as we can see in the three available crystal structures of the MAP2K (Figure 11, p. 48). None of these conformations fitted well into the cryoEM density map. The three α-helices AH1-3 observed in PDB 3vn9 would block the access of the p38α activation loop to the active site of MKK6 . The homology structure of MKK6 obtained by SWISS-MODEL based on the crystal structure of active MKK7 (PDB 6yg1) seems to be closer to the conformation observed of MKK6DDGRA.

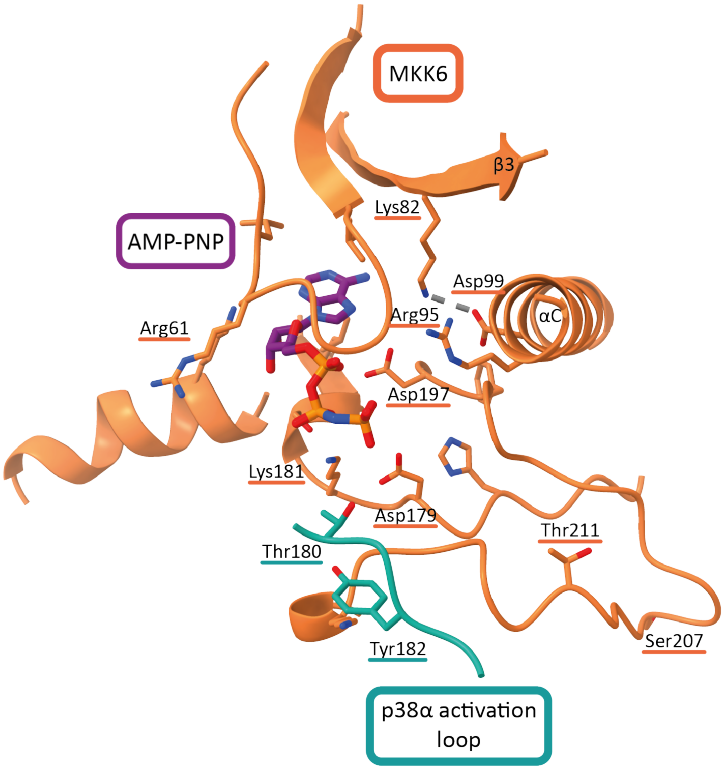


Figure 32: MKK6 active site. The MKK6 homology model based on the crystal structure of active MKK7 (PDB 6yg1) is represented as cartoon and sticks in orange. The p38α activation loop from the refined model is represented in teal.

Focusing on the MKK6 catalytic residues, we can see that the conserved Lys181 can coordinate the γ-phosphate. There are a number of positively charged residues, including Arg61 and Arg95 that surround the MKK6 active site and that could help to stabilize the local negative charges. The MKK6 Asp179 from the HRD motif is the catalytic aspartate that make the hydroxyl group of the phosphor-acceptor a nucleophile. In our model, the p38α activation loop and the Tyr182 are too far away from the nucleotide and MKK6 active site to be in phosphorylation transition state. However, given the resolution of the model, it does not exclude that the actual conformation could be in transition state.

Overall, we can see that the interaction between the two kinases induces multiple conformational changes on both of them, displaying a combination of different conformations observed in individual crystal structures, with some hints of novel conformations and interacting interfaces. Two scenarios are conceivable regarding the catalytic stage of the MKK6/p38α dimer. The dimer could be in transition state conformation, with p38α Tyr182 placed in MKK6 active site, the ADP would sit into the MKK6 ATP binding site and AlF_4^- would mimic the γ-phosphate leaving group, between the two. The catalytic residues are in the right position for phosphorylation and the distances and orientation of the two kinases are compatible with this hypothesis. Another possibility would be the p38α Tyr182 is already phosphorylated, and the dimer is in a post-catalytic state.

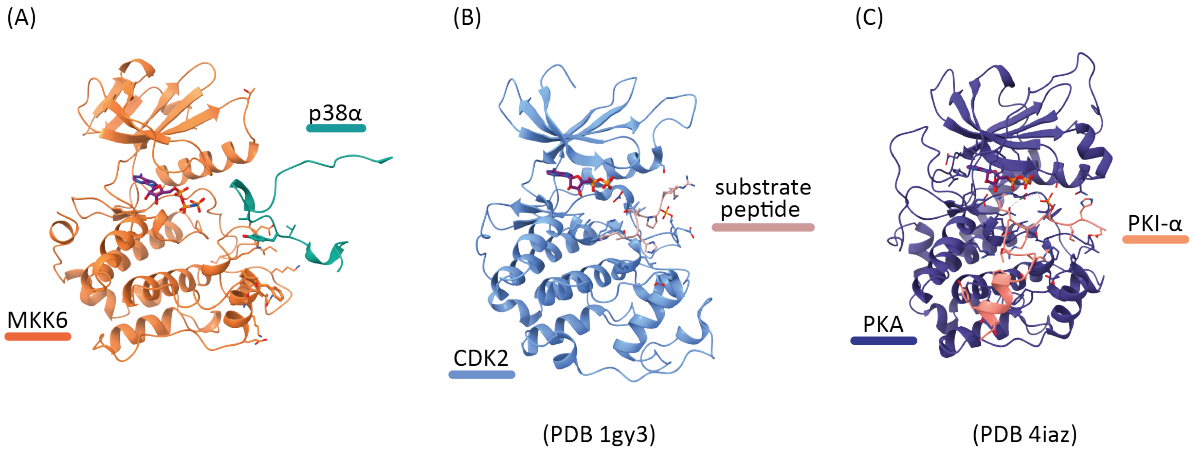


Figure 33: Comparison of kinase-substrate binding to the kinase active site. (A) CryoEM structure of MKK6DDGRA (in orange) and p38α T180V activation loop (in teal). (B) crystal structure of CDK2 and an optimal peptide substrate (PDB 1gy3). (C) Crystal structure of PKA and PKI-α peptide (PDB 4iaz).

The surprising feature of this model is that there is very little interaction between the p38α activation loop and the MKK6 active site (Figure 33A). Structures of other protein kinase-substrate complexes usually display interactions between residues adjacent to the phospho-acceptor and the kinase active site (examples in Figure 33B and C). In the case of the MKK6-p38α complex, most of the interactions between the two kinases are distal from the active site. In addition to the well-studied KIM/p38α docking site interaction, we presented a novel interaction interface between the MKK6 αG-helix and the p38α hydrophobic pocket. There is possibly some interaction occurring between the two N-lobes as well. In this conformation, the p38α activation loop is free to move between the two kinases and to present either of its phosphorylation sites. Thr180 or Tyr182, to the MKK6 active site.

A few asymmetric kinase homodimers have been published, where one kinase acts as the "enzyme" and the other one as the "substrate" in the transphosphorylation reaction (Beenstock, Mooshayef and Engelberg, 2016). Three of them, the PAK1 homodimer (PDB 3q4z) (Wang, Wu and Wang, 2011), the IRAK 4 homodimer (PDB 4u97) (Ferrao *et al.*, 2014) and the Rad53 homodimer (PDB 4pds) (Wybenga-Groot *et al.*, 2014) are also in a face-to-face conformation, in which the activation loop of the "substrate" kinase is placed into the active site of the "enzyme" kinase (Figure 34A, B and C respectively). However, they differ from the MKK6-p38α heterodimer in the relative positioning of the two kinases:

- in the PAK1 homodimer, the main interactions occur between the “substrate” kinase activation loop and the “enzyme” kinase active site (Figure 34A).
- The interactions between the two IRAK4 monomers are concentrated between their C-lobes (Figure 34B).
- The Rad53 dimer is in a head-to-tail conformation (Figure 34C).

The architecture of the complex of BRAF–MEK1–14-3-3 was also derived from a 5 Å cryoEM map in which BRAF is in an active conformation to phosphorylate its substrate MEK1 and is in a similar relative orientation as our model of the MKK6-p38α dimer (PDB 6nyb) (Park *et al.*, 2019) (Figure 34D). In all the examples cited above, the αG helix of the “enzyme” kinase plays an important role in the interaction. Besides the similarities with other face-to-face kinase dimers, the MKK6-p38α heterodimer structure displays specific features that help the understanding of the specificity of the interaction between a MAP2K and its substrate MAPK.

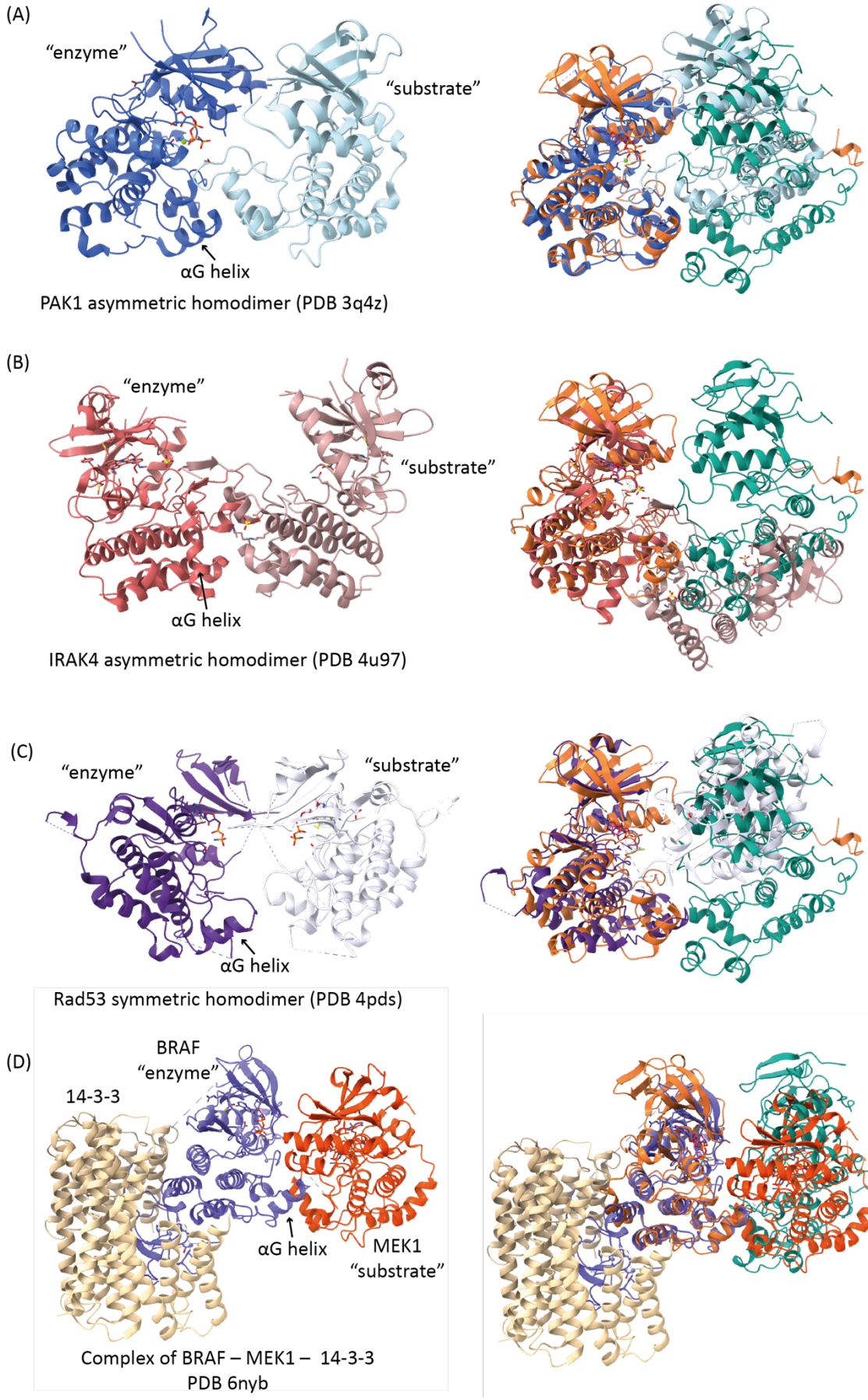


Figure 34: Comparison of the MKK6-p38α dimer with other kinase dimers. For the asymmetric homodimers (A, B and C), the "enzyme" kinase is represented in dark color and the "substrate" kinase in light color. The αG helix of the "enzyme" kinases are indicated by an arrow. The right panel shows the overlap of the published kinase dimer structures with our structure of the MKK6 (in orange) – p38α (in cyan) dimer.

Discussion and perspectives

The strategy to stabilize a transient interaction is crucial

The structural study of the interaction between p38α and its upstream activating MAP2K MKK6 is difficult due to the transient nature of the interaction. We successfully managed to express and purify a stable complex by engineering an MKK6 chimera, in which we replaced its N-terminal KIM motif by the sequence of *T. gondii* GRA24 KIM1 that has a 100-fold higher affinity for p38α docking site. The success of this motif swap in our project could inspire the stabilization of other stable recombinant protein complexes.

Careful optimisation of expression and purification was necessary to obtain a pure and stable sample, containing a 1:1 active complex, in a homogeneous phosphorylation state. Maintaining the p38α activation loop unphosphorylated was crucial for the study of its activation by MKK6.

The benefits of an integrative structural biology approach

Integrative structural biology approaches are establishing themselves in the field, and were particularly crucial for the success of this project. Contrary to crystallography, it is difficult to screen different conditions and sample compositions with electron microscopy. We exploited SAXS as a tool for screening different complex assemblies to evaluate their conformation and flexibility prior to cryoEM studies, combined with biochemical, biophysical and MS methods, for sample optimization and quality control. Besides, the comparison of the different 3D models successively obtained by SAXS, negative stain EM and cryoEM shows that low resolution information obtained by the first two techniques can still lead to relevant hypothesis (Figure 35).

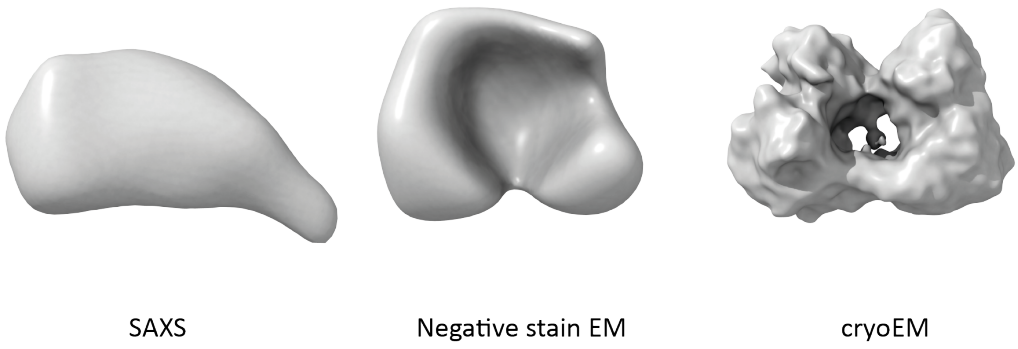


Figure 35: Comparison of the density maps of the MKK6DDGRA/p38α complex successively obtained by SAXS, negative stain electron microscopy and cryoEM. The density map contour levels are set to enclose 120 nm^3 (Porod volume from SAXS measurement).

The cryoEM resolution revolution has opened new possibilities for the study of small objects

Despite the resolution revolution of cryoEM and the tremendous progress for the study of small objects, the size of the MKK6/p38α complex still represents a challenge (Kuhlbrandt, 2014; Herzik, Wu and Lander, 2018). The emergence and development of novel data processing methods is crucial for pushing the boundaries.

The implementation of on-the-fly processing has been particularly helpful for our project. The evaluation of quality of a sample embedded in ice is not easy by eye in the case of small objects. The on-the-fly motion-correction and movie alignment, CTF estimation and particle picking (with WARP software in our case) helped greatly to select sample/grid/ice thickness appropriated for data collection and processing.

We initially processed our data with the Relion v3 software package (for REgularised LIkelihood OptimisatioN) (Zivanov *et al.*, 2018). Given the challenges of our project, we decided to also try other software and chose the CryoSPARC software package (Cryo-EM Single Particle Ab-Initio Reconstruction and Classification) (Punjani *et al.*, 2017). The major difference between Relion and CryoSPARC processing in our project was the *ab initio* modelling. The two software packages use different approaches, and only CryoSPARC gave satisfying results to generate a convincing 3D *ab initio* model of the kinases hetero-dimer.

One of the recent breakthroughs in cryoEM data processing is the implementation of new particle picking methods based on neural networks, such as CrYOLO (Wagner *et al.*, 2019), WARP (GPLv3 license) (Tegunov and Cramer, 2019) and Topaz (GPLv3 license) (Bepler, Noble and Berger, 2019). Particle picking is often a limiting step in case of small objects. In our case, the small size of the complex combined with low signal-to-noise ratio, the potential presence of a combination of assemblies including kinases monomers and dissociated complex, and ice contamination of the grid made particle picking even more challenging. The success of our strategy relied on an iteration of reference-free picking, 2D classification and "*in silico* purification" until a few classes corresponding to the particles of interest (the kinase hetero-dimer assembly) were obtained that were then used to train the Topaz particle picker to work on the full dataset.

Our latest Krios dataset is not optimal to achieve a high-resolution reconstruction. Based on the CryoSPARC ice thickness estimation, we know that the ice was too thick in the dataset collected on the Krios microscope.

Despite these limitations, the 2D classes are well defined (effective classes assigned score between 1.4 and 2.0 in cryoSPARC), and the 3D map already gives insights into the MKK6/p38α interaction. For future data collection, we will focus on better stabilizing the complex in the transition state by replacing ADP by its non-cleavable analogue AMP-CP. Meanwhile, we will also optimise Topaz picking with the use of Topaz Cross Validation.

New insights into the interaction between a MAPK and its upstream MAP2K

The 3D model of the MKK6DDGRA24/p38α T180V complex with ADP + AlF_4^-, despite its low resolution, allows the formulation of novel hypotheses on the interaction between the MAPK and it upstream MAP2K and the activation process.

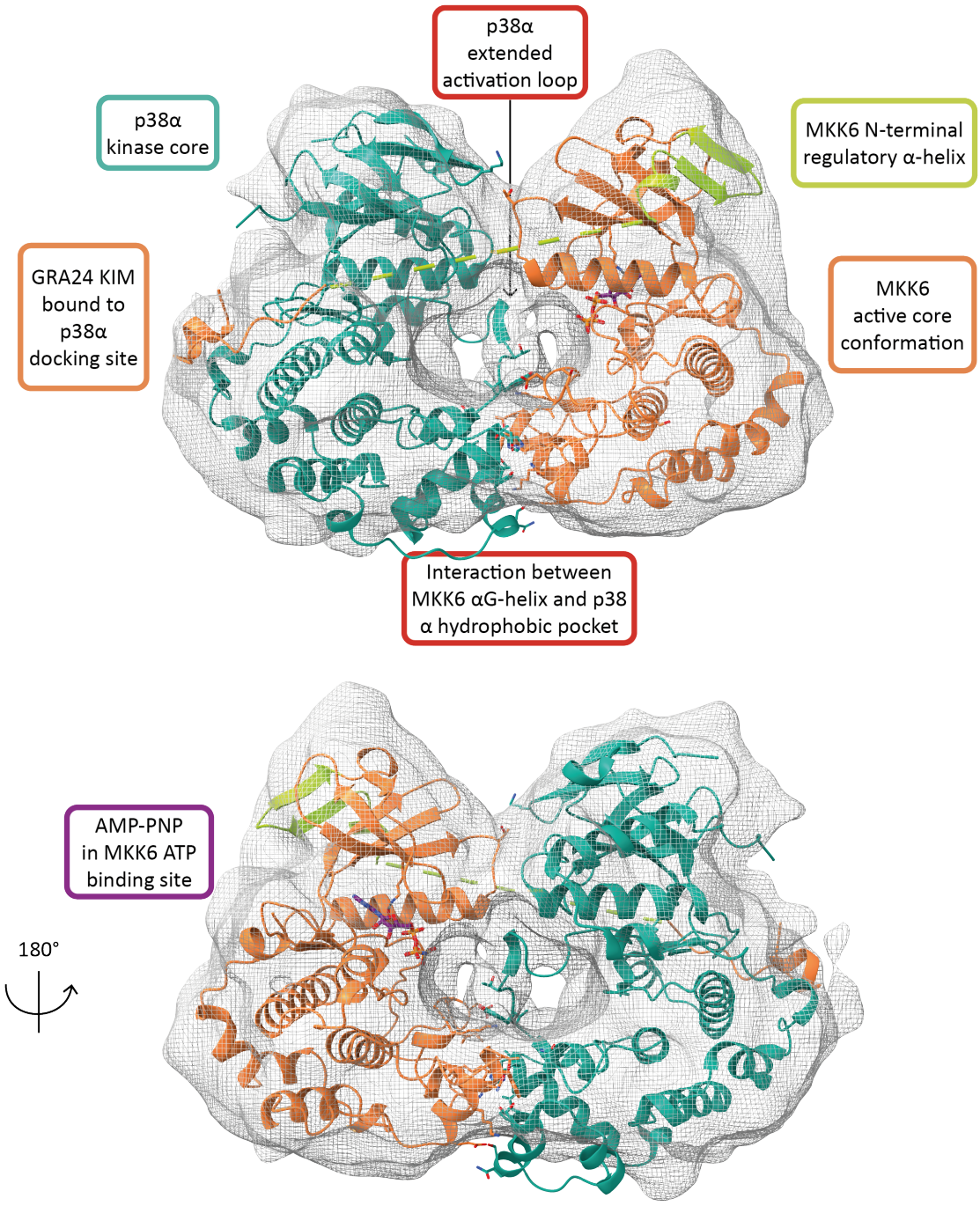


Figure 36: Model of the MKK6DDGRA/p38α T180V complex in the cryoEM density map at 7.9 Å.

Combining the fitting of different conformations of the two kinases observed in crystal structures, the generation of homology models using SwissModel, and rigid-body real space refinement in Coot and Phenix, we generated the first model of a MAP2K/MAPK heterodimer in a putative transition state conformation. We present here the model of MKK6DDGRA interacting with p38α T180V in which (Figure 36):

- MKK6DD kinase core is in an active conformation and the salt bridge between β3-stand and αC-helix is formed.
- An ADP molecule should sit into the ATP binding site. The transition state would be mimicked by the AlF_4^- ion acting as the phosphate leaving group. In our model we represented an AMP-PNP molecule.
- The MKK6 N-terminal extension present an ordered conformation that could be similar to the regulatory α-helix observed in MKK7 and MEK1/2 structures. Based on the homology model built from active MKK7 conformation (PDB 6yg1), it might form a small β-sheet preceded by an α-helix before joining the C-terminus of GRA24 KIM peptide with 15 unmodeled residues. The cryoEM density map points to the presence of an α-helix that would interact with the αC-helix stabilizing the catalytically active conformation. The GRA24 KIM peptide binds to p38α docking site as in PDB 5eta.
- The p38α activation loop is ordered and extend toward the MKK6 active site.
- The kinase core of p38α is in an intermediate conformation similar to the one adopted when binding to KIM peptides (as in PDB 5eta) (Pellegrini, Palencia, *et al.*, 2017).
- The αG-helix of MKK6 interacts with the p38α hydrophobic pocket.
- There are possible interactions between the two N-lobes.

Novel interaction interface between p38α hydrophobic pocket and MKK6 αG-helix

We propose the description of a novel interaction interface between the p38α hydrophobic pocket and the αG-helix of MKK6.

The hydrophobic pocket is located in a flexible region and only formed in certain conformations of the MAPK. It is delimited by the MAPK insert, a region only conserved amongst MAPK and CDK kinases. The MAPK insert is constituted of two α-helices connected by loops to the bottom of the C-terminal lobe of the kinases, conferring some flexibility and a possible clamp movement. It has been described as the lipid binding site in p38α (Diskin, Engelberg and Livnah, 2008), and also corresponds to the DEF binding site described in ERK1/2 to bind FxF(P) motifs present in some interacting partners (Jacobs *et al.*, 1999). The DEF site presence was also suggested in p38α/β, and JNK proteins (Sheridan *et al.*, 2008; Liu *et al.*, 2016). The p38α hydrophobic pocket is flexible, and its opening depends on the kinase conformation and interactions (Figure 37A).

The residues involved in the interaction between MKK6 αG-helix and p38α hydrophobic pocket are highly conserved amongst MAP2Ks and MAPKs respectively (Figure 37B).

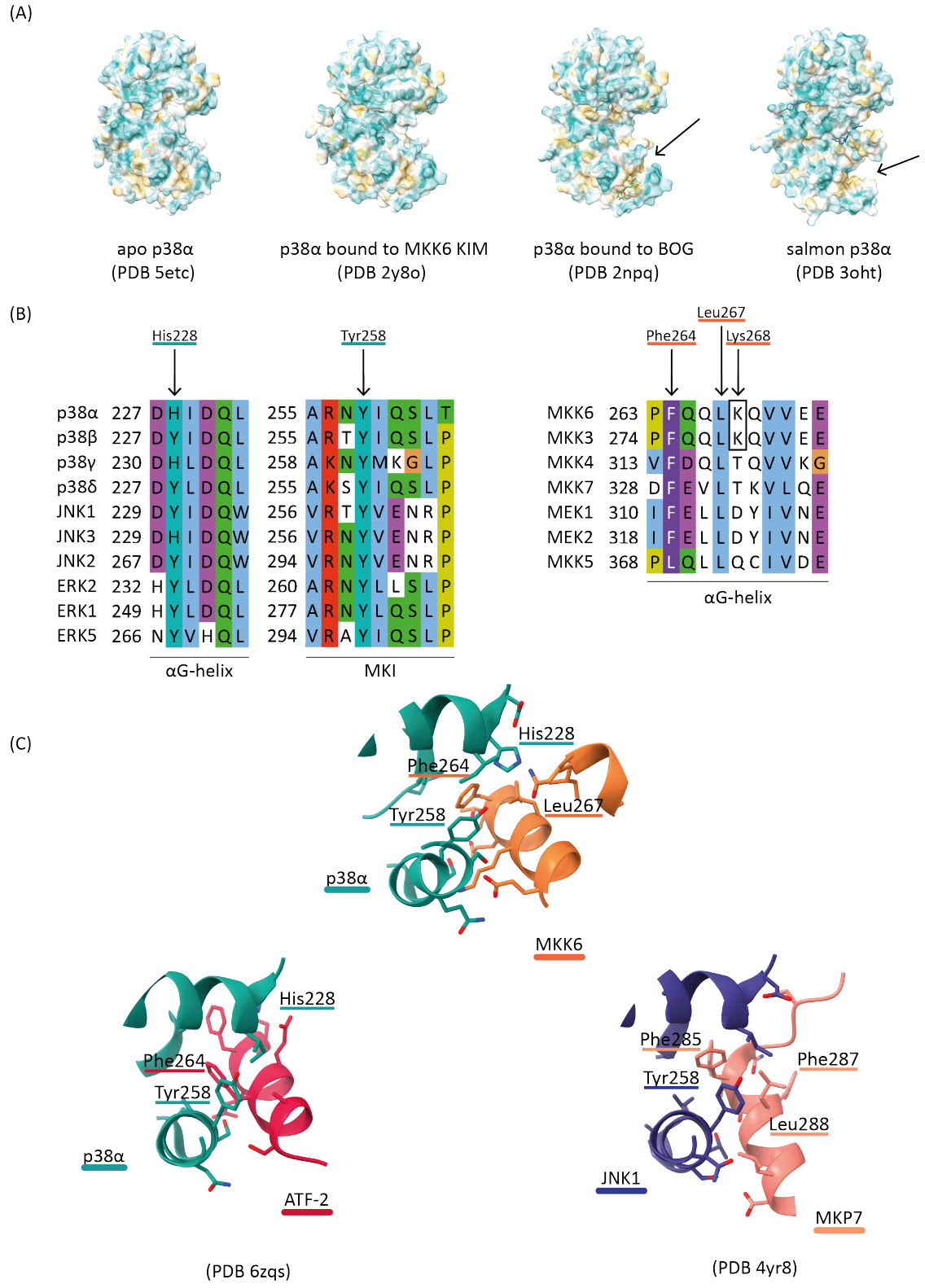


Figure 37: Focus on the interaction between MKK6 αG-helix and p38α hydrophobic pocket (A) Cartoon and surface representation of four p38α structures (apo p38α (PDB 5etc), p38α bound to MKK6 KIM (pdb 2y8o), p38α bound to BOG (PDB 2npq) and salmon p38α (PDB 3oht)). The surface is coloured by hydrophobicity (cyan for most hydrophilic to yellow for most hydrophobic). An arrow points to the open hydrophobic pocket. (B) Sequences of human MAPKs and MAP2Ks aligned with Clustal Ω. The residues involved in the interaction are highly conserved. (C) Comparison between the cryoEM structure of MKK6DDGRA/p38α T180V and the crystal structures of p38α interacting with a peptide from its substrate transcription factor ATF2 (PDB 6zqs), and JNK1 interacting with MKP7 (PDB 4yr8). The interaction between the MAPK Tyr and a substrate Phe is conserved.

Two structures display a similar binding mode between MAPKs and partners. The closest one to our structure is the structure of biphosphorylated p38α bound to an ATF2 transcription factor peptide (Kirsch *et al.*, 2020) (PDB 6zqs, Figure 37C). The binding involves the same p38α Tyr258 and a phenylalanine in an α-helix of ATF2. Hydrophobic interactions and hydrogen bonds stabilize the interaction, involving the p38α His228 as well.

A very similar interaction is also present between JNK1 and the phosphatase MKP7 (Liu *et al.*, 2016) (PDB 4yr8, Figure 37C). It is the only crystal structure of a MAPK kinase bound to an FxF peptide. The first phenylalanine of this motif in MKP7 (Phe285) interacts with the conserved tyrosine in the MKI (Tyr259 in JNK1, Tyr258 in p38α). The second phenylalanine from the motif is not directly involved in the binding. However, as in MKK6, a leucine positioned 3 residues upstream of Phe285 is also involved in the interaction (Leu288 in MKP7, Leu267 in MKK6).

The definition of the FxF(P) is loose, the structure of JNK1 bound to MKP7 is the only structural evidence of such interaction, and the second phenylalanine of the motif does not participate directly in the interaction (PDB 4yr8, Figure 37C). We suggest the existence of another type of motif, F-x-x-φ, adding the requirement for it to be located inside an α-helix.

In the context of the MKK6/p38α interaction, the binding of the MKK6 αG-helix into the p38α hydrophobic pocket could help stabilise the conformation of the two kinases facing each other and facilitate phosphorylation. It could also protect the activation loops from phosphatases (both the phosphorylated MKK6 loop, to maintain it in the active state, and the p38α loop until dual phosphorylation is achieved). It could participate in the specificity determination of the pathway. Interestingly, the αG-helix is preceded by a proline-rich region insertion found only in MEK1/2.

Based on all these hypotheses, we can propose a mechanism where the interaction of the KIM motif of MKK6 to p38α docking site induces conformational changes on both kinases. MKK6 adopts a catalytically active conformation. The hydrophobic pocket is then formed in p38α and accommodates the hydrophobic residues of the αG-helix of MKK6. This additional interaction interface stabilises the face-to-face conformation of the two kinases, allowing the activation loop of p38α to extend into MKK6 active site. The dual phosphorylation of p38α induces conformational changes in p38α toward an active conformation. These conformational changes could affect the hydrophobic pocket and stimulate its closing, participating in the release of MKK6 and the dissociation of the complex.

So far, the interaction to the hydrophobic pocket of MAPKs have been described (or suggested by the presence of FxF(P) motifs) for:

- MKP phosphatases (including MKP3, 5 and 7), in which this binding is believed to help with the proper orientation of the phosphatase catalytic residues with the TxY motif in the MAPK activation loop, and could participate to the specificity between MKPs and their MAPK substrates (Zhang, Wu and Wang, 2011).
- MAPK substrates, and more specifically transcription factors, including ATF2, Elk-1, c-Fos, c-Myc, and JunD (Sheridan *et al.*, 2008).
- scaffolding proteins, including KSR1 (Alexa, Varga and Reményi, 2010).

We suggest that the hydrophobic pocket of MAPKs could have a larger role, including the interaction with their upstream activated MAP2Ks. It could have a role in the specificity and strength of the interactions. For instance, the FL MKK6 affinity for p38α is higher than the

affinity of the sole MKK6 KIM peptide for the MAPK (Kumar, Page and Peti, 2021); and the KIM motif definition is not sufficient to explain the discrimination of binding partners between the different MAPKs. It would be interesting to explore if this interaction is shared in other MAP2K/MAPK pairs.

We will first pursue the characterization of the specific interaction between the p38α hydrophobic pocket and MKK6 αG-helix by different approaches. Using recombinant p38α and MKK6 αG-helix peptide, we will measure the affinity of the interaction by ITC, and attempt co-crystallization. We will also prove the importance of such interaction, by applying point mutations on the identified key residues on both kinases and evaluating their effect *in vitro* on the recombinant complex, and *in cellulo* using a reporter cell assay. Small hydrophobic molecules have been developed to bind to the p38α hydrophobic pocket (Tzarum *et al.*, 2012; Bührmann, Hardick, *et al.*, 2017; Bührmann, Wiedemann, *et al.*, 2017). It would be interesting to evaluate their effect on the activity of MKK6 towards p38α.

Transition state or post-catalytic state?

The activation loop of p38α appears to be ordered and oriented towards the MKK6 active site. We consider two possible scenarios regarding the state of the complex.

The first is the presence of the desired transition state complex in which AlF_4^- mimics the phosphate group leaving the nucleotide accommodated in the MKK6 ATP binding site to phosphorylate the p38α Tyr182 residue. Complementary studies could help validate this hypothesis. Analysing the sample with fluorine NMR would allow us to validate the formation of AlF_4^- in our complex as it has been demonstrated in other complexes (Jin, Molt and Blackburn, 2017).

The second scenario corresponds to a “post-catalytic” state in which the p38α Tyr182 is already phosphorylated and the p38α/MKK6 dimer is still stable. This would imply that MKK6 uses ADP to phosphorylate p38α and could bring crucial structural insights to this novel mechanism that is developed in Chapter II. The resulting AMP molecule might still be bound in the MKK6 binding site.

The cryoEM density map obtained so far cannot differentiate between one or the other scenario. Another possibility would be that several catalytic states are present in the sample, and higher resolution might differentiate them into several classes. We will proceed with new cryoEM grid preparation of the MKK6DDGRA/p38α T180V sample with AlF_4^- and AMP-CP, a non-cleavable ADP analogue, to avoid any phosphorylation and favour the transition state complex. We will collect a new dataset on a Titan Krios electron microscope with the aim of improving the final resolution of the 3D model.

The two kinases both undergo important conformational changes upon interaction. A higher resolution cryoEM map will bring crucial information on the activation of a MAPK by its upstream MAP2K. It would give insights on the phosphorylation mechanism if the transition state conformation is trapped. Even if we reached the limitation of this dataset, our model of the MKK6DDGRA/p38α T180V complex with ADP + AlF_4^- is the first model of a MAPK interacting with its activating MAP2K. It has revealed new interaction interfaces between the two kinases and paves the way for hypothesis on MAP kinase activation.

Chapter II: Novel mechanism of activation of p38α by MKK6 using ADP as a source of phosphate

Introduction

The interaction and phosphorylation of MAPKs by their upstream MAP2Ks has been extensively studied. However, molecular details of the mechanism remain unknown. Along with the structural characterization of the interaction of MKK6 and p38α presented in Chapter I, we observed a novel mechanism of phosphorylation that occurs in the presence of ADP. This was initially observed by Erika Pellegrini in her PhD thesis (Pellegrini, 2013), but definitive experimental proof was not made at the time.

The canonical activation of p38α by MKK6 results from the docking interaction of active MKK6 to the p38α docking site, triggering an allosteric conformational change, followed by the dual phosphorylation of the TGY motif of the p38α activation loop by MKK6. It can occurs through two types of mechanism (Patwardhan and Miller, 2007):

- a processive mechanism, in which MKK6 binds to the p38α and phosphorylates both phospho-sites before dissociating
- a distributive mechanism, in which each phosphorylation requires a separate binding event between MKK6 and the p38α.

In both scenarios, the phosphorylation implies that the binding of ATP, followed by the nucleophilic attack of the ATP γ-phosphate, and the release of the resulting ADP molecule occur sequentially, two times. The main difference resides in the release of the monophosphorylated p38α and the subsequent rebinding event.

These two possibilities involve different consequences regarding the kinetic of activation of a MAPK. A processive mechanism follows classical Michaelis–Menten kinetics. A distributive mechanism introduces a dependence on enzyme (MAP2K) concentration, where small changes in enzyme (MAP2K) proportion result in large changes in phosphorylated product (MAPK) proportion. Conversely, an increase in substrate (MAPK) concentration would result in competition between non-phosphorylated and monophosphorylated MAPK, decreasing the accumulation of biphosphorylated MAPK, and therefore providing protection against MAPK over-expression (Witzel and Blüthgen, 2018). Distributive phosphorylation would allow proofreading to increase the specificity of activation of signalling pathways (Ouldridge and Rein Ten Wolde, 2014).

An additional level of complexity is added by the deactivation by phosphatases. Phosphatases can be Ser/Thr specific or Tyr specific, and therefore participate in a distributive dephosphorylation; but there are also some dual-specificity phosphatases (DUSPs) that can act in either a processive or distributive manner (Ha *et al.*, 2019).

Another consideration is the effect of protein localization on the process. Activated MAPKs usually translocate to the cell nucleus to phosphorylate a number of transcription factors, affecting the local concentration of biphosphorylated MAPK product.

The most probable scenario is that MAPK dual-phosphorylation is a distributive mechanism *in vitro*, but tends to convert to pseudo-processive kinetics in the crowded cellular environment, since rebinding after the first phosphorylation event is enhanced by slow diffusion in these conditions, as demonstrated for ERK (Aoki *et al.*, 2011). Molecular crowding is a fundamental parameter in cells, and attempts have been made to reproduce it *in vitro*, e.g. using PEG (Aoki *et al.*, 2013). Humphreys and co-workers proposed that p38α phosphorylation by MKK6 is distributive *in vitro* by fitting MS data to kinetic models (Humphreys *et al.*, 2013; Piala, Humphreys and Goldsmith, 2014). More recently, Wang and co-workers published an extensive study allowing most kinetic constants to be measured (Wang *et al.*, 2019). They showed that the dual-phosphorylation of p38α by MKK6 is a partially processive mechanism, in which monophosphorylated intermediates can either dissociate from MKK6 or proceed to the second phosphorylation step. One of the advantages of combining processive and distributive mechanisms of the dual-phosphorylation by MAP2K and dephosphorylation by phosphatases is that the cell can generate a multitude of response types, including switch-like/ultrasensitive behaviour, oscillations, biphasic responses, or steady state (Sun *et al.*, 2014; Suwanmajo and Krishnan, 2015).

Along with these observations arises the question of a preferential order of phosphorylation between the two phospho-sites Thr180 and Tyr182. It could theoretically be either random, or sequential, following a precise order. In the latter case, the activation of the MAPK would be associated with the second phosphorylation event. Some MS studies showed that p38α is phosphorylated by MKK6 first on Tyr182 and then on Thr180 (Humphreys *et al.*, 2013; Piala, Humphreys and Goldsmith, 2014). The same sequence of events was also observed in ERK2 (Haystead *et al.*, 1992). In their study, Wang and co-workers showed that both p38α monophosphorylated forms can be produced by MKK6 catalysis, but the phosphorylation of Tyr182 is four-fold faster than Thr180 phosphorylation during the first phosphorylation step (Wang *et al.*, 2019). It seems to be compensated by an accelerated second phosphorylation step for the $T^{P}GY$ form, partially explaining the accumulation of the TGY^{P} intermediate and not the $T^{P}GF$ one.

The role of monophosphorylated forms of p38α seems to depend on the context, and the relative contribution of each phospho-site to MAPK activity is still debated. Some studies show that tyrosine phosphorylation is required for activity, others claimed that it is “only” a stepping stone in the MAPK activation (Zhou *et al.*, 2002; Bell and Engelberg, 2003). Monophosphorylated p38α at the threonine site displays some activity, reduced compared to the biphosphorylated form, but still relevant in some biological contexts (Askari *et al.*, 2009; Mittelstadt *et al.*, 2009). On the contrary, monophosphorylated p38α at the tyrosine site seems to be inactive.

Vázquez and co-workers studied Sty1, the *Schizosaccharomyces pombe* orthologue of p38/JNK MAPKs, and Pmk1, the ERK orthologue; they observed that these MAPK T180A activation loop mutants (human p38α numbering) were strongly phosphorylated at the tyrosine site in cells, but also that the binding to their respective MAP2Ks Wis1 and Pek1 was increased (Vázquez *et al.*, 2015). They imagined that the MAPK, once phosphorylated at the Tyr, would be trapped with its MAP2K until dual phosphorylation is achieved. This affinity increase could compete

with MAPK substrates and explain the low activity observed for MAPK monophosphorylated on the Tyr.

The current experimental data obtainable by phospho-proteomics are still insufficient to properly model *in vivo* phosphorylation. This mechanism involves many parameters, including rapidly evolving temporal and spatial distribution of kinase enzymes, protein substrates and nucleotides, negative feedback involving phosphatases, and molecular crowding (Robichon, 2020).

The aim of this project was to characterize the mechanism of phosphorylation of p38α by MKK6 using ADP as a source of phosphate.

We present our findings concerning this novel and unusual phosphorylation event: MKK6 is able to phosphorylate and activate p38α using the β-phosphate of ADP and ATP *in vitro*. We suggest that this non-canonical phosphorylation route confers adaptation to stress in ATP deprived conditions in order to maintain signalling. It could also be favoured by pseudo-processivity in crowded cellular environment.

Materials and methods

Protein expression and purification

Recombinant protein constructs

p38α refers to the human wild-type Mitogen-activated protein kinase 14 canonical isoform CSBP2 (Uniprot MK14_HUMAN, Q16539-1). p38α K53R refers to the p38α kinase dead mutant.

MKK6 refers to the human wild-type Dual specificity mitogen-activated protein kinase kinase 6 isoform 1 (Uniprot MP2K6_HUMAN, P52564-1). MKK6DD refers to the constitutively activated mutant S207D T211D of MKK6 that mimic the phosphorylated activation loop.

Plasmids encoding p38α fused to a His6 tag and 3C protease cleavage site and MKK6 S207D T211D fused to a twin StrepII tag and 3C cleavage site were ordered from Genscript (gene synthesis, cloning in expression vector and mutagenesis).

Lambda Phosphatase plasmid was ordered from Addgene (a gift from John Chodera & Nicholas Levinson & Markus Seeliger ; Addgene plasmid # 79748 ; http://n2t.net/addgene:79748 ; RRID:Addgene_79748) (Albanese *et al.*, 2018).

Recombinant protein constructs used in this chapter are presented in Table 7.

Table 7: Recombinant protein constructs

Protein	**Construct**	**Protein sequence** **(tag and protease recognition cleavage site in bold and cleavage site indicated by ↕)**	**MW (cleaved protein MW)**
p38α (1-360)	N-ter His6 tag + 3C cleavage site pET-28b+ (Kan)	MGSS**HHHHHH**SSG**LEVLFQ↕GP**MSQERPTFYRQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVKK LSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQIL RGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCI MAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLD SDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,594 kDa (41,448)
p38α K53R (1-360)	N-ter His6 tag + 3C cleavage site pET-28b+ (Kan)	MGSS**HHHHHH**SSG**LEVLFQ↕GP**MSQERPTFYRQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLRVAVRK LSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQIL RGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGYVATRWYRAPEIMLNWMHYNQTVDIWSVGCI MAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFANVFIGANPLAVDLLEKMLVLD SDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSLTYDEVISFVPPPLDQEEMES*	43,622 kDa (41,476)
MKK6DD	N-ter Twin strep tag + 3C cleavage site pFastBac1	MGSQSKGKKRNPGLKIPKEAFEQPQTSSTPPRDLDSKACISIGNQNFEVKADDLEPIMELGRGAYGVVEKMRHVPSG QIMAVKRIRATVNSQEQKRLLMDLDISMRTVDCPFTVTFYGALFREGDVWICMELMDTSLDKFYKQVIDKGQTIPEDI LGKIAVSIVKALEHLHSKLSVIHRDVKPSNVLINALGQVKMCDFGISGYLVDDVAKDIDAGCKPYMAPERINPELNQKGY SVKSDIWSLGITMIELAILRFPYDSWGTPFQQLKQVVEEPSPQLPADKFSAEFVDFTSQCLKKNSKERPTYPELMQHPFF TLHESKGTDVASFVKLILGD**LEVLFQ↕GPWSHPQFEKGGGSGGGSGGSAWSHPQFEK***	41,345 kDa (38,321)
Lambda phosphatase	none pET13S-A (Spec)	MRYYEKIDGSKYRNIWVVGDLHGCYTNLMNKLDTIGFDNKKDLLISVGDLVDRGAENVECLELITFPWFRAVRGNHE QMMIDGLSERGNVNHWLLNGGGWFFNLDYDKEILAKALAHKADELPLIIELVSKDKKYVICHADYPFDEYEFGKPVDH QQVIWNRERISNSQNGIVKEIKGADTFIFGHTPAVKPLKFANQMYIDTGAVFCGNLTLIQVQGEGA	25,219

Plasmid amplification, protein expression in *E. coli* and in insect cells, purification, SDS-PAGE gels, and mass spectrometry experiments were performed as described in Chapter I (p. 51).

Nucleotides and analogues

To study phosphorylation events, we used a range of nucleotides and non-cleavable analogues (Figure 38).

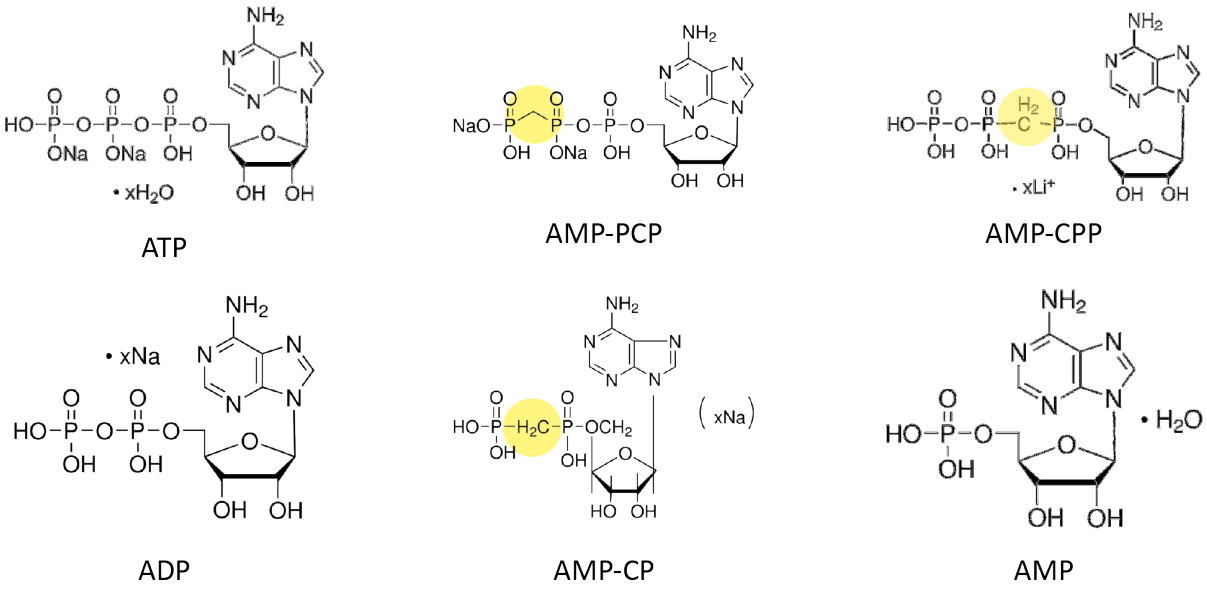


Figure 38: Nucleotides and their analogues.

Phosphorylation assays with native-PAGE gels

Pre-cast 4-20% gradient Tris-Glycine gels (ThermoFisher Scientific) were used. Protein samples (monomer or complex) were prepared at 2.5 µM in 10 µl on ice in assay buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP). Nucleotides were added to a final concentration of 10 mM, and samples were incubated for 20 min at RT. Native sample buffer (50% glycerol, 10 mM Tris pH 6.8, bromophenol blue traces – 4 µl) was added to each sample and each sample (15 µl final) was loaded, and the gel run in Tris-Glycine Native running buffer (2.5 mM Tris Base, 19.2mM glycine pH 8.3) at 125 V for 1h. Gels were stained with InstantBlue (Expedeon) or Coomassie staining (when for MS analysis) and imaged with a BioRad GelDoc system.

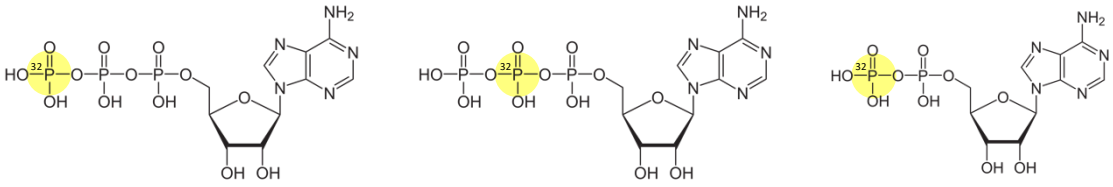
The phosphorylation state of each band was confirmed by LC-MS/MS analysis.

Phosphorylation assays with radiolabelled nucleotides

Protein samples (monomers or complex) were prepared at 0.1 mg/ml in assay buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol, 0.5 mM TCEP) on ice. Radiolabelled

nucleotide sources (ATP γ-^{32}P (Perkin Elmer), ATP β-^{32}P, and ADP β-^{32}P (Hartmann Analytic GmbH) (Figure 39) were diluted 1:10 into water and cold nucleotide (ATP or ADP) to achieve the desired cold nucleotide concentration. The reaction was set up by adding 1 µl of nucleotide to 10 µl of protein sample and incubated for 20 min at RT. The reaction was stopped by adding 4 µl of SDS sample buffer (0.4% bromophenol blue, 0.4 M DTT, 0.2 M Tris pH 6.8, 8% SDS, 40% glycerol) and boiling for 5 min at 95 ℃. Samples were centrifuged at 15 000 g for 5 min and loaded on Pre-cast 4-20% gradient Tris-Glycine gels (ThermoFisher Scientific). The gel run in Tris-Glycine running buffer (2.5 mM Tris Base, 19.2mM glycine pH 8.3, 1% SDS) at 220 V for 40 minutes. Precision Plus Protein Dual Xtra standards were used as molecular weight markers (BioRad). The gels were exposed to a storage phosphor screens (GE) O/N and imaged using a Typhoon scanner (GE Health). Gels were then stained with InstantBlue (Expedeon) and scanned.

Images were analysed using ImageJ (Schneider, Rasband and Eliceiri, 2012) or Image Lab (Bio-Rad) software.



ATP γ-^{32}P ATP β-^{32}P ADP β-^{32}P

Figure 39: Radiolabelled nucleotides.

Results

MKK6 is one of the activators of p38α along with MKK3 and MKK4. Phosphomimetic mutations on its activation loop (MKK6DD for S207D T211D) confer constitutive activation to the MAPKK (Wilsbacher and Cobb, 2001). We discovered that MKK6DD is able to phosphorylate p38α in the presence of ADP as the sole source of phosphate.

p38α is phosphorylated on its activation loop Thr180 and Tyr182 residues by MKK6DD in the presence of ADP

ADP can be used as a phosphoryl source

To assess the ability of MKK6DD to phosphorylate p38α in presence of different nucleotides, we ran NATIVE-PAGE gels to separate the different phosphorylation states of p38α. Bands were cut out from the gel and analysed by LC-MS/MS to confirm their phosphorylation states. We complemented this analysis with native ESI-QTOF MS experiments as well as LC-MS/MS detection of phosphorylated sites on samples in solution.

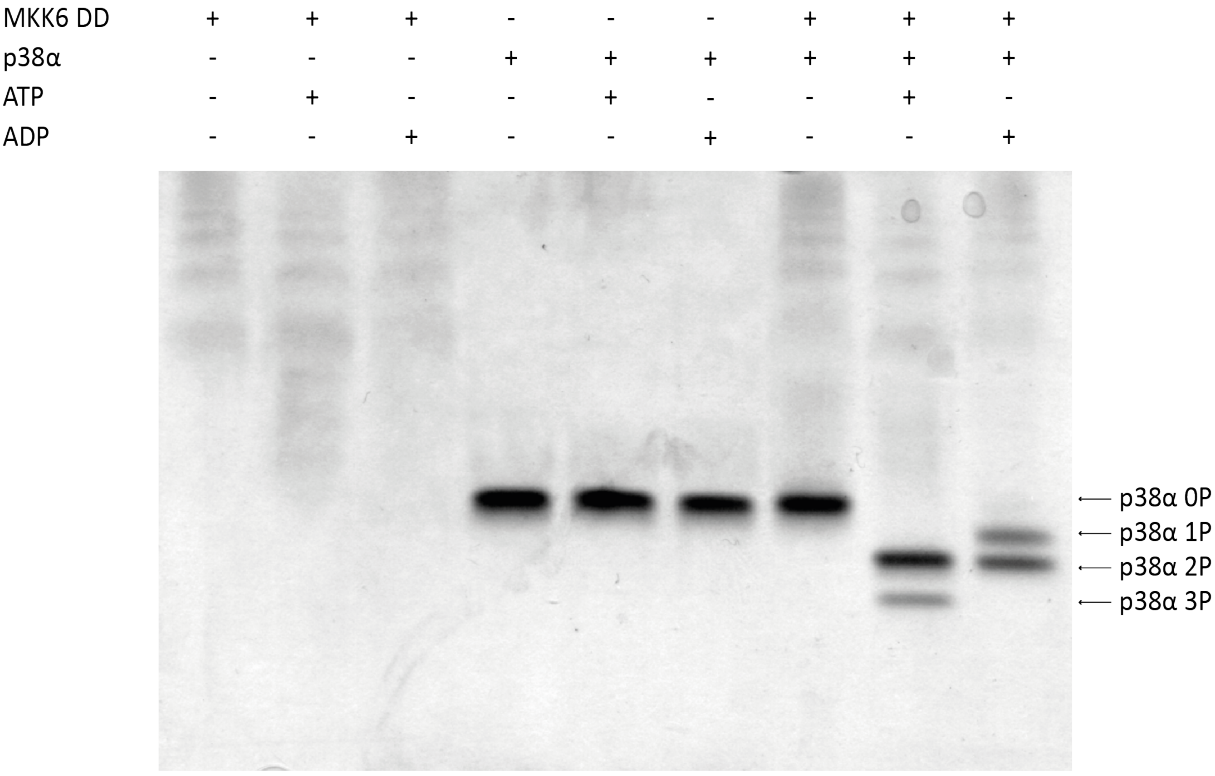


Figure 40: Native PAGE gel of MKK6DD + p38α WT in presence of 10 mM ATP, ADP, or without nucleotides. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P and triphosphorylated 3P).

We first performed endpoint experiments by incubating the protein samples with an excess of nucleotides (Figure 40). MKK6DD bands tend to be smeared, and p38α runs as well-defined bands depending to the number of phosphorylated residues it carries.

In the presence of ATP, p38α presented a total of three phosphorylated sites. It was phosphorylated by MKK6DD on both phosphorylation sites of its activation loop $T^{P}GY^{P}$ (Thr180 and Tyr182), as confirmed by LC-MS/MS detection of phosphorylated sites (Figure 41). Extra minor phosphorylation on other residues of p38α were also observed, resulting in the presence of a band corresponding to 3P on Native PAGE gels as confirmed by MS. No unphosphorylated or monophosphorylated forms of p38α were detected.

In the presence of ADP, we could observe both mono- and biphosphorylated forms in similar proportions (Figure 40 and Figure 41). Interestingly, both monophosphorylated p38α $T^{P}GY$ and TGY^{P} forms were present. More phosphorylated peptides were detected for Thr180 than for Tyr182.

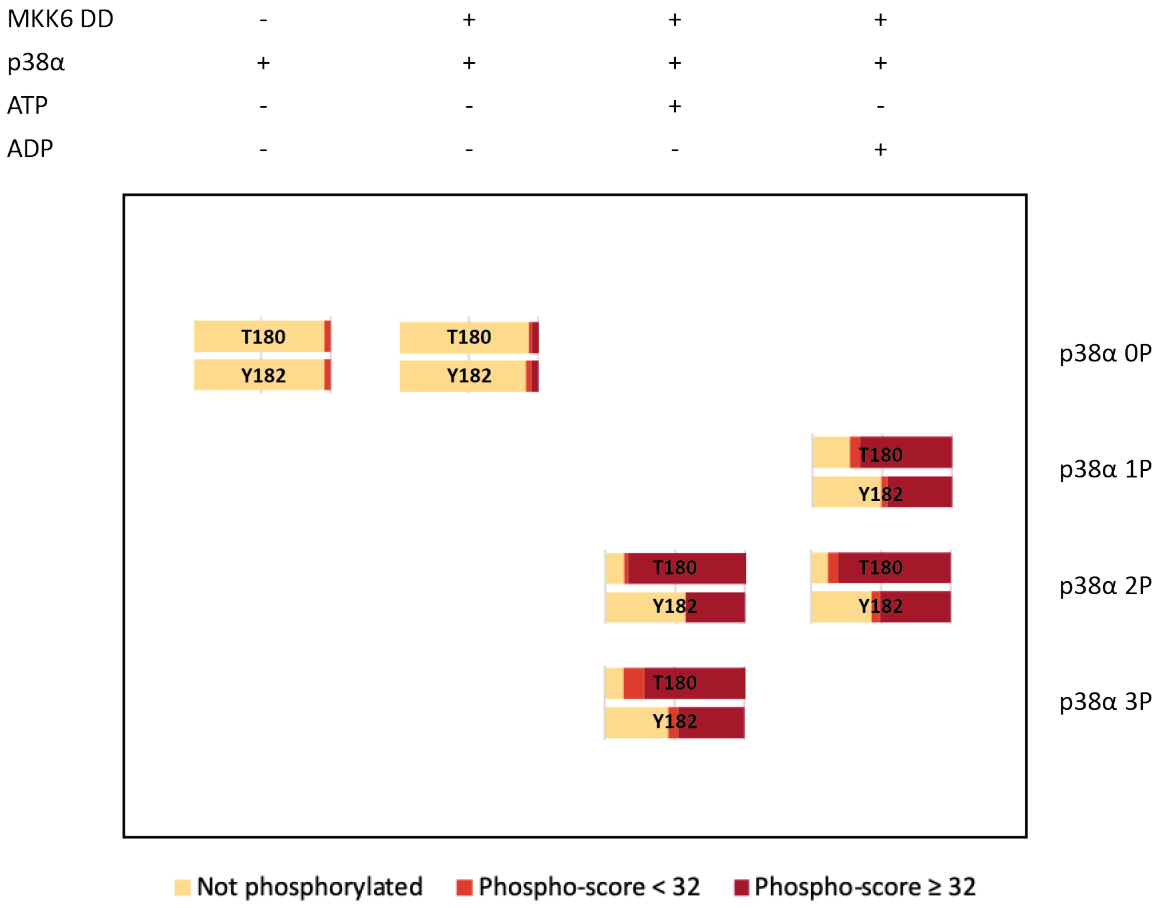


Figure 41: LC-MS/MS analysis of bands extracted from a Native PAGE gel of MKK6DD + p38α WT in the presence of 10 mM ATP, ADP or without nucleotides. The results are displayed in the same layout as the gel. For each analysed band, the results of the phosphorylation state of Thr180 (top) and Tyr182 (bottom) of p38α are represented as a percentage of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (orange) and phosphorylated (red) by MASCOT analysis. Bands that were significantly less abundant are represented in faint colours.

We then tried to assess if complete dual phosphorylation can be achieved with ADP as a phosphate source by testing higher ADP concentrations and longer incubation times, both on ice and at RT (Figure 42). We did not observe an increase in the biphosphorylated forms by incubating up to 7 hours. When incubating O/N, no bands were detected because of protein denaturation. Further study will be necessary to determine if a limiting factor for achieving complete dual phosphorylation of p38α with ADP exists, or if it results from lower affinity and slower kinetics.

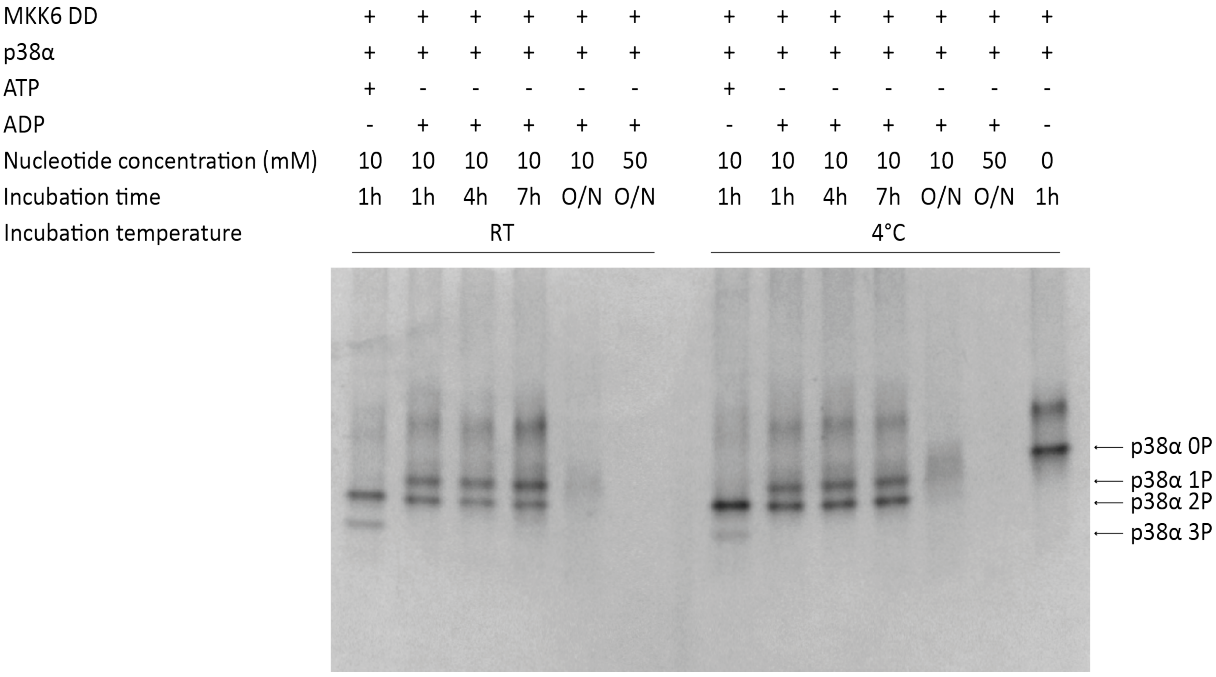


Figure 42: Native PAGE gel of MKK6DD + p38α WT in presence of ATP, ADP, or without nucleotides. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P and triphosphorylated 3P).

ADP is less efficient than ATP to phosphorylate p38α

The dual-phosphorylation of p38α by MKK6 with ATP is highly efficient and difficult to trap in intermediate states. We performed enzymatic reactions with limiting concentrations of nucleotides to trap intermediate phosphorylation states and compare dynamics between ATP and ADP (Figure 43). Kinases (5 µM) were incubated with nucleotides for 20 minutes at RT. In the experiment conditions, monophosphorylated p38α (on the activation loop, either $T^{P}GY$ or TGY^{P} as confirmed by LC-MS/MS) appeared at a concentration of 1 µM of ATP, or 100 µM of ADP (100x higher). Biphosphorylated p38α (on the activation loop $T^{P}GY^{P}$ as confirmed by LC-MS/MS) appeared from a concentration of 100uM of ATP, or of 10 mM of ADP. The transition from the mono- to biphosphorylated form of p38α was almost complete at 100 µM ATP, and at higher concentrations, an extra band appeared corresponding to a triphosphorylated form of p38α, on various sites as shown by MS.

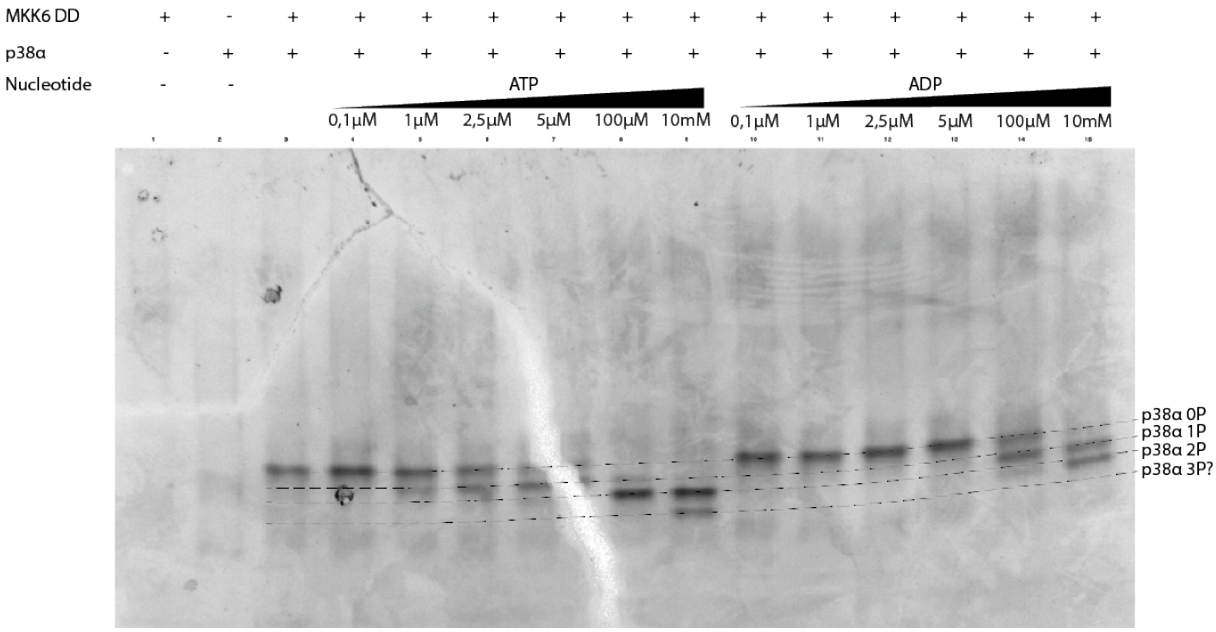


Figure 43: Native PAGE gel of MKK6DD + p38α WT with increasing concentrations of ATP and ADP. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P, and triphosphorylated 3P).

The relative proportion of each phosphorylated form for each condition was quantified and plotted in Figure 44. It appears that in limited nucleotide concentration, p38α phosphorylation first occurs on only one amino acid of the activation loop, and when ~50% of the p38α population is monophosphorylated, biphosphorylated forms start appearing, both when using ATP or ADP as a source of phosphate.

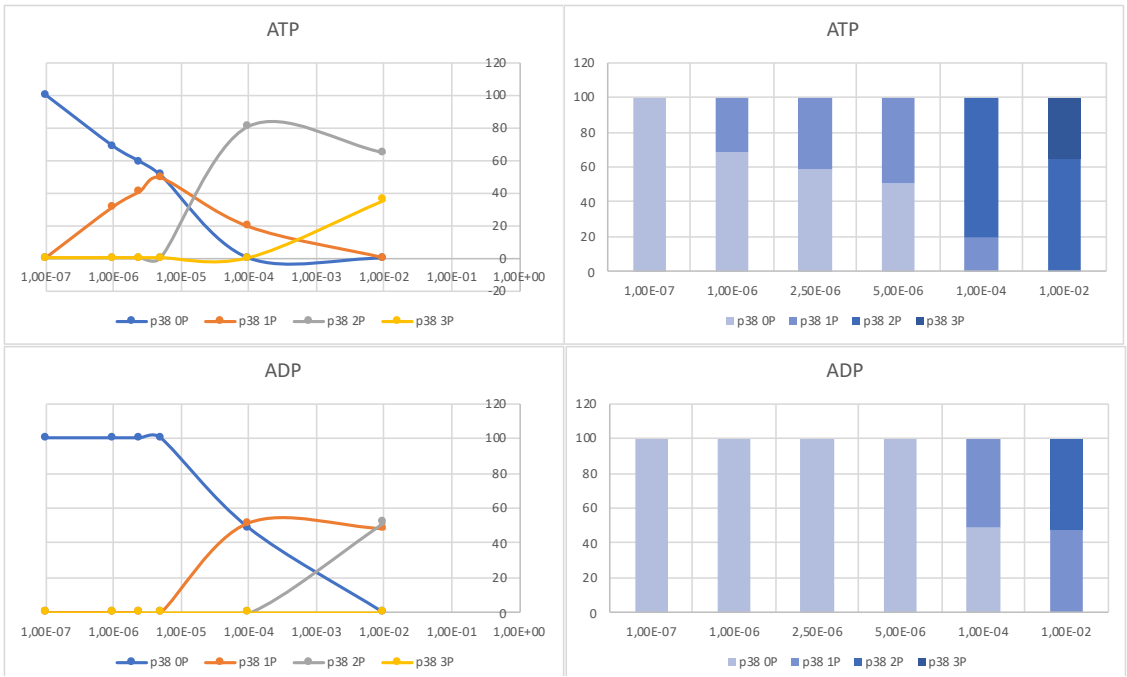


Figure 44: Quantification of p38α bands from the native PAGE gel of Figure 43.

This delay in the formation of biphosphorylated p38α was previously observed (Humphreys *et al.*, 2013), and is indicative of a distributive mechanism.

p38α phosphorylation with ADP is due to MKK6DD activity

We also measured the intact mass of MKK6DD + p38α WT samples incubated with different nucleotides (AMP, ADP and ATP, as well as ultra-pure ADP (uADP, Cell Technology) to rule out potential ATP contamination of our regular ADP stock) with native ESI-QTOF MS experiments (Figure 46A). The same samples were also analysed by LC-MS/MS, phosphorylated sites were detected using MASCOT (Appendix 3A, p. 213).

To assess the possibility that this phosphorylation event in the presence of ADP is due to possible auto-phosphorylation activity of p38α when in the presence of MKK6DD, we performed the same assays using a kinase-dead p38α mutant (p38α K53R) (Figure 46B and Appendix 3B, p. 213).

We encountered a few deviations from theoretical mass of the different proteins. The p38α WT and K53R proteins were detected with a mass deviation of ~14 Da greater than the expected mass. This could correspond to an acetylation. As in Chapter I, p38α constructs (WT and K53R) contained a minor population (~10%), with a mass deviation of ~44 Da greater than the expected mass. MKK6DD was present in a major and a minor population (~15%), with respective mass deviations of ~120 Da and ~70-80Da less than the expected mass. These deviations were not explained by measurement error (the sensitivity of ESI-QTOF MS measurement is +/- 2 Da), and all the plasmids were controlled by sequencing. All the populations behave similarly with increasing phosphorylated site number when adding ADP and ATP (+ 80 Da for each extra phosphorylated site).

The p38α WT results are similar to the data obtained from the Native PAGE gel analysis (Figure 40, p. 106): without nucleotide, or in presence of AMP, the majority of p38α WT population is unphosphorylated. With ADP and uADP, p38α WT is present in a mixture of both mono- and biphosphorylated forms. With ATP, most of p38α WT is at least biphosphorylated.

The p38α K53R phosphorylation pattern with different nucleotides is the same as for p38α WT with ADP and uADP, indicating that the ability to use ADP as a phosphate donor comes from MKK6DD (Figure 46B and Appendix 3B, p. 213). The phosphorylation of p38α appears to be slightly less efficient with ATP for the K53R mutant compared to the WT, with the presence of a monophosphorylated peak. There is no peak for triphosphorylated form of p38α K53R. This suggests that this extra phosphorylation on other site than the activation loop results from the biphosphorylated p38α activity and not MKK6DD activity (Figure 45).

MKK6DD is mainly non-phosphorylated when incubated with p38α WT or K53R without nucleotide or in the presence of AMP (Figure 46). Some phosphorylation is systematically detected on its Thr28 (Appendix 3, p. 213), probably occurring during the insect cell expression as previously observed in Chapter I.

When incubated with p38α WT in the presence of ATP, MKK6DD is present in a combination of mono-, bi- and triphosphorylated forms (Figure 46A). It is phosphorylated on several sites of it disordered N-terminal extension, and a few other sites (Appendix 3B, p. 213). These sites are barely phosphorylated with the dead-kinase mutant p38α K53R (Figure 46B and Appendix

3B), showing that most phosphorylated sites observed in MKK6DD result from p38α activity (Figure 45).

The same phosphorylated sites as with ATP are detected on MKK6DD when incubated with p38α WT in the presence ADP or uADP but in lower proportions (Figure 46A and Appendix 3A, p. 213). Interestingly, these phosphorylation events do not occur with the dead-kinase mutant p38α K53R. This could indicate that p38α might also be able to use ADP in some extent to phosphorylate MKK6DD once phosphorylated and activated with ADP by the MAP2K.

Overall, the comparison between the phosphorylation patterns obtained with p38α WT or kinase dead mutant K53R shows that MKK6DD is mainly responsible for p38α phosphorylations, and conversely p38α is responsible for most of MKK6 observed phosphorylations. Some sites can be phosphorylated using either ADP or ATP, other sites are only phosphorylated in the presence of ATP. We can therefore classify the different phosphorylation sites as described in Figure 45. Interestingly, besides the p38α activation loop residues Thr180 and Tyr182 that are highly phosphorylated by MKK6 with both ADP and ATP, two additional sites are phosphorylated by MKK6 with both ADP and ATP in a lesser extent: p38α Tyr258 and Thr263, located in the MAPK insert. p38α Tyr258 is one of the key residues mediating the interaction between the αG-helix of MKK6 and the p38α hydrophobic pocket characterized in Chapter I.

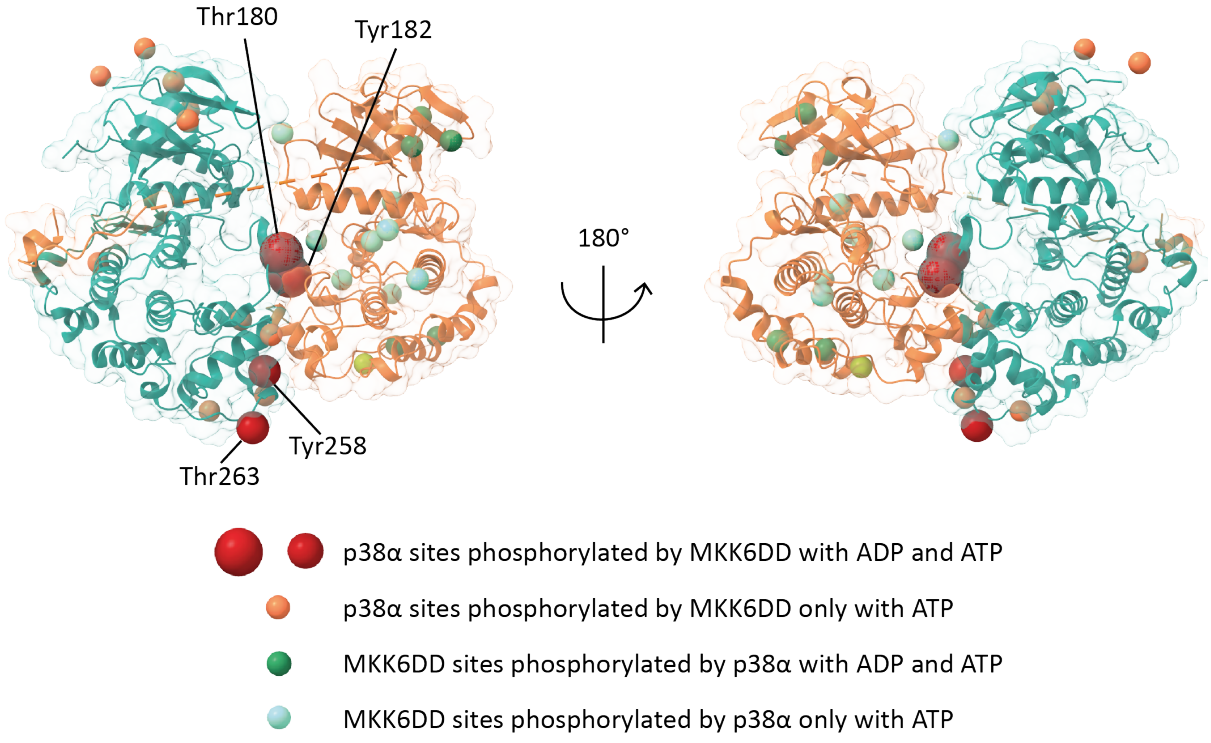


Figure 45: Phosphorylation sites of MKK6DD and p38α detected by MS and represented on the cryoEM model of MKK6DDGRA/p38α T180V presented in Chapter I. The phosphorylated sites are represented as balls (their size (small, medium and big) corresponds to the relative proportions of the phospho-site observed in MS).

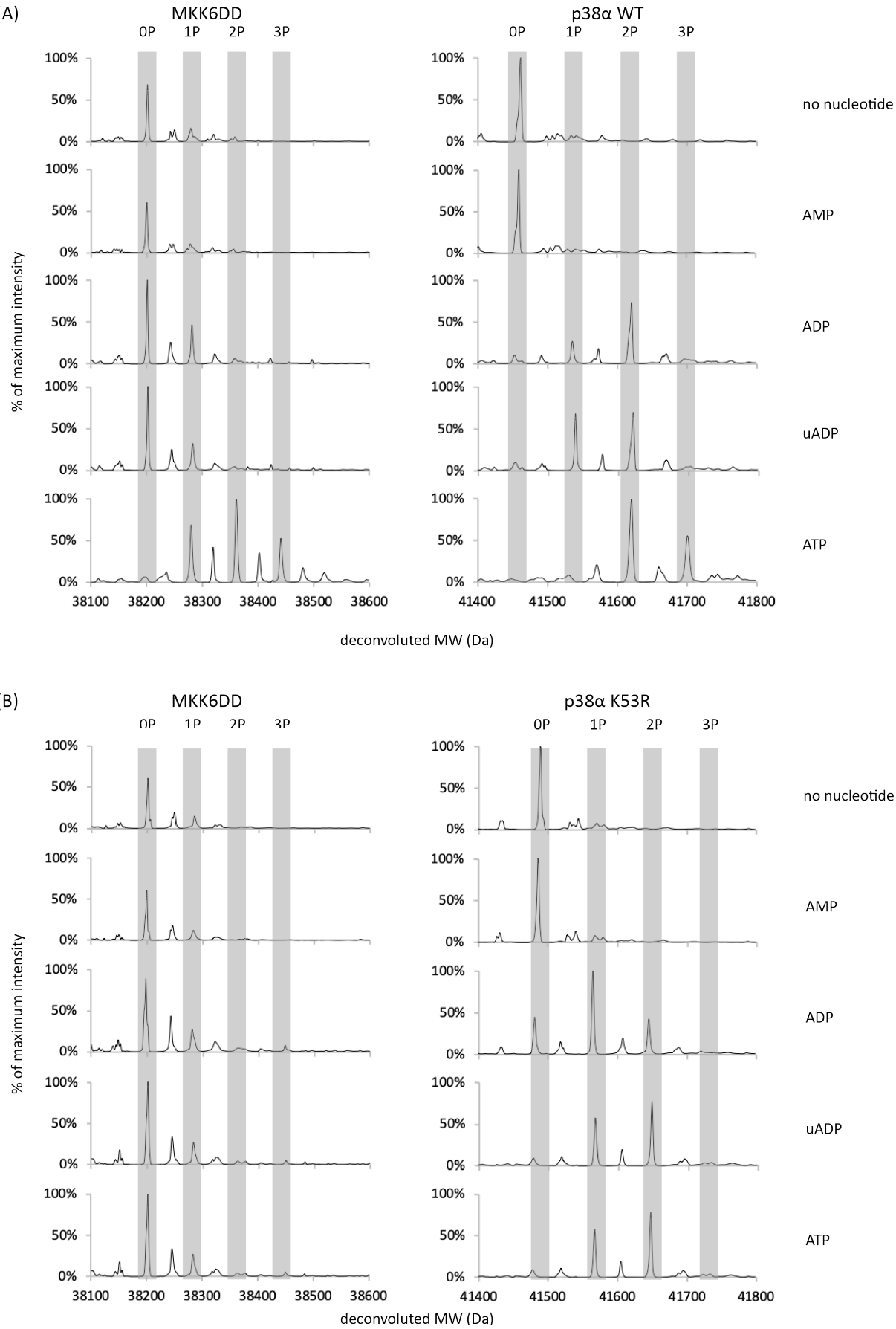


Figure 46: native ESI-QTOF MS spectra of MKK6DD with p38α WT (A) and K53R (B) samples in the presence of 10 mM AMP, ADP, uADP, ATP or without nucleotide. For each sample, peak for each MKK6DD and p38α phosphorylation form is highlighted.

We then focused our analysis of the LC-MS/MS experiments on the two phosphorylation sites of p38α activation loop Thr180 and Tyr182 when incubated with MKK6DD and different nucleotides (Figure 47). The observations were consistent with the previous experiment with

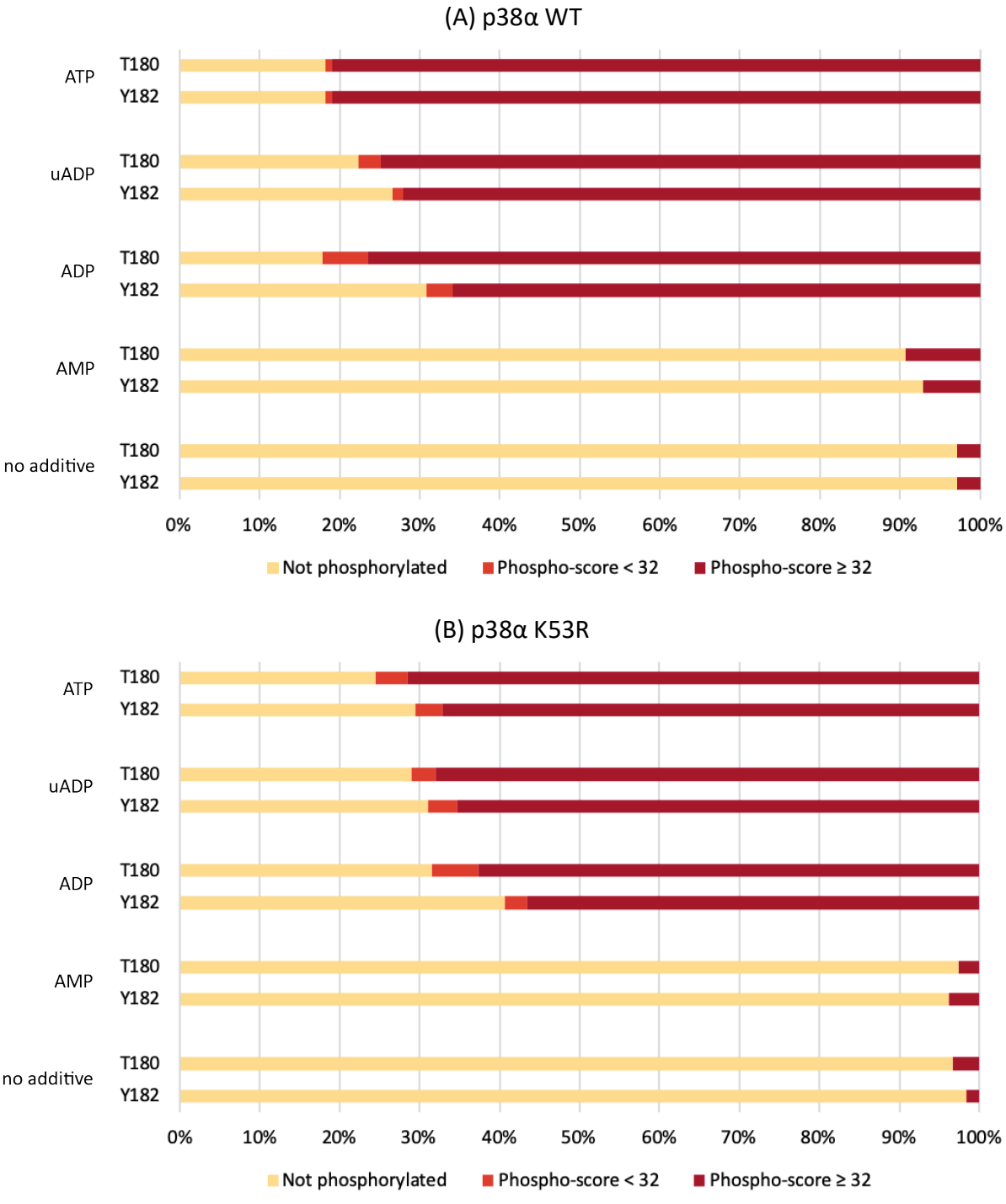


Figure 47: Phosphorylation state of p38α Thr180 and Tyr182 from the LC-MS/MS analysis of MKK6DD with p38α WT (A) and K53R (B) samples in presence of 10 mM AMP, ADP, uADP, ATP or without nucleotide. For each analysed sample, the results of phosphorylation state of Thr180 (top) and Tyr182 (bottom) of p38α are represented in % of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (orange) and phosphorylated (red) by MASCOT analysis.

p38α WT and K53R mutant behaving similarly. Low levels of phosphorylation of the 2 sites was detected in the absence of nucleotide or with AMP. In the presence of ADP and uADP, both phosphorylation sites presented phosphorylation. When ATP was added, both phosphorylation sites were also mainly phosphorylated.

These MS measurements and analyses are not quantitative. Therefore we cannot draw conclusions from the comparison of close percentages of detected phospho-peptides or peak intensities between samples. We can only say that:

- both Thr180 and Tyr182 can be phosphorylated in the presence of ADP
- MKK6DD is responsible for these phosphorylations
- there is a combination of monophosphorylated $T^{P}GY$, monophosphorylated TGY^{P}, and biphosphorylated $T^{P}GY^{P}$ p38α present.

The use of β-phosphate is not required to achieve dual phosphorylation of p38α

One of the hypotheses emerging from our initial observation was that the route to achieve p38α dual phosphorylation might use a single ATP molecule, hydrolysing the γ-phosphate to phosphorylate a first residue in the activation loop, and subsequently hydrolysing the β-phosphate of the ADP produced in the first step to phosphorylate the second residue in the activation loop.

To assess if the use of β-phosphate is required to achieve dual phosphorylation of p38α, we tested the enzymatic reaction in the presence of AMP-CPP, an ATP analogue that cannot be hydrolysed between the α- and β-phosphate (Figure 38, p. 104). The behaviour was similar with AMP-CPP as with ATP, resulting in a complete dual phosphorylation of p38α (Figure 48). This indicates that β-phosphate incorporation is not required and supports the hypothesis that the use of β-phosphate might only be an alternative mechanism of phosphorylation.

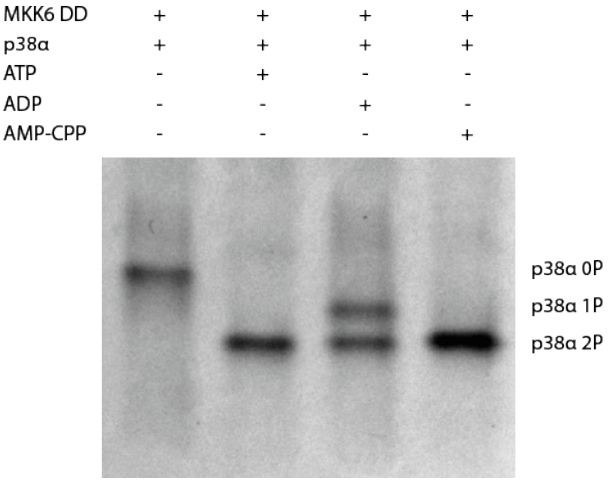


Figure 48: Native PAGE gel of MKK6DD + p38α WT in the presence of 10 mM ATP, ADP, AMP-CPP or without nucleotide. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (non-phosphorylated 0P, monophosphorylated 1P and biphosphorylated 2P).

MKK6 is able to use the β-phosphate of ADP and ATP to phosphorylate p38α

Direct evidence of β-phosphate incorporation

To confirm that MKK6 can use ADP as a phosphate donor, we performed the enzymatic reaction with ^{32}P radiolabelled nucleotides: ATP γ-labelled, ATP β-labelled and ADP β-labelled (Figure 39, p. 105). As expected, we observed that MKK6DD is highly active and phosphorylates p38α using the ATP γ-phosphate (Figure 49). We observed significant signals for phosphorylation using the β-phosphate of ADP, as well as the β-phosphate of ATP.

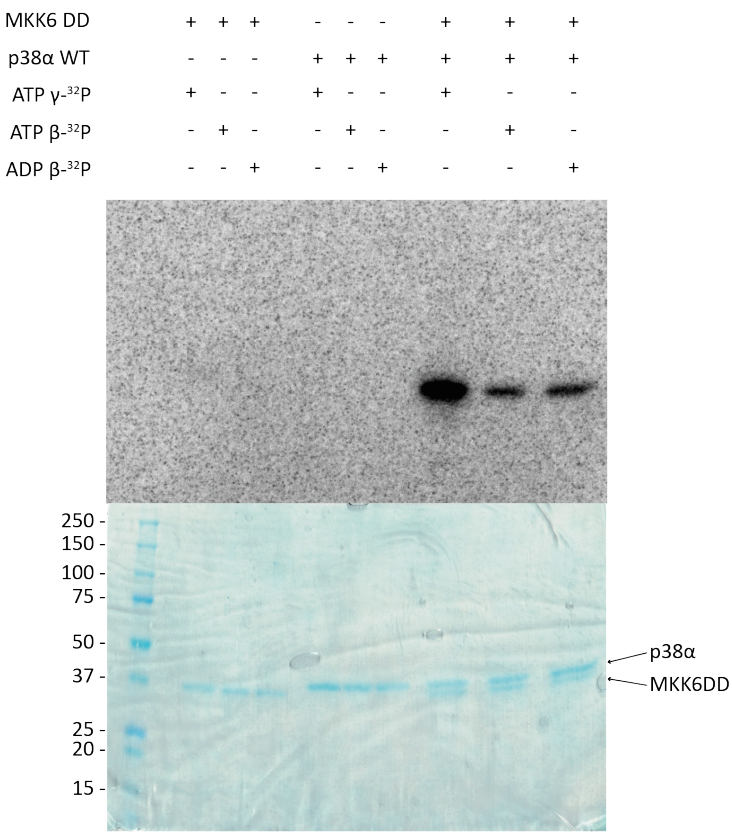


Figure 49: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with 1 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and run on SDS-PAGE gel. The corresponding SDS-PAGE gel was stained after exposure and data acquisition as control.

Our hypothesis is that MKK6 prioritises the use of the γ-phosphate of ATP, but can also use the β-phosphate of ADP coming from the ATP substrate in a much less efficient manner. Therefore, the use of ADP for phosphorylation could also occur in presence of ATP, explaining the observed incorporation of ATP β-phosphate.

We performed the same assay using the kinase-dead p38α mutant (p38α K53R) (Figure 50). Both p38α WT and K53R mutant were phosphorylated using the β-phosphate of ADP.

In this assay, a low signal was obtained for both p38α WT and K53R with ATP β-^{32}P. We addressed this issue in the following section.

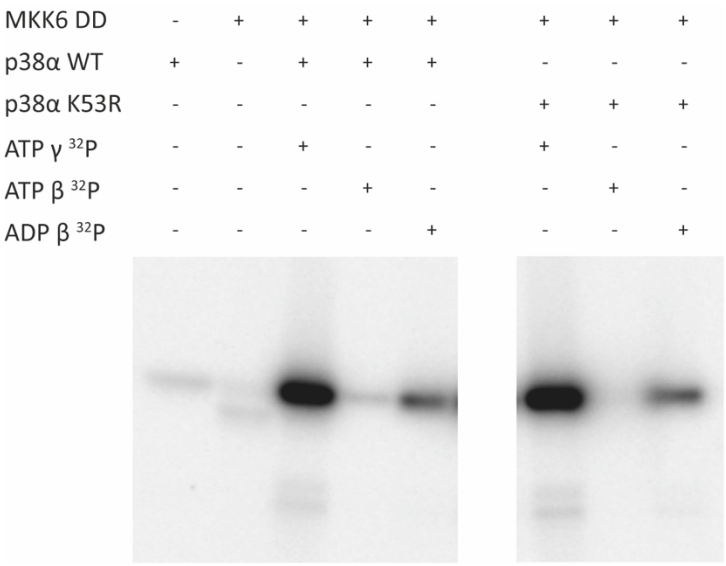


Figure 50: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT or K53R were incubated with 10 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or ß-^{32}P ATP, ADP with ß-^{32}P ADP), and run on SDS-PAGE gel.

Is the nucleophilic attack of β-phosphate dependent on the nucleotide concentration?

The levels of phosphorylation of p38α observed by radiolabelled nucleotide incorporation were dependent on the total concentration of cold nucleotide, especially for ATP β-labelled and ADP β-labelled. Indeed, as radiolabelled nucleotide concentrations are often low and not easily quantifiable, they are usually complemented with cold nucleotides. Figure 51 shows the results of the same experiment with supplement of 10 mM cold nucleotide (Figure 51A), or without cold nucleotide supplement (Figure 51B), indicating that the total nucleotide concentration influences the phosphorylation dynamics and could bias our interpretations.

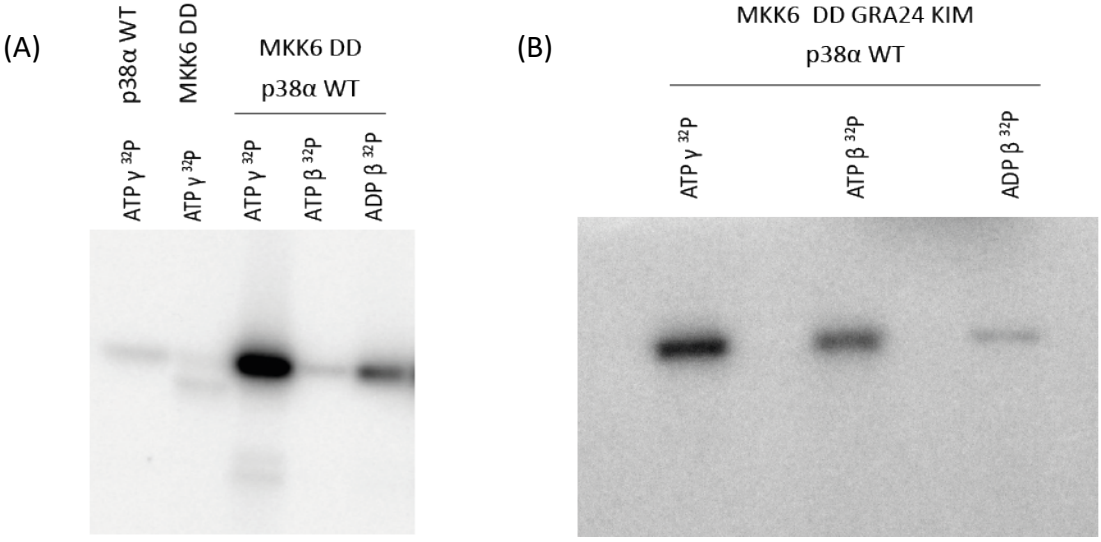


Figure 51: In vitro phosphorylation assay with radiolabelled nucleotides and different cold nucleotide concentrations. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with (A) 10 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or ß-^{32}P ATP, ADP with ß-^{32}P ADP), and (B) with only radiolabelled nucleotide, and run on SDS-PAGE gel.

To better understand these changes depending on the cold nucleotide concentration, we repeated the experiments with a fixed concentration of radiolabelled nucleotide and increasing concentration of cold nucleotide (Figure 52).

When looking at the incorporation of radiolabelled γ-phosphate from ATP onto p38α, we observe a decreasing signal with increasing cold ATP concentration (Figure 52A). This result is coherent with the expected competition effect between the radiolabelled and cold γ-phosphate.

On the other hand, signal with ADP β-labelled increases with increasing concentrations of cold ADP, supporting the low affinity of MKK6 for ADP and the need for a minimum concentration for enzymatic process (Figure 52C). The signal drops significantly in presence of 10 mM cold ATP, this could result from a potential competition effect between the radiolabelled and cold ATP at highest concentrations.

However, results with β-labelled ATP and increasing concentrations of cold ATP are more confusing (Figure 52B). The incorporation of the β-phosphate of radiolabelled ATP seems similar for a concentration of cold ATP from 0 to 1 mM. And no signal is detected in the presence of 10mM cold ATP. This could indicate that the incorporation of β-phosphate does not compete with the incorporation of γ-phosphate.

Surprisingly, we observe a similar pattern for the incorporation of β-radiolabelled phosphate from ADP with increasing cold ATP concentration as with increasing cold ADP concentrations (Figure 52D).

As extra controls, all samples were blotted (1 µl) on a Whatman paper and exposed on phosphor screen to compare their signal intensity, and showed a similar signal, as the concentration of radiolabelled nucleotide is the same in all the samples. The SDS-PAGE gels were stained with Instant blue after screen exposition to control that protein levels were similar for each condition. No anomalies were detected (Appendix 4, p. 215).

These results are difficult to interpret and show the limitations of the use of radiolabelled nucleotides in some biochemical assays. As it is difficult to control the concentrations and proportions of hot and cold nucleotide in a sample, as well as the declining activity of different sources, we cannot compare the efficiency of incorporation of the γ-phosphate of ATP and β-phosphate of ADP and of ATP, as they seem to be influenced by the total nucleotide concentrations in different ways.

Looking at Figure 51A or B separately, one could be tempted to develop some hypotheses on the higher incorporation of β-phosphate of ADP over β-phosphate of ATP, or the opposite.

However, these assays are conclusive proof that MKK6 is able to phosphorylate p38α using the β-phosphate of ADP and ATP, with a lower efficiency than with the canonical use of γ-phosphate of ATP.

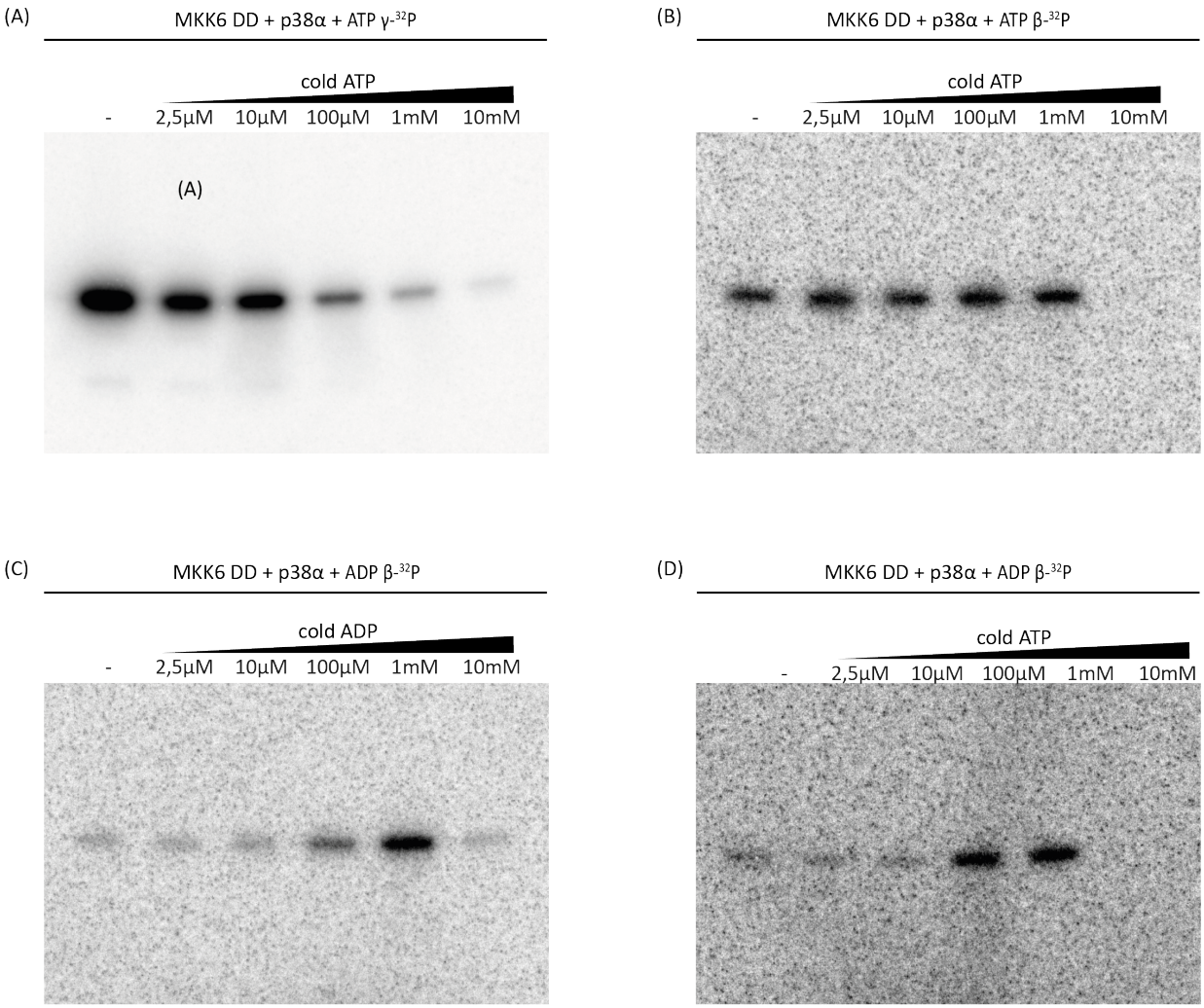


Figure 52: Influence of cold nucleotide concentration on the incorporation of radiolabelled phosphorus. In vitro phosphorylation assay with radiolabelled nucleotides and increasing cold nucleotide concentrations. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with (A) cold ATP with γ-^{32}P ATP, (B) cold ATP with ß-^{32}P ATP, (C) cold ADP with ß-^{32}P ADP, and (D) cold ATP with ß-^{32}P ADP, and run on SDS-PAGE gel.

Can other kinases also use ADP as a source of phosphate?

To explore the specificity of this non-canonical phosphorylation using ADP as a source of phosphate, we compared the phosphorylation of p38α by MKK6 with the auto-phosphorylation of Receptor Interacting protein 2 kinase (RIP2K). RIP2 is an adaptor protein belonging to the innate immune signalling. It comprises a kinase domain, RIP2K, belonging to the TKL kinase group, a different phylogenetic group from p38α and MKK6 (Figure 1, p. 22). Activation of RIP2K occurs via auto-phosphorylation (Pellegrini, Signor, *et al.*, 2017).

We performed enzymatic assays with radiolabelled nucleotides (Figure 53). RIP2K auto-phosphorylates in the presence of ATP as previously described. No phosphorylation was detected using ATP β-^{32}P or ADP β-^{32}P.

RIP2 cannot use ADP as a source of phosphate for auto-phosphorylation. This intriguing use of ADP by a kinase as a source of phosphate presents some specificity. Further characterization is necessary to understand if it is specific to MKK6, or if other kinases share this potential. Looking at other pairs of MAP2K/MAPK will provide a first idea of the specificity. It will also be interesting to see if the signal can be transduced along the MKK6/p38α axis only using ADP: can MAP3Ks phosphorylate MKK6 using ADP? Is p38α active when phosphorylated with ADP? Can p38α use ADP to phosphorylate its substrates? These are questions we will try to address in the continuation of this project.

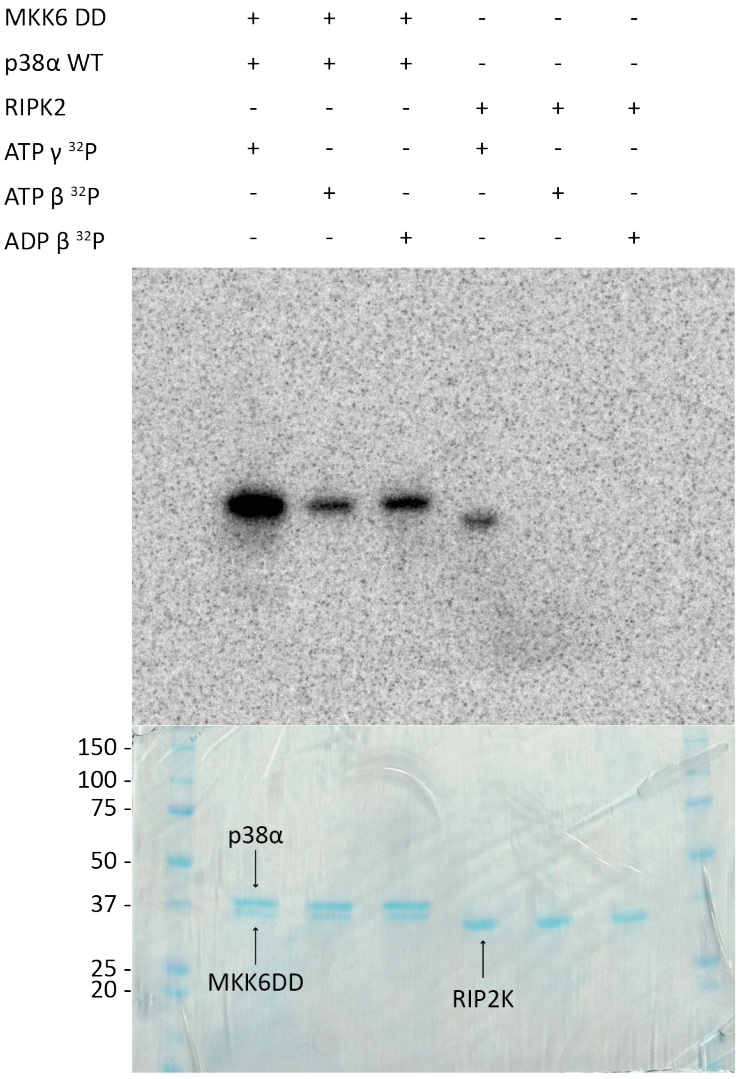


Figure 53: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD + p38α WT, or RIP2K were incubated with 1 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and run on SDS-PAGE gel. The corresponding SDS-PAGE gel was stained after exposure and data acquisition as control.

Discussion and perspectives

We have demonstrated that a novel mechanism of phosphorylation of p38α by MKK6 that uses ADP as a phosphate source exists, and we established direct evidence of the incorporation of the β-phosphate of ADP using radiolabelled nucleotides. Our preliminary results seem to suggest that MKK6 can use both ATP and ADP under normal conditions, but with a much higher affinity and efficiency for ATP. The β-phosphate phosphorylation could exist as a background mechanism, independent of γ-phosphate use. This non-canonical phosphorylation path raises a number of questions and hypotheses.

ADP kinases are rare and cannot use ATP

Phosphorylation using ADP as a source of phosphate is chemically feasible, and releases the same amount of energy as with ATP, but is rarely observed in kinases (Kengen *et al.*, 1994). A small number of atypical kinases, dependent on ADP for phosphorylation, have been described in the literature.

The ADP-dependent sugar kinases were initially discovered in archaea, but are also present in eukaryotes (Ito *et al.*, 2001; Ronimus and Morgan, 2004; Guixé and Merino, 2009). These kinases are part of the ribokinase superfamily and phosphorylate glucose and fructose-6-phosphate using ADP. They can bind ATP but are unable to use it for phosphorylation. Several structures of ADP-dependent sugar kinases, including the *Thermococcus litoralis* ADP-dependent glucokinase (Rivas-Pardo *et al.*, 2013) and the mouse homologue (Richter *et al.*, 2016), have been determined.

The comparison of the nucleotide binding site of ADP- and ATP- dependent glucokinases reveals that the residues recognising the α- and β- phosphate of ADP are conserved and correspond to the ones that recognise the β- and γ- phosphate of ATP. The difference is observed at the site of recognition of adenosine and ribose. In both cases, the nucleotide and sugar substrate are tightly bound to the kinase, placing the atoms perfectly for nucleophilic attack. This tight binding excludes the possibility of other conformations/assemblies that could also be catalytically active with ATP, as the γ-phosphate simply cannot be accommodated (Figure 54B and C).

Known ADP-dependent kinases are mainly limited to sugar kinases. The only ADP kinase able to phosphorylate amino acids described to this date is the *Thermococcus kodakarensis* serine kinase SerK (Makino *et al.*, 2016). SerK is an archaeal ADP-dependent serine kinase involved in cysteine biosynthesis.

There is no mention in the reviewed literature of kinases able to use both ATP and ADP as a source of phosphate. MKK6 is the first protein kinase described to be able to phosphorylate its substrate using either ATP or ADP.

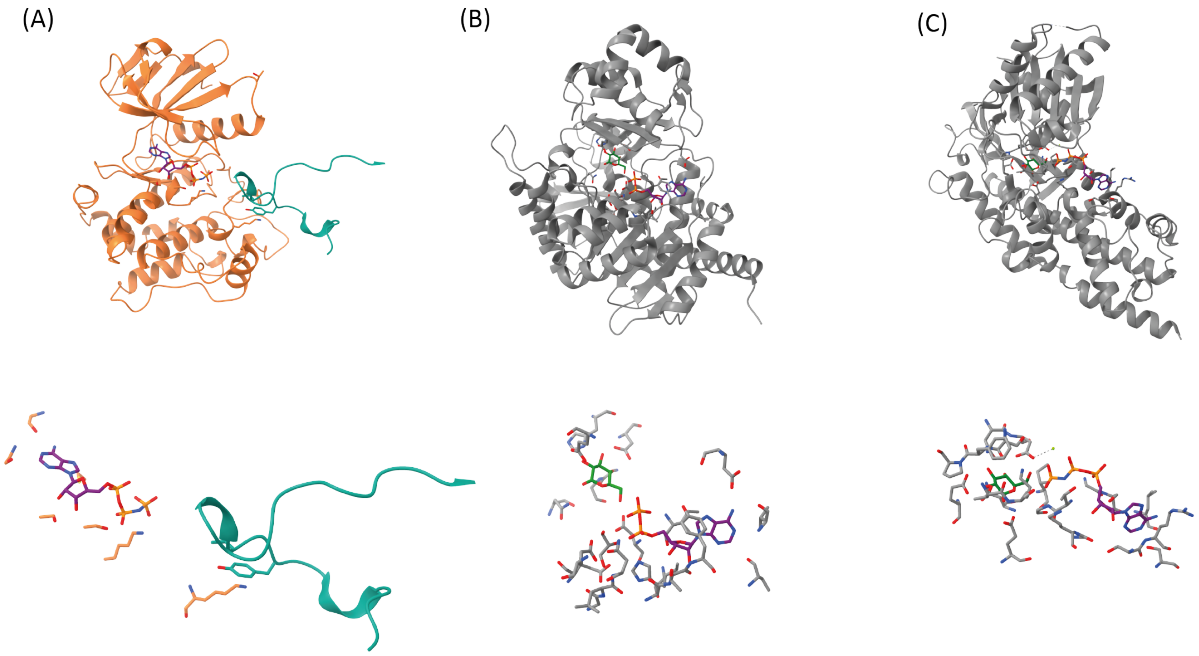


Figure 54: Comparison of kinases with their substrate and nucleotide. In the top panel the kinases are represented as cartoon, the substrates and nucleotides as sticks. The bottom panel is a zoom on the kinase active site, with only key kinase residues displayed as sticks, the substrates and nucleotide s are represented as sticks. (A) MKK6DDGRA + AMP-PNP + p38α activation loop substrate (B) Thermococcus litoralis ADP-dependent glucokinase + ADP + α-D-glucopyranose substrate (PDB 1gc5) (C) human ATP-dependent glucokinase + AMP-PNP + β-D-glucopyranose substrate (PDB 3fgu).

The structural flexibility of the MKK6/p38α catalysis conformation might have a role in the catalysis of both nucleotides

Structural insight into the mechanism of activation of p38α by MKK6, as pursued in Chapter I, provides insight into the structural elements conferring its ability to use both γ-phosphate of ATP and β-phosphate of ADP to phosphorylate p38α on both the threonine and the tyrosine of its activation loop.

The nucleotide binding site of sugar kinases is different than in eukaryotic protein kinases. In the latter family, the ATP binding pocket is deeper and located in between the two lobes. The adenosine and ribose bind deep in the pocket, and the phosphate groups are coordinated by residues closer to the surface of the protein.

There is little interaction between the p38α activation loop and the MKK6 active site (Figure 54A), conferring flexibility in the placement of one or the other phospho-acceptor residue. Most interactions regulating the specificity are distal to the active site. Other examples of kinase-substrate structures in the transition state present several specific binding interactions between the active site surroundings and the substrate residues adjacent to the phospho-acceptor site (Figure 33, p. 89).

In the MKK6/p38α model, both kinases are facing each other, interacting between their C-lobes below the p38α activation loop and between their N-lobes above it. The activation loop is enclosed in between, with space to move inside this "cage" with almost no interaction between MKK6 and the activation loop of p38α itself. In this way, we can imagine that multiple positions are possible for each of the phospho-acceptor residues of the p38α activation loop, adjusting their placement to the nucleotide (ADP or ATP) present in the MKK6 active site. Combined with the flexibility of MKK6 catalytic residues, it could explain the ability of using both ADP or ATP to phosphorylate either Thr180 or Tyr182.

The advantages of using both ADP and ATP for dual-phosphorylation

The use of ADP over ATP would confer an advantage in the accommodation of phosphate groups during catalysis, especially when the substrate is already monophosphorylated on one residue. In this way, the local charge balance would be the same when phosphorylating the already monophosphorylated p38α with ADP as for the first phosphorylation of non-phosphorylated p38α with ATP (Cliff *et al.*, 2010). In case of the first phosphorylation of non-phosphorylated p38α with ADP, the presence of a phosphate group less in the MKK6 active site could also be of interest.

Active MAP2Ks have high affinity for ATP. Phosphorylation kinetics should require lower affinity for ADP, in order to release the ADP product and free the active site for ATP binding. There is only limited information in the reviewed literature regarding measured affinity for ADP, and nucleotide affinity is greatly dependent on the conformational state, the phosphorylation state, and the binding to substrates of scaffolds. However, most kinases seem to have similar affinity for ADP and ATP according to (Becher *et al.*, 2013), and ADP release is a time-limiting step in phosphorylation catalysis (Schwartz and Murray, 2011). Product inhibition by ADP has been reported for several kinases (Machius *et al.*, 2001; Kato *et al.*, 2005; Jagannathan, Kaur and Datta, 2010). The ability to use both ADP and ATP as substrates for phosphorylation could help overcome this limiting step. This hypothesis is even more convincing in the case of a processive mechanism for dual-phosphorylation of the protein substrate.

Processive or distributive?

Whether the dual phosphorylation of p38α by MKK6 proceeds via a processive or a distributive mechanism is still debated in the community. The use of the γ-phosphate of ATP for the first phosphorylation and the β-phosphate of the resulting ADP for the second phosphorylation could favour a fully processive mechanism where the same nucleotide stays bound donating both phosphate groups. This would be favoured by the stability of the total charge in the active site and would potentially gain in efficiency, skipping the steps of release of the ADP produced in the first phosphorylation and rebinding of a second ATP molecule.

In vitro assays with limited nucleotide concentration indicate a distributive mechanism (Figure 43 and Figure 44, p. 109). At low concentrations of ATP, there is first a pool of monophosphorylated p38α and only at higher concentrations a biphosphorylated population

appears. However, cellular molecular crowding can convert this type of distributive mechanism to pseudo-processive due to the low diffusion of the substrate MAPK (Ouldridge and Rein Ten Wolde, 2014). We could therefore hypothesize that the slower diffusion in a crowded environment also affects nucleotides. Consequently, ADP produced by ATP consumption would stay bound to the MAP2K active site longer than *in vitro*. This could increase the use of ADP for phosphorylation by balancing the lower affinity for ADP over ATP, and would allow faster catalysis if nucleotide diffusion becomes a limiting factor.

Based on our MS data, there seems to be no preference for the phosphorylation of one site over the other, with both ATP and ADP. In these experiment conditions, there is no preferential order of phosphorylation with both monophosphorylated forms p38α $T^{P}GY$ and p38α TGY^{P} co-existing.

However, more research on this topic needs to be undertaken. The kinetics of dual-phosphorylation between two kinases is complex, involving a large number of parameters that are difficult to control and/or reproduce *in vitro*. The cellular context, the relative concentrations of kinases and nucleotides, other interacting partners, scaffolding proteins and molecular crowding are all believed to have an influence (Alexa, Varga and Reményi, 2010; Pan *et al.*, 2012; Aoki *et al.*, 2013).

ADP phosphorylation could help maintain active signalling in stress conditions

This alternative route of phosphorylation could confer an adaptive advantage to stress in case of ATP deprivation to maintain the signalling pathway.

p38α is activated by hypoxia-related stress. For instance, it is triggered by ischemia during myocardial infarction or stroke (Kumphune, Chattipakorn and Chattipakorn, 2012; Park and Rongo, 2016). It is associated with tissue damage and contributes to ischemic injury, necrosis and apoptosis resulting in heart failure in the case of cardiac ischemia, or neurodegeneration after a stroke.

In case of hypoxia, ATP cellular levels are diminished (Lee, Chandel and Simon, 2020). Using ADP for signal transduction would allow cells to:

- save ATP and the cellular metabolic demand that cannot be met by mitochondria reduced oxidative phosphorylation in low O_2 conditions
- maintain signal transduction to trigger the appropriate cell response to hypoxia, even if ATP is lacking.

This adaptation to low ATP levels in hypoxia was also suggested as an advantage of the ADP-dependant mouse glucokinase (Richter *et al.*, 2016). In the case of archaeal ADP-dependant sugar kinase, it was suggested that the use of ADP over ATP was an adaptation to high temperatures at which ADP is more stable (Guixé and Merino, 2009).

MKK6 could use ADP instead of ATP to orientate the p38α signalling to specific responses

In the presence of ADP and MKK6DD, p38α is present in both mono- and biphosphorylated forms, in an apparent 50:50 ratio (Figure 40, p. 106). One of our initial hypotheses was that there might be an order to follow to achieve dual phosphorylation with ADP, i.e. that MKK6 cannot use ADP to phosphorylate either p38α $T^{P}GY$ or p38α TGY^{P}, and that there is an order of phosphorylation to follow (first Tyr then Thr, or first Thr then Tyr, respectively). However MS analysis of the monophosphorylated p38α band in NATIVE PAGE gels indicated a mix of both p38α $T^{P}GY$ and p38α TGY^{P}. The kinetic and biochemistry of p38α phosphorylation with ADP by MKK6 need to be explored further to unravel if there is a limiting step to achieve complete dual phosphorylation, or if we could not observe it in our experimental setup due to low affinity and slow process.

However, the presence of both mono- and biphosphorylated p38α forms could also be a hint for specific effects *in vivo*. For instance, in the non-canonical activation pathway involving the phosphorylation of p38α Tyr323 by ZAP70, p38α auto-phosphorylates only on its Thr180 (Salvador *et al.*, 2005). Monophosphorylated p38α exhibits different substrate affinity and activity compared to the classically activated biphosphorylated p38α in T cells (Mittelstadt *et al.*, 2009).

The generation of monophosphorylated p38α $T^{P}GY$ and p38α TGY^{P} forms could limit p38α substrates in certain circumstances where ADP phosphorylation by MKK6 would be favoured over ATP.

Further work will be needed to test and evaluate the activity of p38α phosphorylated by MKK6 using ADP. Careful separation of the different phosphorylated forms, monophosphorylated p38α $T^{P}GY$ and TGY^{P}, as well as biphosphorylated p38α $T^{P}GY^{P}$, would be necessary to assess their individual activity and substrate specificity.

Is the ability to use ADP for phosphorylation shared by other MAP2Ks or other kinases?

The specificity of this novel phosphorylation mechanism still needs to be explored. By extending these experiments to different families, both closely and distantly related, it is hoped that the limits of this mechanism can be delineated between kinase families beyond the MAPK and RIP2K tested here. The first step will be to look at other MAP2K/MAPK pairs. The sequences and structures are highly conserved amongst these families, but the different pathways have different roles and implications in signalling.

Some MS experiments comparing the phosphorylation of p38α WT and K53R dead-kinase mutant by MKK6 using ADP suggested that p38α might also be able to use ADP to phosphorylate MKK6 in return (Figure 45 and Figure 46, p. 111-112). Further work is required to evaluate this possibility. In the scenario where ADP would be used to maintain signalling in ATP deprived condition, it would be coherent that p38α would share this capacity as well to extend the signalling after its activation by MKK6.

It would also be interesting to look at close relative of the MAP2K members in the STE kinase family, but also to dual-specificity kinases that might be evolutionary distant. The common features of dual-specificity kinases remain unclear. They are able to phosphorylate both Ser and Thr as well as Tyr residues. This implies that they are able to accommodate and orientate different sizes of phospho-acceptor into their active site, all correctly aligned with the γ-phosphate of the ATP donor. One of the questions that emerges from this observation is the following: could the adaptability to the different sizes of phospho-acceptor be also involved in the adaptability of different sizes of phospho-donor?

The alternative route of phosphorylation of p38α by MKK6 using ADP as a source of phosphate could have a number of implications *in vivo.* It is important to place such events in a cellular context, where molecular crowding and local concentrations of enzymes and substrates are varying locally and from cell to cell.

Chapter III: Interactions between *T. gondii* GRA24 and human MAPKs

Introduction

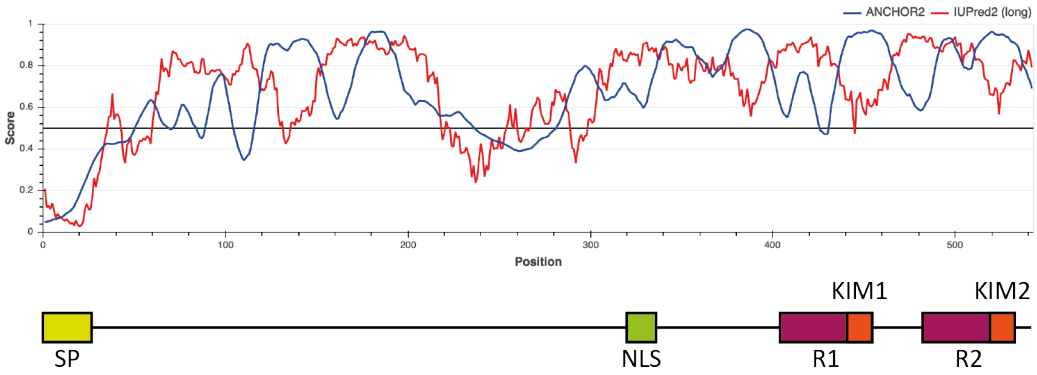
Some pathogens interact with the mammalian MAPK network using different strategies to enable host invasion. *T. gondii* infection triggers the activation of the three main MAPK families: p38α, ERK1/2 and JNK (Valère *et al.*, 2003).

GRA24 is the *T. gondii* effector responsible for p38α activation. It is excreted from dense granules, and exported beyond the PV to the host cell nucleus (Braun *et al.*, 2013). It directly interacts with p38α acting as a shuttling and scaffolding protein to induce p38α auto-phosphorylation. In this way, *T. gondii* bypasses the MKK-dependent phosphorylation to induce p38α activation and IL-12 production.

GRA24 is a 57 kDa intrinsically disordered protein that contains an N-terminal signal peptide (SP), a putative nuclear localization sequence (NLS), and two repeats at its C-terminus (R1 and R2) (Figure 55A) (Dosztányi *et al.*, 2005; Dosztányi, Mészáros and Simon, 2009). These two repeats carry two similar KIM motifs localized in intrinsically disordered regions. These motifs are unusual and required the definition of a new class of KIM (Figure 55B and Table 1, p. 33). They combine the N-terminal definition of the phosphatase HePTP class, with a long and helical N-terminus containing an extra hydrophobic residue φ_u, and a C-terminal description similar to the DCC class, with a potential polyproline II helix-like geometry.

The structure of the GRA24 KIM1 peptide bound to p38α was determined, demonstrating the interaction of the new KIM motif with the MAPK (Figure 55C) (Pellegrini, Palencia, *et al.*, 2017). In the proposed model, GRA24 interacts with two molecules of p38α through the binding of its KIM motifs to their docking site. This docking interaction induces an allosteric conformational change in p38α molecules from the inactive conformation to a closed conformation that tends towards the active one, in which the activation loop is disordered and the active site can accommodate an ATP molecule. Through this conformational change and the close proximity between the two p38α molecules, GRA24 enables p38α to auto-phosphorylate in a *trans* conformation.

(A)



(B)

φ_U-x_2-θ-θ-G-x_{3-6}-φ_L-(P-x/x-P)-φ_A-x-φ_B

KIM1: L L E R R G V S E L P P L Y I

KIM2: L L E R R G V S E L P P L R I

(C)

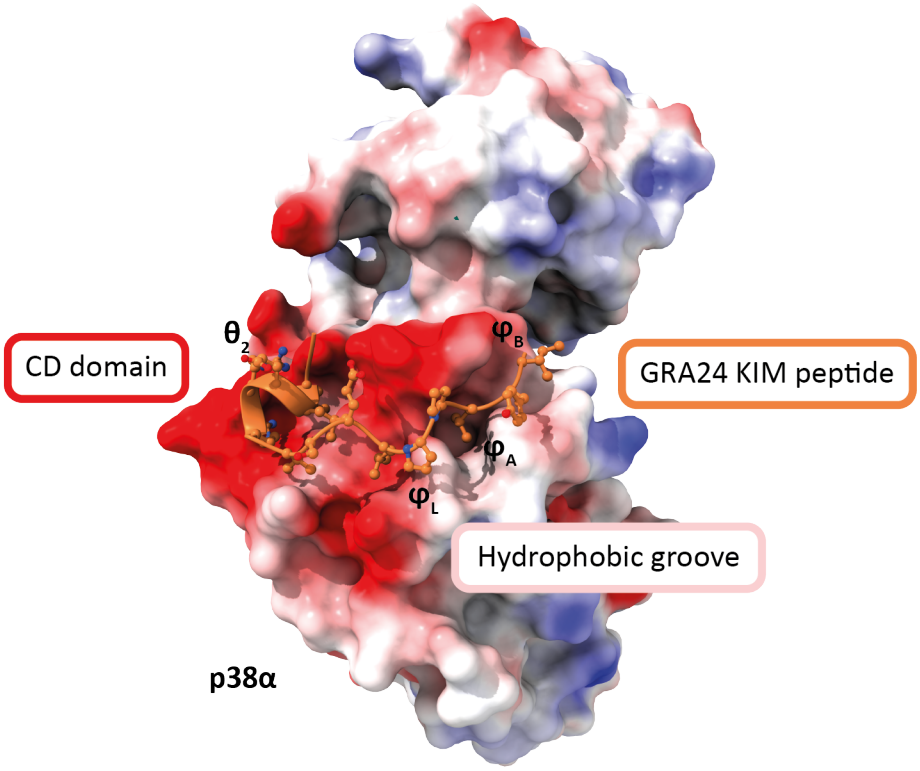


Figure 55: GRA24 and its interaction with p38α (A) Diagram of GRA24 domains (SP signal peptide, NLS nuclear localization signal, R1 and R2 repeats containing KIM1 and KIM2 motifs) with corresponding IUPred2 (intrinsically disordered regions prediction) and ANCHOR2 (prediction binding regions that can undergo disorder-to-order transition upon binding) scores (B) GRA24 KIM class description and KIM1/2 sequences (C) Crystal structure of GRA24 KIM1 peptide bound to the docking site of p38α (PDB 5eta). The surface of p38α is displayed coloured by its electrostatic potential (from -10 (red) to 10 kTe^{-1} (blue)). GRA24 KIM peptide is represented as cartoon with sticks side chains.

GRA24 KIM peptide affinity for p38α is significantly stronger than most other binding partners described in the literature. It has a K_D of 1.6 µM, which is a 100-fold higher that the affinity of the MKK6 KIM peptide for p38α (Pellegrini, Palencia, *et al.*, 2017). GRA24-dependent p38α

activation is sustained, in contrast to the canonical activation that is more transient. This direct and strong interaction and activation of p38α by GRA24, coupled with the long-lasting secretion of the effector, has several advantages. Thanks to high-affinity, other p38α binding partners will not compete efficiently. It likely avoids canonical negative feedbacks by phosphatases. In infected macrophages, GRA24-activated p38α upregulates transcription factors EGR-1 and c-Fos, that induce the production of pro-inflammatory cytokines, including IL-12 (Braun *et al.*, 2013).

The activation of p38α by GRA24 triggers inflammation during the tachyzoite proliferation and dissemination stage to repress parasites over-proliferation that would kill the host before the establishment of the chronic phase, and thus facilitate long-term parasitism.

T. gondii infection activates the three main MAPK pathways and GRA24 is the responsible for p38α auto-phosphorylation. The effectors responsible for the activation of ERK1/2 and JNK proteins remain unknown.

Collaborators recently showed by immune-precipitation assays that GRA24 also interacts with ERK1/2 during host invasion. ERK1/2 activation has different signalling outcomes than their p38 homologues. It is involved in cell growth and differentiation.

ERK1/2 activation could have a role in ocular toxoplasmosis in which visual impairment is due to neovascularization in the eye (Quan *et al.*, 2020). *T. gondii* infects retinal cells and induces vascular endothelial growth factor (VEGF) production. It was described that ERK1/2 phosphorylation is triggered by VEGF and in response stimulates VEGF expression, creating a positive feedback loop allowing *T. gondii* growth and replication. However, the mechanism of activation of the VEGF and ERK1/2 pathway by *T. gondii* remains unknown. GRA24 is therefore an interesting candidate as a potential ERK1/2 activator.

The aim of this project was to characterize the docking interaction between the GRA24 KIM peptide and ERK1/2, and to evaluate if GRA24 is the effector responsible for ERK1/2 activation during *T. gondii* infection.

We show that GRA24 KIM peptide binds to ERK1 in a similar way to what have been described for p38α. However, we also observed that the phosphorylation of ERK1/2 induced by *T. gondii* infection is not dependent on GRA24. We suggest that the effect of GRA24 on ERK1/2 could occur downstream of the MAPK phosphorylation.

Materials and methods

Protein expression and purification

Recombinant protein constructs

p38α refers to the human wild-type Mitogen-activated protein kinase 14 canonical isoform CSBP2 (Uniprot MK14_HUMAN, Q16539-1). ERK1 refers to the human wilt-type mitogen-activated protein kinase 3 isoform 1 (Uniprot MK03_HUMAN, P27361-1). JNK1 refers to the human wild-type Mitogen-activated protein kinase 8 isoform α (Uniprot MK08_HUMAN, P45983-1).

Plasmids encoding MAPKs (p38α (10-360), ERK1 FL, and JNK1 (5-364)) fused to a His6 tag and TEV protease cleavage site were ordered from Genscript (gene synthesis and cloning in expression vector).

Lambda Phosphatase plasmid was ordered from Addgene (gift from John Chodera & Nicholas Levinson & Markus Seeliger ; Addgene plasmid # 79748 ; http://n2t.net/addgene:79748 ; RRID:Addgene_79748) (Albanese *et al.*, 2018).

Recombinant protein constructs used in this chapter are presented in Table 8.

Table 8: Recombinant protein constructs

Protein	**Construct**	**Protein sequence** **(tag and protease recognition cleavage site in bold and cleavage site indicated by ↕)**	**MW** **(cleaved protein MW)**
p38α (10-360)	N-ter His6 tag + TEV cleavage site pETM-11 (Kan)	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMARQELNKTIWEVPERYQNLSPVGSGAYGSVCAAFDTKTGLR VAVKKLSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLT DDHVQFLIYQILRGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDDEMTGYVATRWYRAPEIMLN WMHYNQTVDIWSVGCIMAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARNYIQSLTQMPKM NFANVFIGANPLAVDLLEKMLVLDSDKRITAAQALAHAYFAQYHDPDDEPVADPYDQSFESRDLLIDEWKSL TYDEVISFVPPPLDQEEMES*	43,482 kDa (40,484)
ERK1 (FL 1-379)	N-ter His6 tag + TEV cleavage site pETM-11 (Kan)	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMAAAAAQGGGGGEPRRTEGVGPGVPGEVEMVKGQPFDVG PRYTQLQYIGEGAYGMVSSAYDHVRKTRVAIKKISPFEHQTYCQRTLREIQILLRFRHENVIGIRDILRASTLEA MRDVYIVQDLMETDLYKLLKSQQLSNDHICYFLYQILRGLKYIHSANVLHRDLKPSNLLINTTCDLKICDFGLARI ADPEHDHTGFLTEYVATRWYRAPEIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPGKHYLDQLNHILGILGSP SQEDLNCIINMKARNYLQSLPSKTKVAWAKLFPKSDSKALDLLDRMLTFNPNKRITVEEALAHPYLEQYYDPT DEPVAEEPFTFAMELDDLPKERLKELIFQETARFQPGVLEAP*	46,262 kDa (43,264)
JNK1 (5-364)	N-ter His6 tag + TEV cleavage site pETM-11 (Kan)	MK**HHHHHH**PMSDYDIPTT**ENLYFQ↕G**AMAKRDNNFYSVEIGDSTFTVLKRYQNLKPIGSGAQGIVCAAYD AILERNVAIKKLSRPFQNQTHAKRAYRELVLMKCVNHKNIIGLLNVFTPQKSLEEFQDVYIVMELMDANLCQV IQMELDHERMSYLLYQMLCGIKHLHSAGIIHRDLKPSNIVVKSDCTLKILDFGLARTAGTSFMMTPYVVTRYY RAPEVILGMGYKENVDLWSVGCIMGEMVCHKILFPGRDYIDQWNKVIEQLGTPCPEFMKKLQPTVRTYVE NRPKYAGYSFEKLFPDVLFPADSEHNKLKASQARDLLSKMLVIDASKRISVDEALQHPYINVWYDPSEAEAPP PKIPDKQLDEREHTIEEWKELIYKEVMDLE*	44,908 kDa (41,910)
Lambda phosphatase	none pET13S-A (Spec)	MRYYEKIDGSKYRNIWVVGDLHGCYTNLMNKLDTIGFDNKKDLLISVGDLVDRGAENVECLELITFPWFRAV RGNHEQMMIDGLSERGNVNHWLLNGGGWFFNLDYDKEILAKALAHKADELPLIIELVSKDKKYVICHADYP FDEYEFGKPVDHQQVIWNRERISNSQNGIVKEIKGADTFIFGHTPAVKPLKFANQMYIDTGAVFCGNLTLIQV QGEGA	25,219

Peptides

GRA24 peptides were ordered from LifeTein in lyophilized aliquots of 1-4mg at 98% purity and stored at -80°C (Table 9). The peptides were centrifuged and dissolved into the appropriate buffer before usage and stored in small aliquots at -80°C.

Table 9: Peptides

Name	Sequence	MW (Da)
GRA24 KIM1	GLLERRGVSELPPLYI	1812
MKK6 KIM	SKGKKRNPGLKIPKEA	1751

Plasmid amplification, protein expression in *E. coli*, purification, SDS-PAGE gels, and mass spectrometry experiments were performed as described in Chapter I (p. 51).

ITC

The ITC runs were performed with a MicroCal PEAQ-ITC (Malvern Panalytical) using 200µM GRA24 (syringe), and 20µM of p38/ERK1/JNK1 (cell), T=20°C, 13 x 3µL injections, stirring speed of 750rpm, feedback high.

The p38α and ERK1 data were collected in triplicate and the JNK1 data in duplicate at the initial concentrations (200µM/20µM) indicating non-binding for the GRA24 vs JNK1, and a single measurement at higher concentration (486µM/42µM) for the GRA24 vs JNK1.

The buffer vs buffer, buffer vs GRA24, and p38 or ERK1 vs buffer runs were performed and used as compiled controls.

Data were analysed using the Microcal PEAQ-ITC Analysis software in the single-binding site mode with buffer-buffer and ligand-buffer controls.

X-ray crystallography

Crystallization optimisation

Samples were prepared by mixing a 3:1 molar ratio of GRA24 peptide with ERK1 (at a concentration of ~10mg/ml) and incubating at room temperature for 30 minutes. Sparse matrix crystallization screens were carried out at the High-Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble (Dimasi *et al.*, 2007; Dupeux *et al.*, 2011).

Crystallization plates were setup using the sitting drop method, where drops consisted of 100 nl protein sample and 100 nl reservoir solution. Plates were kept at 20°C and visualised according to an imaging schedule with visible and UV light.

Promising conditions were refined. Refinement plates were prepared using a Formulator (Formulatrix) and sent to the HTX lab for set up.

Crystals were harvested using CrystalDirect technology (Zander *et al.*, 2016).

Data collection

Data were collected at MASSIF1 at the ESRF (Bowler *et al.*, 2015) using automatic protocols for the location and optimal centring of crystals (Svensson *et al.*, 2018). Strategy calculations accounted for flux and crystal volume in the parameter prediction for complete datasets.

Structure determination and refinement

Data were processed with XDS (Kabsch, 2010) and programs from the CCP4 suite (Collaborative Computational Project Number 4) (Winn *et al.*, 2011). The structure was determined by molecular replacement using MolRep. The crystal structure of ERK1 (PDB 2zoq) was used as search model with the peptide residues and all water molecules removed. Refinement was carried out alternately using Phenix (Afonine *et al.*, 2012) and by manual rebuilding with COOT (Emsley and Cowtan, 2004). Tight NCS restraints between monomers in the asymmetric unit were used during refinement. Models were validated using Molprobity (Chen *et al.*, 2010) and figures were produced with PyMol (Schrodinger LLC, 2015).

Cellular assays

Parasites and host cells

Human foreskin fibroblast (HFF) primary cells and ARPE-19 cell line (retinal pigment epithelia (RPE) cell line derived in 1986 by Amy Aotaki-Keen from the normal eyes of a 19-year-old male who died from head trauma in a motor vehicle accident) were cultured in DMEM (Invitrogen) supplemented with 10% heat inactivated FBS (Invitrogen), 10 mM Hepes buffer pH 7.2, 2 mM L-glutamine, and 50 µg/ml penicillin and streptomycin (Invitrogen). Cells were incubated at 37°C in 5% CO2. The *Toxoplasma* strains used in this study were RH*ku80* WT and *RHku80Δgra24* (Braun *et al.*, 2013).

Reagents

Antibodies against Phospho-p38 MAPK (Cell Signaling Technology), Phospho-ERK MAPK (Cell Signaling Technology), EGR1 (Cell Signaling Technology), H4 Acetylated (EMD Millipore) and TgQRS (van Rooyen *et al.*, 2014) were used in Western blotting. Secondary antibodies were conjugated to alkaline phosphatase (Promega).

Cell fractionation

For cytosolic and nuclear fractionation analysis, cells were washed in ice-cold PBS and harvested using a cell scraper. After centrifugation, cells were resuspended in buffer D (10 mM Hepes, pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.34 M Sucrose, 10% Glycerol, 1 mM DTT, and 1× Roche protease inhibitor cocktail) and lysed for 8 min with Triton X-100 at a final concentration of 0.1%. Cell lysates were centrifuged at 1,300 g for 5 min and the supernatant containing the cytosolic fraction was collected. The pellet, containing the nuclear fraction, was washed once in buffer D and resuspended in cell extraction buffer (Invitrogen) for protein extraction. The supernatant and pellet fractions were clarified by centrifugation at 20,000 g for 10 min at 4°C. Supernatants were mixed with protein sample buffer (Invitrogen) for Western blot analysis.

Western blots

Proteins were separated by SDS-PAGE, transferred to a poly-vinylidene fluoride membrane (Immobilon-P; Millipore) by liquid transfer and Western blots were probed using appropriate primary antibodies followed by phosphatase-conjugated goat secondary antibodies (Promega). Signals were detected using NBT-BCIP (Amresco).

Results

Collaborators showed that, in addition to the defined interaction with p38α (Braun *et al.*, 2013; Pellegrini, Palencia, *et al.*, 2017), the *T. gondii* GRA24 effector protein also interacts with ERK1/2, another member of the MAPK family. GRA24 contains two similar KIM motifs, GRA24 KIM1 (GLLERRGVSELPPLYI) and GRA24 KIM2 (GLLERRGVSELPPLRI) through which it interacts with p38α. Our hypothesis is that the GRA24 mode of action is similar with ERK1/2.

Interaction between the GRA24 KIM peptide and ERK1, p38α and JNK1 MAPKs

Three representative members of the MAPK family, ERK1, p38α and JNK1, were expressed in *E. coli* and purified (Figure 56).

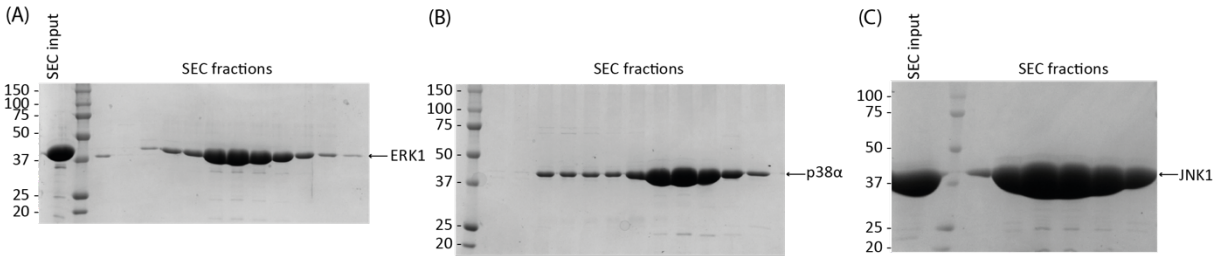


Figure 56: SDS-PAGE gels of purified MAPKs (A) ERK1 (B) p38 (C) JNK1

Isothermal Titration Calorimetry (ITC) experiments were performed to assess the affinity of the GRA24 KIM1 peptide for the 3 MAPKs. The results confirmed high affinity of GRA24 KIM1 peptide for p38α, as previously described (Pellegrini, Palencia, *et al.*, 2017), and for ERK1, confirming the pull down experiments of our collaborators that identified ERK1 as a potential binding partner of GRA24 during infection by *Toxoplasma gondii* (Figure 57). No binding was detected between GRA24 KIM1 and JNK1 (at least within a low- to mid-micromolar range).

Table 10: binding affinity measured by ITC

	K_D (µM)	**sd (µM)**
ERK1 vs GRA24 KIM1	0.31	0.035
p38α vs GRA24 KIM1	1.02	0.163
JNK1 vs GRA24 KIM1	Not detected	

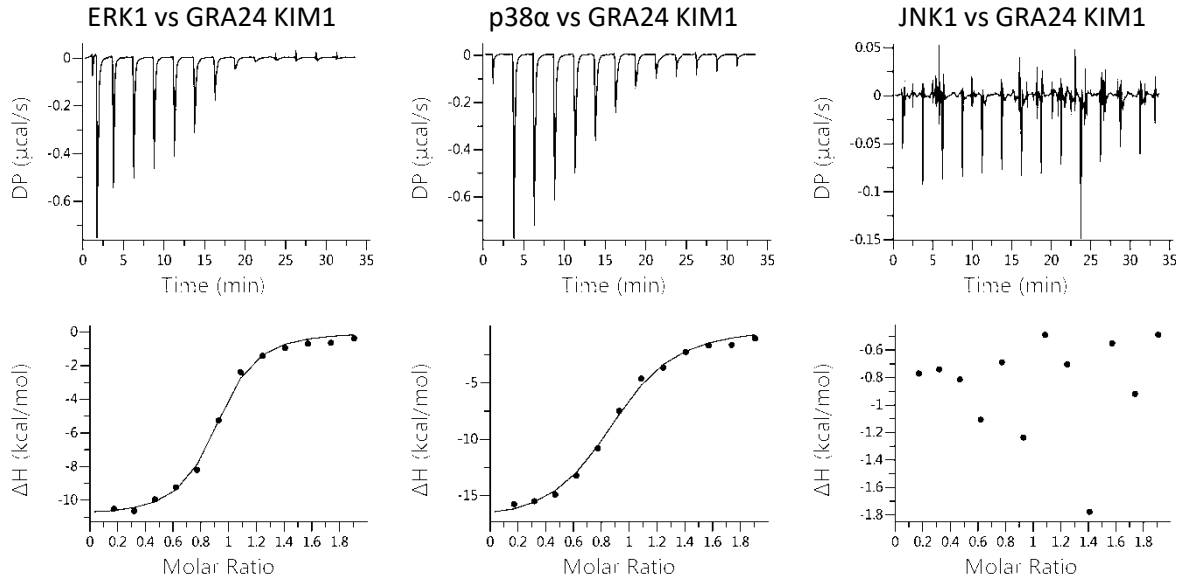


Figure 57: Typical ITC results of binding of GRA24 KIM1 peptide to ERK1, p38α and JNK1.

The GRA24 KIM1 peptide presents a similar mode of binding to ERK1 and p38α, and both show a similar ΔG, pointing to a similar mechanism of binding and conformational changes induced on the MAPK (Figure 58).

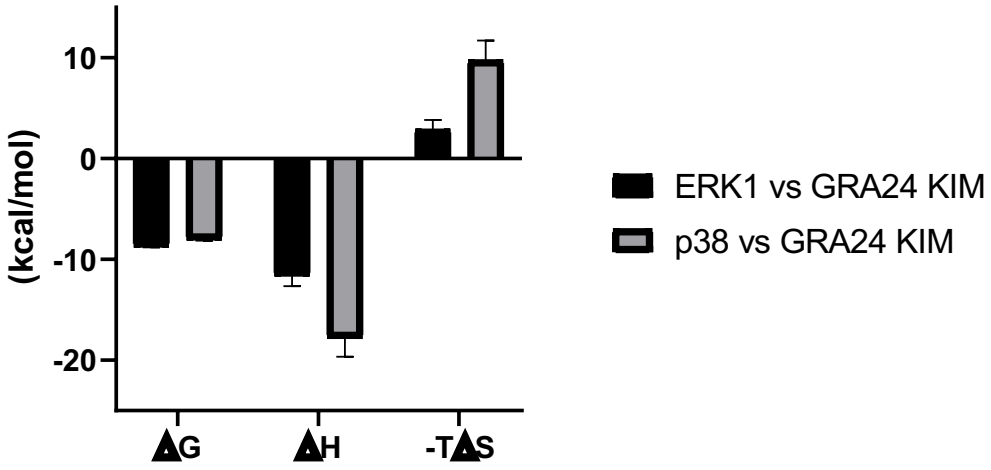


Figure 58: Comparison of ITC results of binding of GRA24 KIM1 peptide to ERK1 and p38α. The plot shows the mean and sd of 3 replicates for each sample.

Crystal structure of ERK1 bound to GRA24 KIM1 peptide

Crystallization optimization

To define the molecular basis of the interaction, we sought conditions for the co-crystallization of ERK1 and GRA24 KIM1 peptide. Sparse matrix screenings of the sample in a 3:1 ratio of peptide:protein was performed using the HTX facility. Initially, samples with or without TCEP were used. Spherulites appeared after a few days in a condition containing 0.1 M MES pH 6.5,

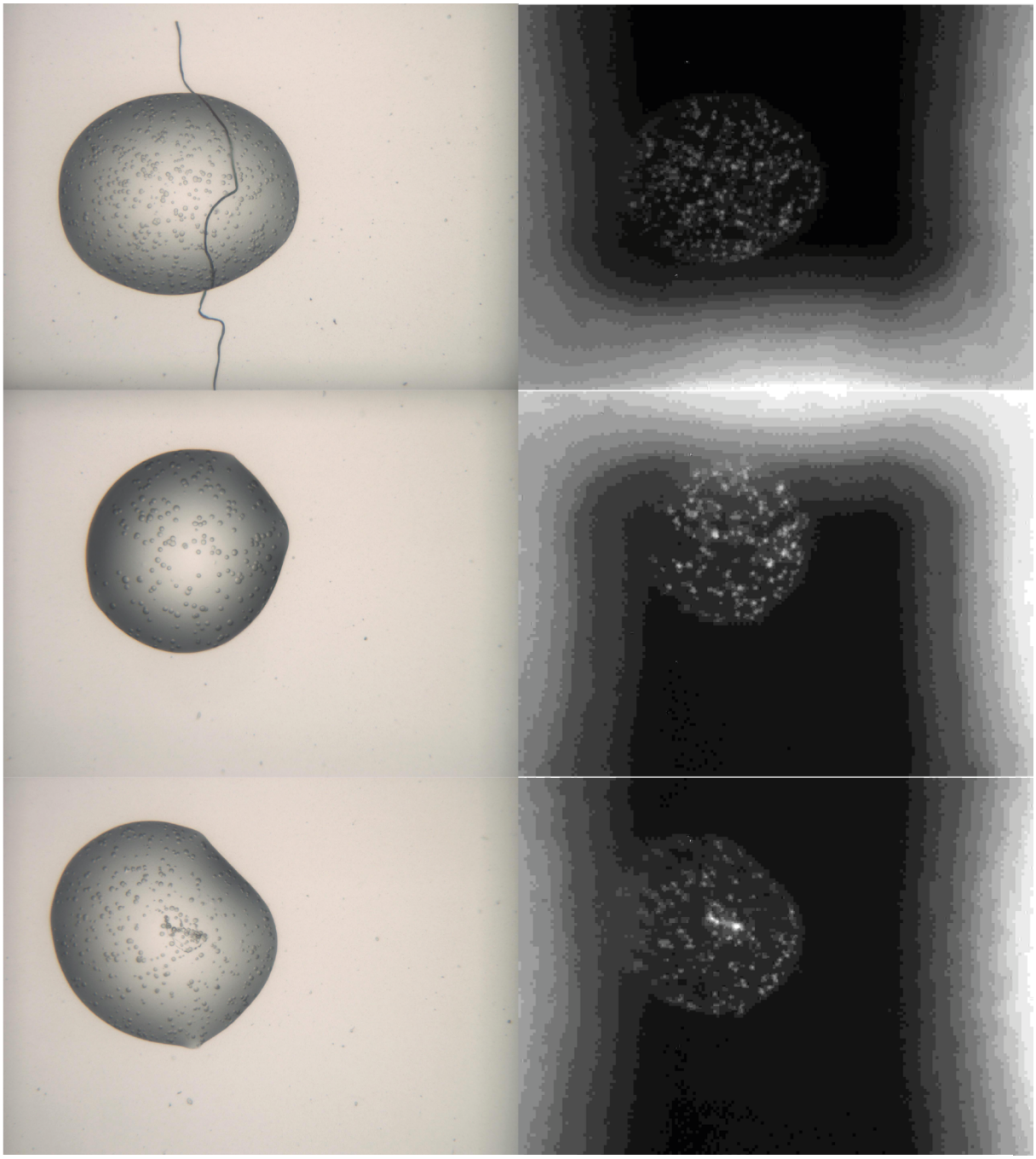


Figure 59: Spherulites of ERK1 with bound GRA24 KIM peptide sample with corresponding UV signal. Sample at 10 mg/ml, setup at 20°C. Crystallization condition: 10% dioxane, 1.6 M ammonium sulfate, 0.1 M MS pH 6.5.

10% dioxane, 1.6 M ammonium sulfate (Figure 59). The spherulites appeared later in the sample containing TCEP than in the ones not containing the reducing agent.

This promising condition was refined in several steps, varying the concentration of dioxane and ammonium sulfate, the pH of the MES buffer, and testing different concentrations of protein. Higher protein concentrations resulted in more crystals. We tried to remove dioxane as it is toxic and difficult to handle, but it was necessary for crystallization. Crystals were harvested and sent to test diffraction at the ESRF beamline MASSIF1.

First, some irregular and poorly diffracting crystals appeared (Figure 60).

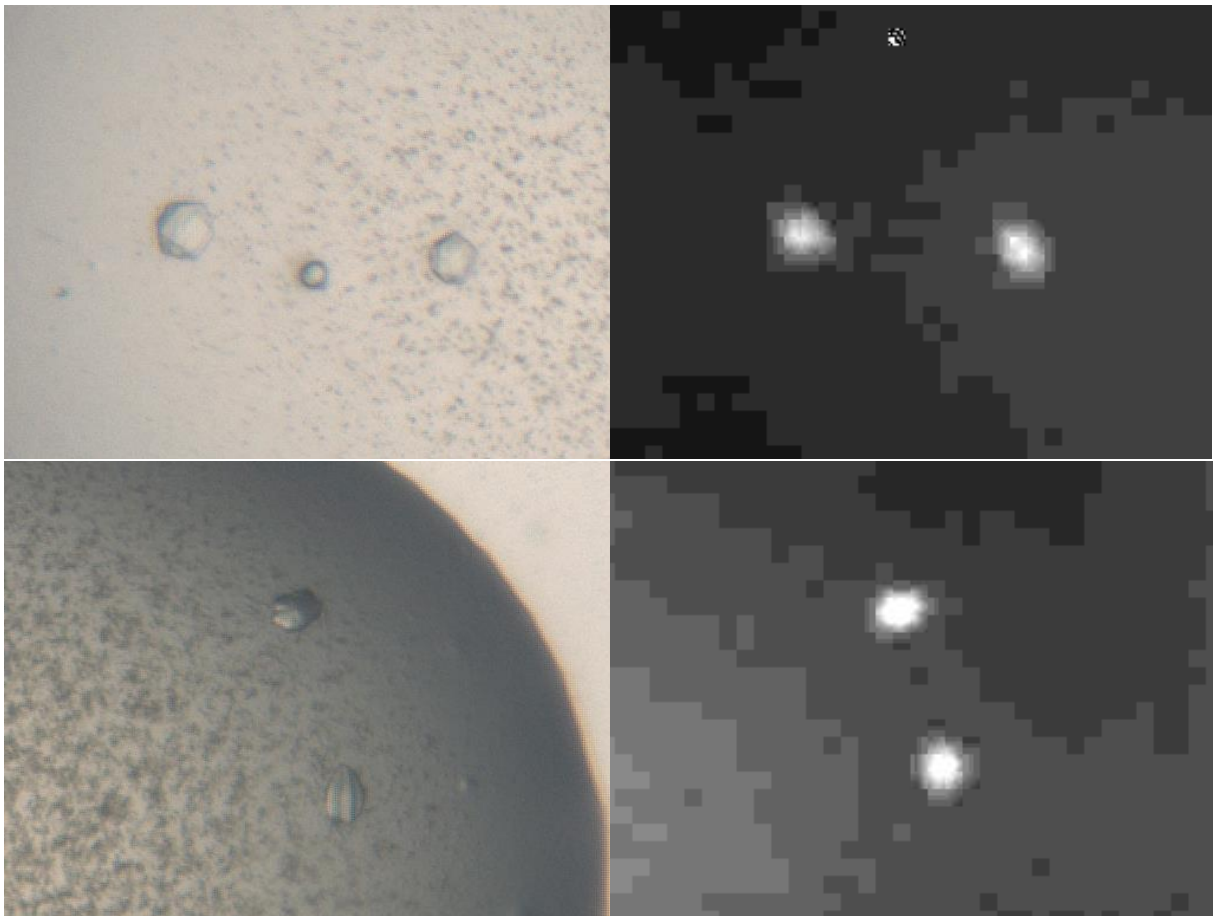


Figure 60: Crystals of ERK1 with bound GRA24 KIM1 peptide sample with corresponding UV signal. Sample at 13.1 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 6.5, 10% dioxane, 1.8 M ammonium sulfate.

With further refinement and higher protein concentration, well-formed hexagonal crystals appeared in several drops (Figure 61).

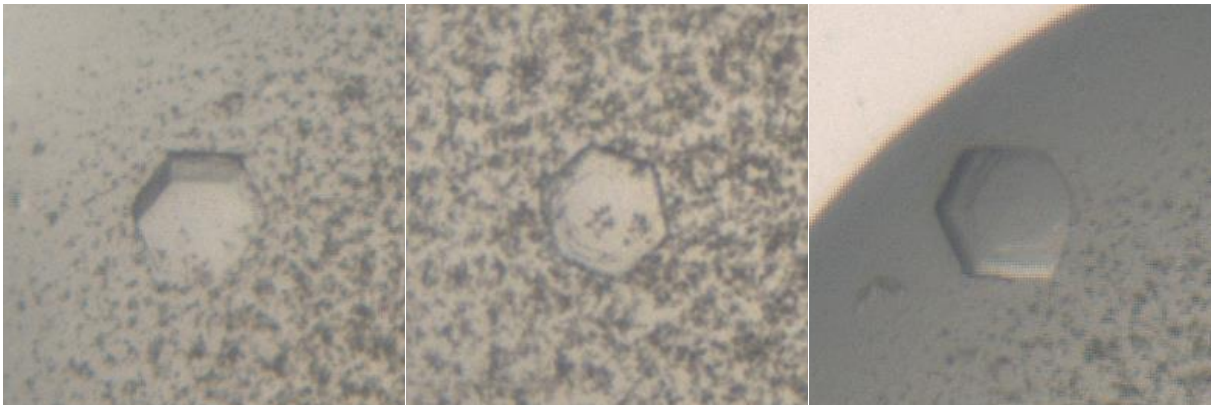


Figure 61: Crystals of ERK1 with bound GRA24 KIM1 peptide sample. Sample at 15 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 6.5-7.5, 10% dioxane, 1.8 M ammonium sulfate.

The best dataset was obtained from a crystal obtained in a condition with 0.1 M MES pH 7, 10% dioxane, 1.8 M ammonium sulfate (Figure 62).

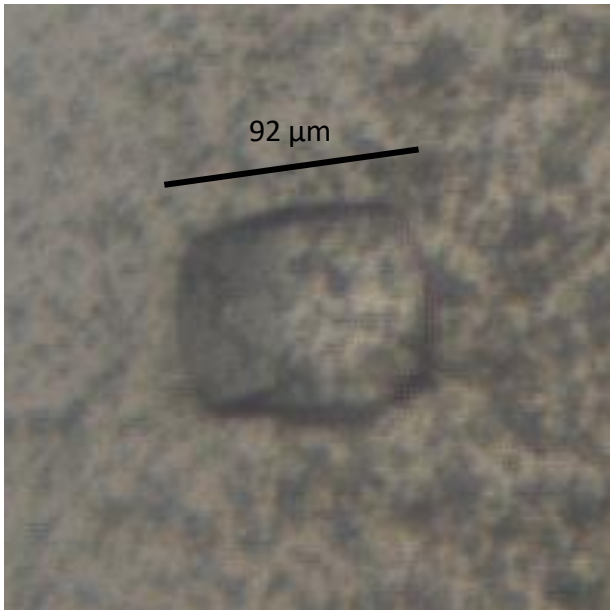


Figure 62: Crystal of ERK1 with bound GRA24 KIM1 peptide that diffracted to a resolution of 2.4 Å. Sample at 17 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 7, 10% dioxane, 1.8 M ammonium sulfate.

Data collection and structure determination

The crystal structure of ERK1 bound to the GRA24 KIM1 peptide was determined at 2.4 Å. Data were collected on beamline MASSIF1 at the ESRF and the phases were calculated by molecular replacement with Phaser using the ERK1 structure PDB 4qtb and refined with a combination of Refmac, Phenix, and Coot (Table 11). The space group determination was not straight forward but was helped by the clear 6-fold symmetry of the crystals (Figure 61). The complex crystallized in the hexagonal space group *P* 65 2 2.

Table 11: Data collection and refinement statistics. Values in parentheses are for the outer resolution shell.

	ERK1 with bound GRA24 KIM1 peptide
Wavelength (Å)	0.96600
Resolution range (Å)	64.78 - 2.40 (2.49 - 2.40)
Space group	*P* 65 2 2
Unit cell **a b c (Å)** **α β γ (°)**	 59.82 59.82 388.70 90 90 120
Total reflections	102849
Unique reflections	17824 (1811)
Multiplicity	5.8 (6.0)
Completeness (%)	99.2 (99.8)
Mean I/sigma(I)	10.5 (1.6)
Wilson B-factor (Å^2)	52.32
R-merge	0.083 (0.913)
R-meas	0.098 (1.10)
R-pim	0.051 (0.579)
CC1/2	0.999 (0.73)
R-work	0.2584 (0.3286)
R-free	0.2823 (0.4162)
Number of non-hydrogen atoms	2929
macromolecules	2869
ligands	20

solvent	40
Protein residues	347
RMS(bonds)	0.003
RMS(angles)	0.58
Ramachandran favored (%)	94.35
Ramachandran allowed (%)	5.06
Ramachandran outliers (%)	0.60
Rotamer outliers (%)	0.00
Clashscore	8.86
Average B-factor	60.57
macromolecules	60.76
ligands	85.00
solvent	34.35
Number of TLS groups	6

Clear density was observed for the peptide at the docking site (Figure 63A). The GRA24 KIM1 peptide binds to the docking site of ERK1, which is comprised of an acidic patch (the CD domain) charged negatively, and a hydrophobic docking groove (Figure 63B). Essential interactions occur at the acidic patch where Arg444 and Arg445 of the peptide are coordinated by Glu98, Asp335 and Asp338 residues of ERK1 (Figure 65A). The C-terminal of the peptide, which comprises leucines (φ_L and φ_A in the KIM nomenclature), prolines and an isoleucine (φ_B), are coordinated by hydrophobic interactions in the docking groove. The Val447 of the peptide, which is not found in any other KIM motif, is sitting in a hydrophobic pocket.

The binding of the GRA24 KIM1 peptide induces a conformational change in comparison to the apo, unphosphorylated and low-activity structure of ERK2 (PDB 5UMO) (Figure 64). The two lobes of the kinase are rotated towards the more closed conformation of the biphosphorylated and active ERK2 (PDB 2ERK). The active conformation of ERK1/2 (PDB 2ERK), as for other MAPKs, is characterized by the dual-phosphorylation of Thr202 and Tyr204 of the TxY motif in the activation loop, maintaining the kinase in an enzymatically active conformation, able to accommodate ATP in its binding site, and to phosphorylate targets. In

the inactive structure of ERK2 (PDB 5UMO), this activation loop is ordered and buried in the C-terminal lobe. In our peptide bound structure, the activation loop is disordered, potentially preparing its TEY motif for phosphorylation. The binding of the GRA24 KIM1 peptide induces allosteric effects that lead ERK1 toward a prone to be activated conformation, similar to that described for the interaction between this same peptide and p38α (Pellegrini, Palencia, *et al.*, 2017).

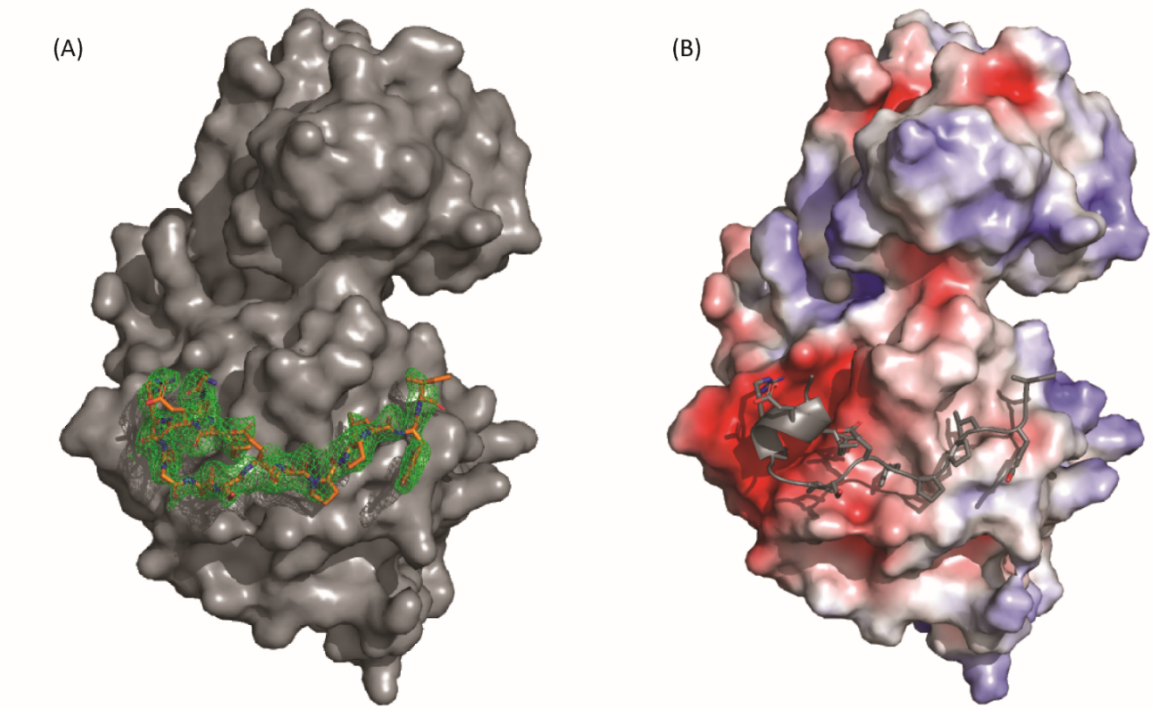


Figure 63: Structure of ERK1 bound to GRA24 KIM1 peptide (A) Simulated annealing omit map electron density (in green) in which the GRA24 KIM1 peptide was built (in orange). (B) The structure of ERK1 is shown as a solvent accessible surface coloured by electrostatic potential (-5 to 5 kTe^{-1}, red to blue) and the GRA24 KIM1 peptide is shown in cartoon representation (in grey).

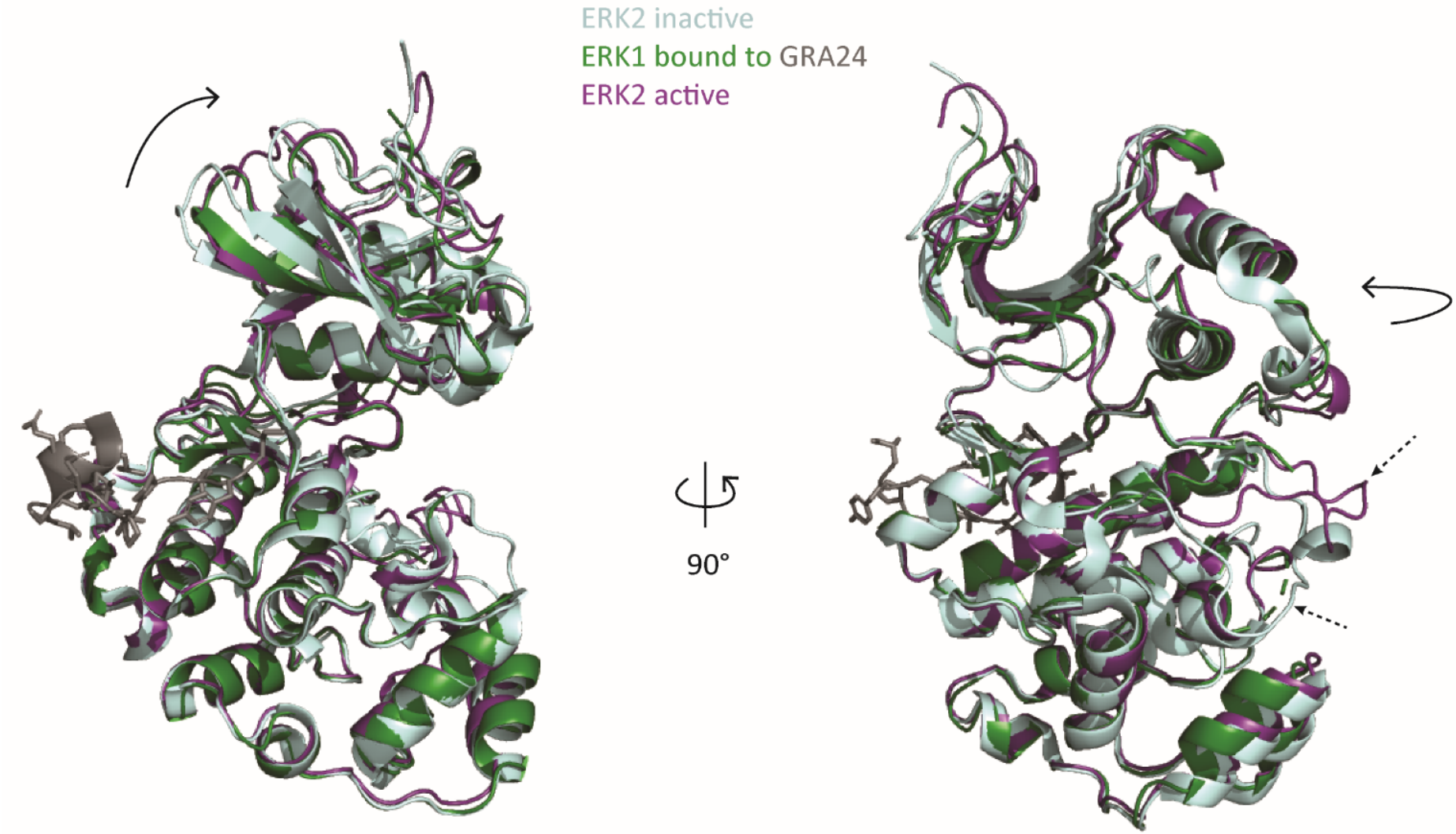


Figure 64: Conformational change induced by GRA24 KIM1 binding. The N-terminal lobe rotates towards the active conformation of the kinase. The activation loop (indicated by dashed arrows) is embedded in the inactive conformation (PDB 5umo), ordered and accessible to solvent in the active conformation (PDB 2erk), and disordered when in complex with GRA24 KIM1 peptide.

Alignment of this structure with other MAPKs structures was based on two core α-helices (residues 140-160 and 224-240) in the C-terminal domain of ERK1.

ITC results showed that the GRA24 KIM1 peptide has a similar affinity for ERK1 as for p38α (Table 10). Comparison between the two structures (with PDB 5eta) shows a similar binding mode, and that the most crucial interactions, especially the asparagines (Asp335 and Aps338 in ERK1 and Asp313 and Asp316 for p38α) and the glutamine (Glu98 in ERK1 and Glu81 in p38α) of the CD domain, are conserved (Figure 65B).

ITC experiments showed no binding of the GRA24 KIM1 peptide to JNK1. Comparison between the docking sites of these two MAPKs explains this absence of binding (Figure 65C). The conserved glutamine in ERK1 and p38α is a lysine (Lys83) in JNK1, this charge inversion would prevent the binding of the arginines of the GRA24 KIM1 peptide. The CD domain in JNK1 (with Asp326 and Glu329) is also in a different conformation than ERK1 CD domain, and the volume of the hydrophobic pocket where Val447 of the GRA24 KIM1 peptide sits in ERK1 is smaller in JNK1. These elements explain the absence of binding of GRA24 KIM1 peptide to JNK1. A different binding mode of KIM peptides is observed in the hydrophobic groove when comparing with JNK1 bound to MKK7 KIM peptide, even if the canonical KIM sequence is similar in this peptide region.

The structure of the docking of *Toxoplasma gondii* GRA24 KIM1 peptide to ERK1 gives insights to the specificity of MAPKs docking sites. It confirms the detected interaction between GRA24 and ERK1 from cellular pull downs. The binding mode is similar to that observed with p38α.

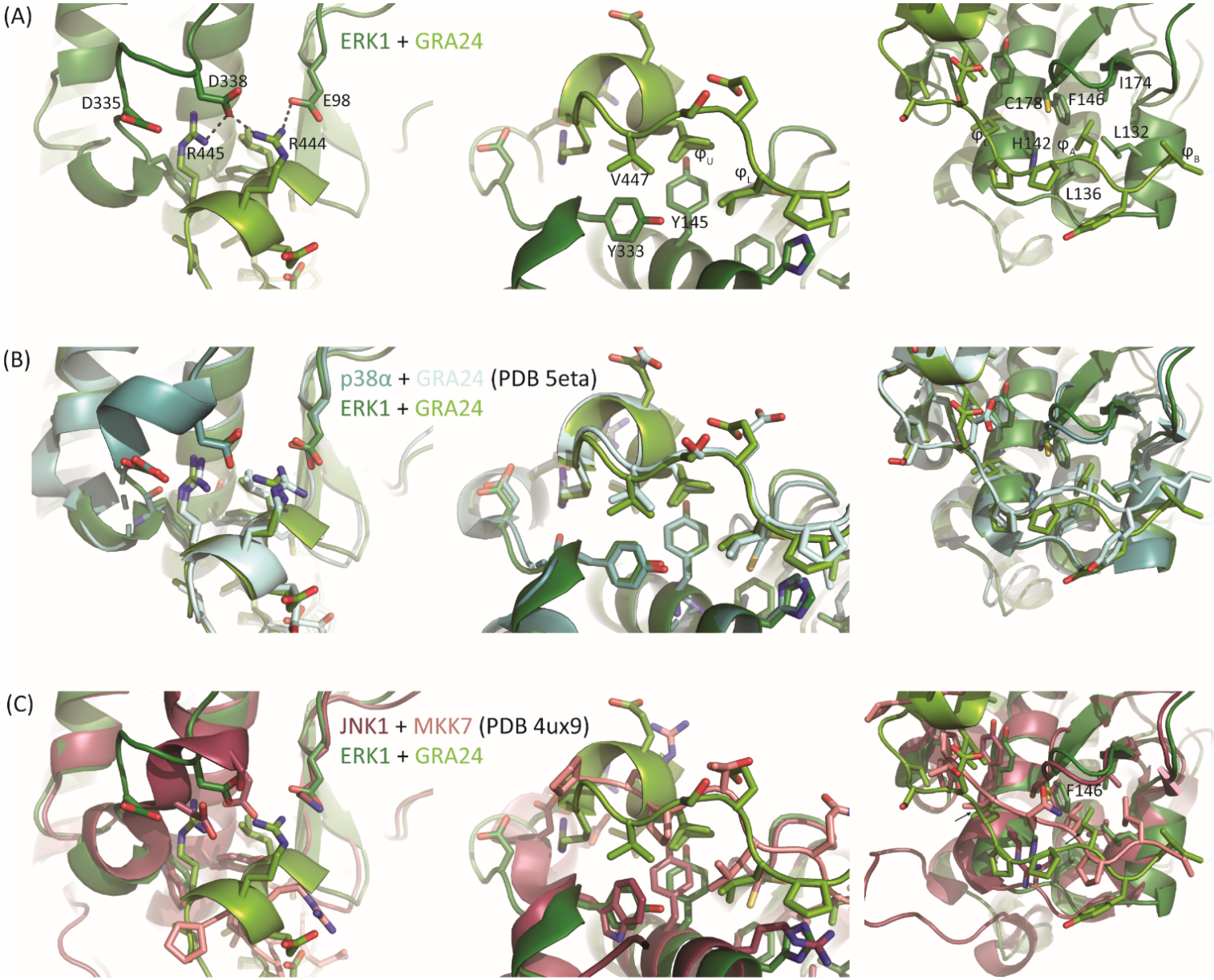


Figure 65: Comparison of the binding mode of KIM peptide to the docking site of ERK1, p38α (PDB 5eta) and JNK1 (PDB 4ux9). (A) GRA24 KIM peptide binds to the docking site of ERK1 through hydrogen bonds with the CD domain and through hydrophobic interactions. (B) GRA24 KIM peptide binds similarly to the conserved docking site of ERK1 and p38α. (C) ERK1 and JNK1 present different docking site conformations, explaining the different binding modes of KIM peptides.

Does GRA24 induces ERK1/2 phosphorylation and activation during host cell infection by *T. gondii*?

In collaboration with Ali Hakimi's group (IAB, Grenoble), cellular assays were performed to study the mechanism of this interaction (Figure 66). Two different human cell lines, human

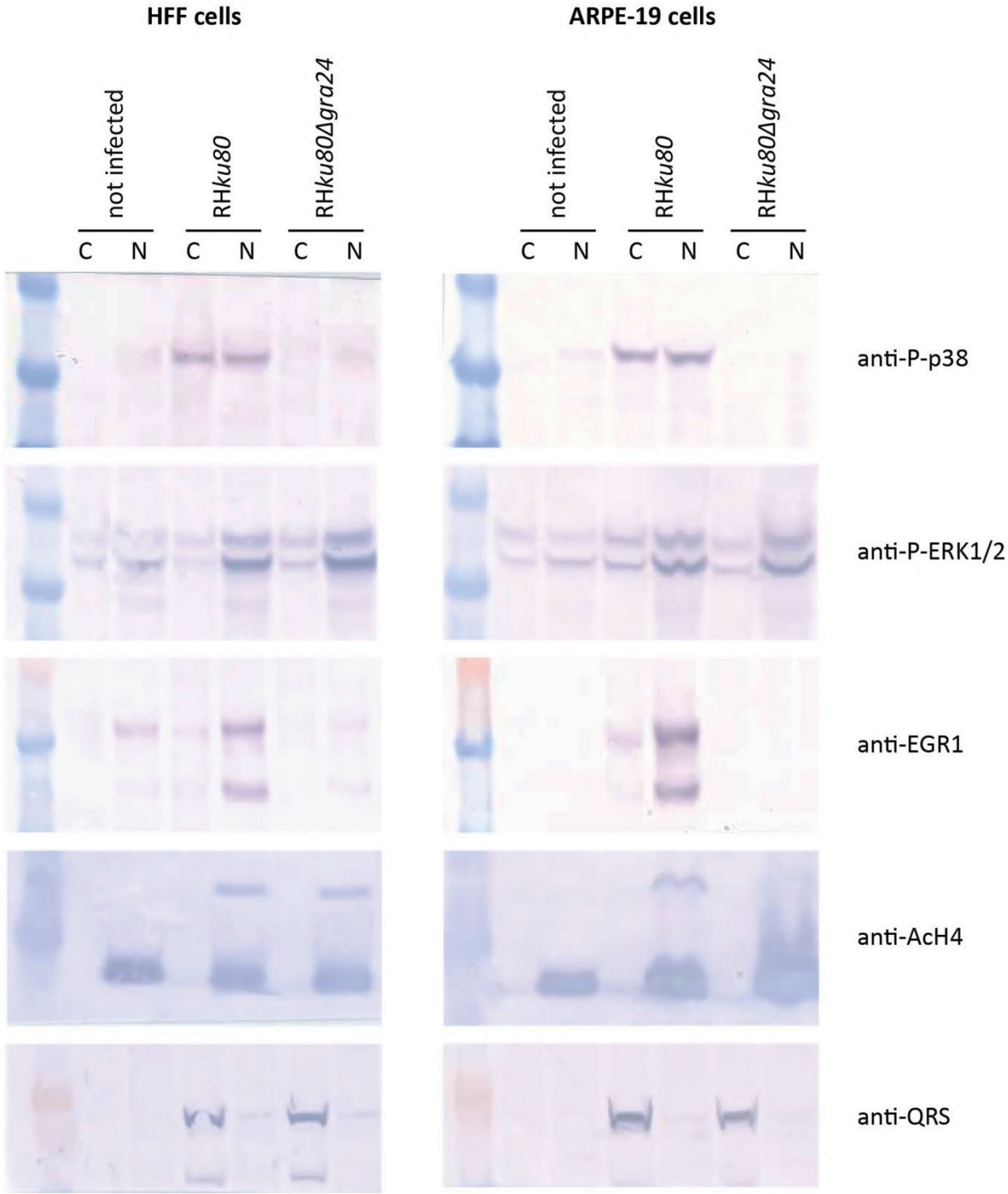


Figure 66: Human cell HFF and ARP19 were infected with T. gondii RHku80 or RHku80Δgra24. Cytosolic and nuclear extracts were blotted with phospho-antibodies of ERK1/2 and p38, and controls (anti-AcH4 as nuclear marker and anti-QRS as infection marker).

foreskin fibroblasts (HFF) and retinal pigment epithelial cells (ARPE-19) were infected by *T. gondii* RH*ku80* or RH*ku80*Δ*gra24.* As previously published, we observed that p38α is phosphorylated in cells infected with *T. gondii* RH*ku80*, but not when infected with the GRA24 knock-out parasite RH*ku80*Δ*gra24*, and that the p38α phosphorylation dependent of GRA24 triggers EGR1 activation (Pellegrini, Palencia, *et al.*, 2017).

We observed that ERK1/2 phosphorylation is also triggered by *T. gondii* infection, but this activation is independent of GRA24, as phosphorylated levels of ERK1/2 are similar in the infected cells with both *T. gondii* RH*ku80* or RH*ku80*Δ*gra24* strains.

In conclusion, despite the strong interaction between GRA24 KIM1 peptide and ERK1, GRA24 is not the *T. gondii* effector responsible for ERK1/2 phosphorylation during parasite infection. However, GRA24 could still have an effect on ERK1/2 different than its activation by phosphorylation, and further study is required to explore the consequences of the GRA24-ERK1/2 interaction during *T. gondii* infection.

Discussion and perspectives

The GRA24 KIM1 peptide binds to ERK1 docking site in a similar fashion to p38α

We showed with *in vitro* experiments that GRA24 binds through its KIM domain to ERK1 in a similar fashion to p38α. The affinity of the GRA24 KIM1 peptide for ERK1 and p38α is similar, and higher than most other MAPK partners. The crystal structure of ERK1 bound to GRA24 KIM1 peptide was elucidated and revealed a similar binding mechanism of the KIM peptide.

Although the study of the structures of p38α (PDB 5ETA) and ERK1 bound to the GRA24 KIM1 peptide explains the binding of GRA24 to these two classes of MAPKs and not to the JNK class, it reinforces the fact that the promiscuity of the binding sites between MAPKs and the KIM motifs definition are not enough to explains the discrimination between MAPKs binding partners, and raises the possibility of unidentified supplementary interaction domains.

GRA24 does not influence ERK1/2 phosphorylation during *T. gondii* infection

T. gondii GRA24 interacts with ERK1/2 in a cellular environment but the phosphorylation of ERK1/2 observed during *T. gondii* infection is independent of the presence of GRA24. Therefore, GRA24 does not promote the auto-phosphorylation of ERK1/2 as previously described for the GRA24-p38α interaction as we initially hypothesized. With these results, we cannot conclude whether the GRA24-ERK1/2 interaction is significant *in vivo* or if GRA24 only binds to ERK1/2 because of the promiscuity between the two MAPKs docking sites.

The same conclusion was recently reported by (Mercer *et al.*, 2020) that observed that GRA24 does not influence ERK1/2 phosphorylation.

The *T. gondii* ERK1/2 activator(s), as well as the JNK one(s), are still to be identified.

Does the interaction between GRA24 and ERK1/2 during *T. gondii* infection have an effect on ERK1/2 signalling?

We can speculate that the interaction between GRA24 and ERK1/2 during parasite infection is not without consequences. Even though GRA24 does not influence the ERK1/2 phosphorylation state, it could still affect ERK1/2 signalling through alternative mechanisms, for example it might influence its cellular localization. However, the assay presented in Figure 66 does not display a difference in cytosolic or nuclear localization of phosphorylated ERK1/2 dependent on GRA24. This hypothesis still remains possible and might be time-dependent.

GRA24 could inhibit phosphorylated ERK1/2 activity by binding competition to the MAPK docking site. It could also block phosphorylated ERK1/2 in an inactive conformation, as has

been observed with the scaffold protein PEA-15 (Mace *et al.*, 2013). We could also imagine that GRA24 could influence ERK1/2 affinity for substrates, orienting the cellular response.

We will proceed with this project by analysing the influence of GRA24 secretion on the host cell phospho-proteome to unravel effects of GRA24 besides p38α auto-phosphorylation induction. We hope that the detection of changes dependent on GRA24 in the phosphorylation state of MAPKs and their substrates will give us insights on the extent of GRA24 effects. The use of specific ERK1/2 inhibitors and cell knockout lines could also help to confirm whether the GRA24 interaction with ERK1/2 is relevant *in vivo*.

The ERK1/2 pathway has a crucial in role in cell proliferation, survival, growth, metabolism, migration and differentiation (Lavoie, Gagnon and Therrien, 2020), and these cellular mechanisms are highly remodelled by *T. gondii* during infection to achieve an equilibrium between host survival and parasite dissemination to achieve long-term survival.

T. gondii triggers the phosphorylation of different members of the MAPK network. The mechanism of p38α has been deciphered, but how ERK1/2 and JNK are activated during infection is still unknown. Current efforts are focused on identifying *T. gondii* proteins containing potential KIM motifs that could be candidates. However, the defined motifs give a lot of hits, and the strategy to filter and select them is not straight forward. Based on experience, potential *T. gondii* effectors interacting with MAPKs should contain one or several KIM motifs, present in intrinsically disordered regions of exported proteins. Whether its sequence is conserved or not among closely-related parasite species such as *Neospora caninum* is open for debate. The case of GRA24 would orientate the strategy to look into *T. gondii* specific sequences. But other apicomplexans also trigger MAPK activation, and some of the responsible effectors might be conserved (Pease *et al.*, 2013; Mota *et al.*, 2016; Li *et al.*, 2019).

During infection, *T. gondii* remodels the phospho-proteome of the cells through the action of different effector proteins. Studying their individual effects is only giving us a partial picture of what is happening during infection. It is likely that their effects are combinatory and tightly regulated by an equilibrium of the different effectors. It is also likely that the expression, release and effects of *T. gondii* interactors are dependent of the type of the infected cell, and on the developmental stage of the parasite.

Chapter IV: Interactions between *T. gondii* GRA16 and human USP7

Introduction

GRA16 is one of the effectors secreted by *T. gondii* during host cell infection and was the first GRA protein described to be exported to the host cytoplasm (Bougdour *et al.*, 2013).

GRA16 is an intrinsically disordered protein of 55 kDa that contains an N-terminal SP, five putative NLS and two internal repeats. (Figure 67) (Dosztányi *et al.*, 2005; Dosztányi, Mészáros and Simon, 2009). GRA16 accumulates in the parasite and the PV, and can also cross the PVM and shuttle to the host cell nucleus. It upregulates the expression of host genes involved in cell-cycle control, metabolism, and the p53 tumour suppressor pathway.

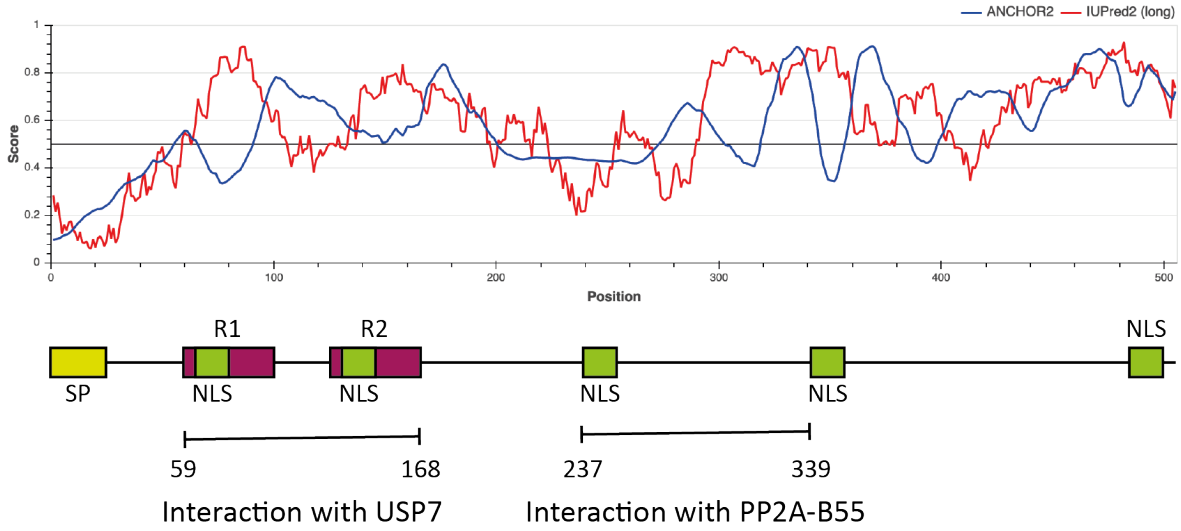


Figure 67: Diagram of GRA16 domains (SP signal peptide, NLS nuclear localization signal, R1 and R2 repeats 1 and 2) with corresponding IUPred2 (intrinsically disordered regions prediction) and ANCHOR2 (prediction binding regions that can undergo disorder-to-order transition upon binding) scores. The two regions identified by Bougdour e al. (2013) to interact with USP7 and PP2A-B55 are indicated.

GRA16 is found in a high-molecular-weight complex involving USP7 (Ubiquitin-specific protease 7, also known as the herpesvirus associated ubiquitin-specific protease HAUSP) and the phosphatase PP2A-B55 (Bougdour *et al.*, 2013).

USP7 is a regulator of the stability of the tumour suppressor protein p53 (Figure 68) (Pozhidaeva and Bezsonova, 2019). Depending on the cellular context, USP7 can promote both the stabilization of p53 by direct deubiquitination, and p53 degradation through the stabilization of MDM2. The levels of p53 depend on the equilibrium between these different mechanisms, and the consequences of USP7 interference are not straightforward.

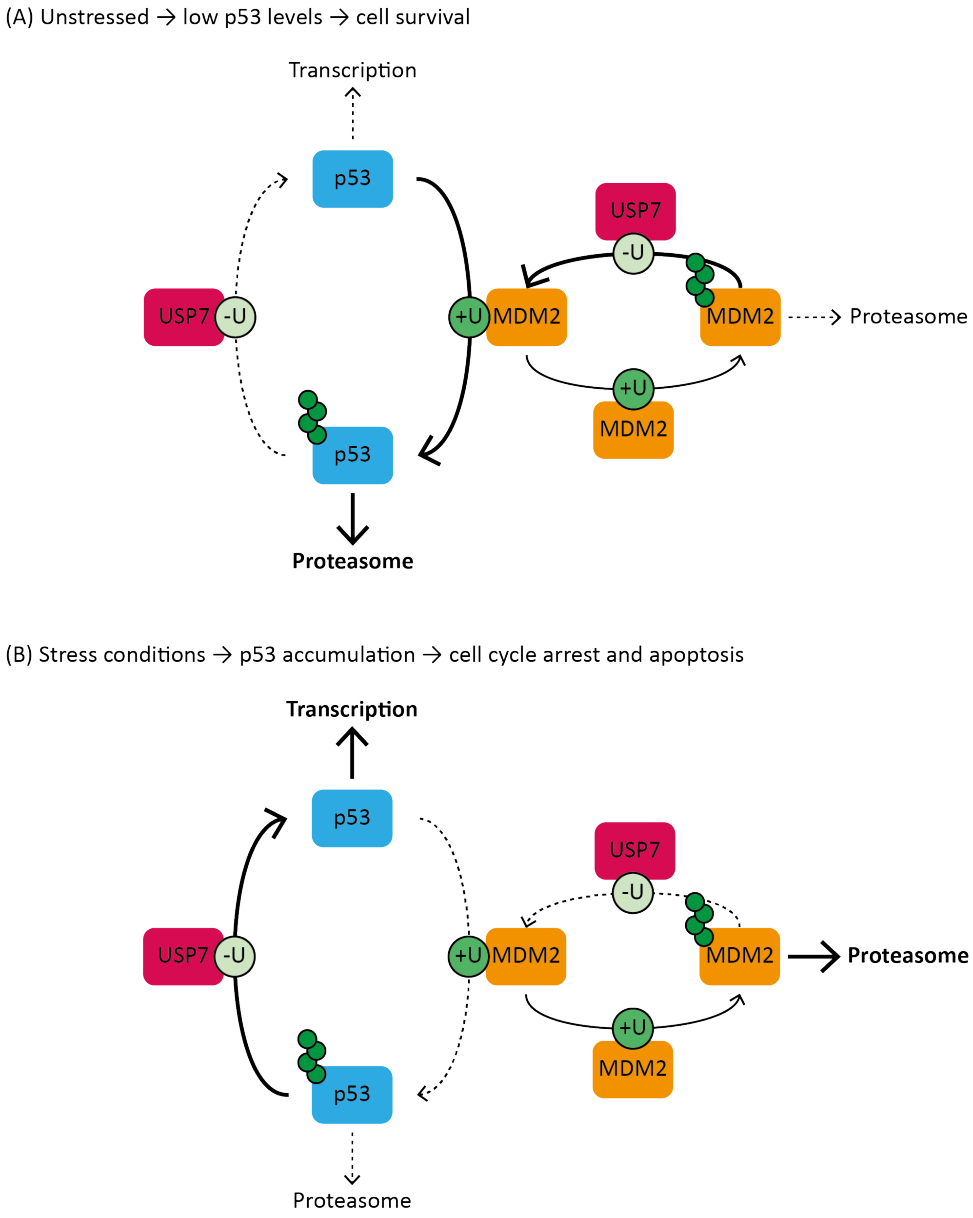


Figure 68: The USP7-MDM2-p53 axis (A) In normal cells, p53 has a short half-life. It is maintained at low levels by an equilibrium between ubiquitination by the E3 ligase Mdm2 and subsequent 26S proteolysis, and deubiquitination by USP7 that promotes its stabilization. Mdm2 also self-ubiquitinates leading to its degradation but can be rescued and stabilized by USP7 deubiquitination. (B) Under stress conditions, USP7 affinity for Mdm2 is lower, resulting in Mdm2 degradation and subsequent p53 accumulation. p53 acts as a transcription factor and induces cell cycle arrest and apoptosis.

USP7 is composed of three domains (Figure 69):

- a TRAF-like domain at its N-terminus
- a catalytic core that holds the deubiquitinase activity
- a C-terminal domain composed of five Ubiquitin-like domains (Ubl)

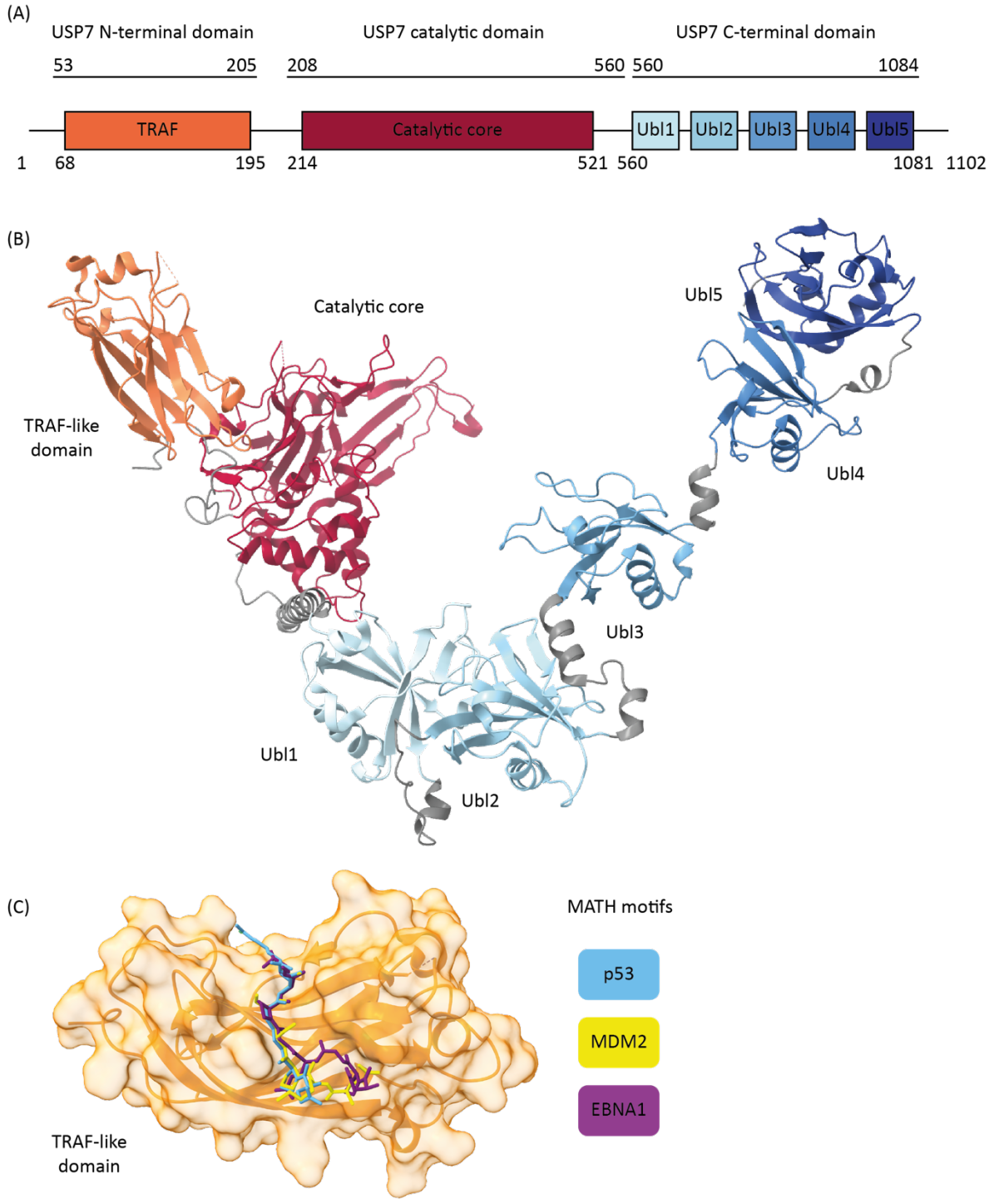


Figure 69: USP7 domains and structures (A) USP7 domains (B) Model of the USP7 structure from combination of PDB 2f1z, 2ylm and 4wph (C) USP7 TRAF-like domain bound to MATH peptides from human interacting partners p53 (PDB 2foo) and MDM2 (PDB 2fop), and viral interacting partners EBNA1 (PDB 1yy6)

The TRAF-like domain binds to MATH motifs located in interacting partners such as p53 and MDM2, but is also targeted by viral proteins such as EBNA1 (Epstein–Barr nuclear antigen 1) (Figure 69C) (Saridakis *et al.*, 2005; Hu *et al.*, 2006; Sheng *et al.*, 2006).

The effects of GRA16 on USP7 are cryptic, but GRA16 could participate in the alteration of the equilibrium levels of p53, potentially affecting the host cell response to stress caused by intracellular infection. The interaction with USP7 could play a role in the evasion from the host immune response, as well as infected host cell immortalization.

GRA16 was recently described to be at the origin of c-Myc upregulation in *T. gondii* infected cells (Panas and Boothroyd, 2020). The stability of c-Myc depends on the phosphorylation of key residues. Phosphatases, including PP2A-B56α, can trigger its degradation by dephosphorylating it (Seshacharyulu *et al.*, 2013). One hypothesis on how GRA16 induces c-Myc accumulation is that GRA16 would bind and stabilize one specific composition of the PP2A complex, containing B55 subunits, and excluding B56α subunits, and therefore would prevent the B56α-dependent degradation of c-Myc.

The effect of GRA16 on hepatocellular carcinoma (HCC) cells have been studied and led to the interesting conclusion that GRA16 inhibits USP7 and possesses p53-dependent anticancer properties (Kim *et al.*, 2019). In this study, USP7 inhibition mediated by GRA16 led to increased levels of the tumour suppressor PTEN (phosphatase and tensin homolog).

This pathway is also targeted by virus infection, as illustrated by the Epstein-Barr virus protein EBNA1, which sequesters USP7 from p53 and leads to its degradation (Saridakis *et al.*, 2005).

The mechanism of interaction between GRA16 and USP7, as well as the consequences of GRA16 binding to the host deubiquitinase, remain unknown.

The aim of this project was to characterize the interaction between *T. gondii* GRA16 and the human deubiquitinase USP7.

The strategy to achieve this goal was to identify interacting regions between *T. gondii* GRA16 and human USP7 and characterize the resulting pairs of interacting GRA16 peptides and USP7 domains using biophysical and structural methods.

We show that the interaction between the two proteins is mediated by the binding of a MATH2 motif in GRA16 to the N-terminal TRAF-like domain of USP7.

Materials and methods

Protein expression and purification

Recombinant protein constructs

USP7 refers to the Ubiquitin-specific protease 7 isoform 1 (also known as HAUSP for Herpesvirus-associated ubiquitin-specific protease) (Uniprot UBP7_HUMAN, Q93009-1).

FL recombinant human USP7 was ordered from Sino Biologicals (ref 11681-H20B1). A DNA sequence encoding full length of human USP7 (NP_003461.2) (Met 1-Asn 1102) was fused with the N-terminal polyhistidine-tagged GST tag at the N-terminus and expressed in insect cells.

Three plasmids encoding the three domains of human USP7 fused to a Twin-Strep tag and 3C protease cleavage site for expression in *E. coli* were ordered from Genscript (gene synthesis and cloning in expression vector):

- USP7 N-terminal domain (53-206)
- USP7 catalytic domain (208-560)
- USP7 C-terminal domain (560-1084)

Recombinant protein constructs used in this chapter are presented in Table 12.

Table 12: USP7 constructs

Protein	**Construct**	**Protein sequence** **(tag and protease recognition cleavage site in bold and cleavage site indicated by ↕)**	**MW** **(cleaved protein MW)**
USP7 N-ter domain (53-206)	N-ter Twin strep tag + 3C cleavage site pET-28a (Kan)	MGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKLEVLFQGPNTAEEDMEDDTSWRSEATFQFTVERFSRLSES VLSPPCFVRNLPWKIMVMPRFYPDRPHQKSVGFFLQCNAESDSTSWSCHAQAVLKIINYRDDEKSFSRRISHLFF HKENDWGFSNFMAWSEVTDPEKGFIDDDKVTFEVFVQADAPHGVAWD*	22,159 kDa (18,213)
USP7 catalytical domain (208-560)	N-ter Twin strep tag + 3C cleavage site pET-28a (Kan)	MGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKLEVLFQGPKKHTGYVGLKNQGATCYMNSLLQTLFFTNQL RKAVYMMPTEGDDSSKSVPLALQRVFYELQHSDKPVGTKKLTKSFGWETLDSFMQHDVQELCRVLLDNVENK MKGTCVEGTIPKLFRGKMVSYIQCKEVDYRSDRREDYYDIQLSIKGKKNIFESFVDYVAVEQLDGDNKYDAGEHG LQEAEKGVKFLTLPPVLHLQLMRFMYDPQTDQNIKINDRFEFPEQLPLDEFLQKTDPKDPANYILHAVLVHSGDN HGGHYVVYLNPKGDGKWCKFDDDVVSRCTKEEAIEHNYGGHDDDLSVRHCTNAYMLVYIRESKLSEVLQAVTD HDIPQQLVERLQEEKRIEAQKRKERQE*	45,044 kDa (41,098)
USP7 C-ter domain (560-1084)	N-ter Twin strep tag + 3C cleavage site pET-28a (Kan)	MGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKLEVLFQGPEAHLYMQVQIVAEDQFCGHQGNDMYDEEK VKYTVFKVLKNSSLAEFVQSLSQTMGFPQDQIRLWPMQARSNGTKRPAMLDNEADGNKTMIELSDNENPWTI FLETVDPLAASGATLPKFDKDHDVMLFLKMYDPKTRSLNYCGHIYTPISCKIRDLLPVMCDRAGFIQDTSLILYEEV KPNLTERIQDYDVSLDKALDELMDGDIIVFQKDDPENDNSELPTAKEYFRDLYHRVDVIFCDKTIPNDPGFVVTLS NRMNYFQVAKTVAQRLNTDPMLLQFFKSQGYRDGPGNPLRHNYEGTLRDLLQFFKPRQPKKLYYQQLKMKITD FENRRSFKCIWLNSQFREEEITLYPDKHGCVRDLLEECKKAVELGEKASGKLRLLEIVSYKIIGVHQEDELLECLSPAT SRTFRIEEIPLDQVDIDKENEMLVTVAHFHKEVFGTFGIPFLLRIHQGEHFREVMKRIQSLLDIQEKEFEKFKFAIVM MGRHQYINEDEYEVNLKDFEPQPGNMSHPRPWLGLDHFNK*	65,538 kDa (61,592)

Peptides

GRA16 peptides were ordered from LifeTein in lyophilized aliquots of 1-4mg at 98% purity and stored at -80°C (Table 13). The peptides were centrifuged and dissolved into the appropriate buffer before usage and stored in small aliquots at -80°C.

Table 13: GRA16 peptides. When the strain is not specified, the peptide sequence is conserved amongst the 3 strains TGGT1, TGME49 and TGVEG.

Name	Sequence	MW (Da)
GRA16_TGGT1_100-118	GFGATVGRVATPRIRSGVV	1900
GRA16_TGGT1_104-114	TVGRVATPRIR	1225
GRA16_TGGT1_129-138	RRPGEVESTL	1143
GRA16_192-206	SSWDSDPLEGTSRDW	1738
GRA16_196-206	SDPLEGTSRDW	1262
GRA16_TGGT1_192-210	SSWDSDPLEGTSRDWQYVP	2225
GRA16_TGME49/TGVEG_192-210	SSWDSDPLEGTSRDWLYVT	2214
GRA16_TGGT1_196-210	SDPLEGTSRDWQYVP	1750
GRA16_TGGT1_220-254	LTGLGGIGRKFAPLYVRDRKFDLLQFVNLTRSKKQ	4037
GRA16_TGGT1_256-270	LLMSSKSPSLRRLLM	1732

Plasmid amplification, protein expression in *E. coli*, SDS-PAGE gels, and mass spectrometry experiments were performed as described in Chapter I (p. 51).

Protein purification

Cells were collected by centrifugation at 4000 g for 15 minutes and resuspended into 100 ml lysis buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5 % glycerol, 0.5 mM TCEP, with a Pierce protease inhibitor EDTA-free tablet (Thermo Scientific) and a trace of DNaseI (Sigma)) for every litre of culture. The solution was passed 3 times through a microfluidizer (Hyland Scientific, M-110L) running at a pressure of 18000 psi for lysis. Lysate was centrifuged at 40 000 g for an hour. The supernatant was loaded onto a pre-packed 5 ml StrepTactin XT

column (IBA), equilibrated according to the supplier's protocols with wash buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5 % glycerol, 0.5 mM TCEP), at a flow rate of 1 ml/min and washed until UV signal returned to zero. Tagged-protein was eluted with 50 ml elution buffer (wash buffer with 50 mM Biotin). SDS PAGE gel was run to identify fractions containing protein, which were pooled and quantified using a NanoDrop ND1000 to measure absorption at 280 nm. GST 3C protease was added to the sample (1 mg for every 100 mg of protein) in the pooled fractions. The sample was dialysed overnight at 4 °C against wash buffer.

After dialysis, the sample was run through the StrepTactin XT column again and flow-through fractions were collected. SDS PAGE was run, and protein containing fractions pooled and quantified. The sample was then concentrated down to 250 µl 10 kDa Amicon membrane filters concentrators (Merck) and loaded onto a size-exclusion Superdex 75 10/300 GL column (GE Healthcare) equilibrated in wash buffer beforehand and running at 0.3 ml/min. SDS PAGE was used to assess purity and identify protein containing fractions, which were pooled and quantified as before. The pooled fractions were concentrated again to reach the desired concentration (50 mg/ml for crystallisation and approximately 1 mg/ml for ITC).

Pure protein samples were either used fresh or stored at -80°C.

Peptide microarrays

JPT Peptide Technologies' PepStar™ peptide microarrays (including peptide synthesis, array printing, labeling of protein with Dylight 650, array incubation, and data analysis) were ordered

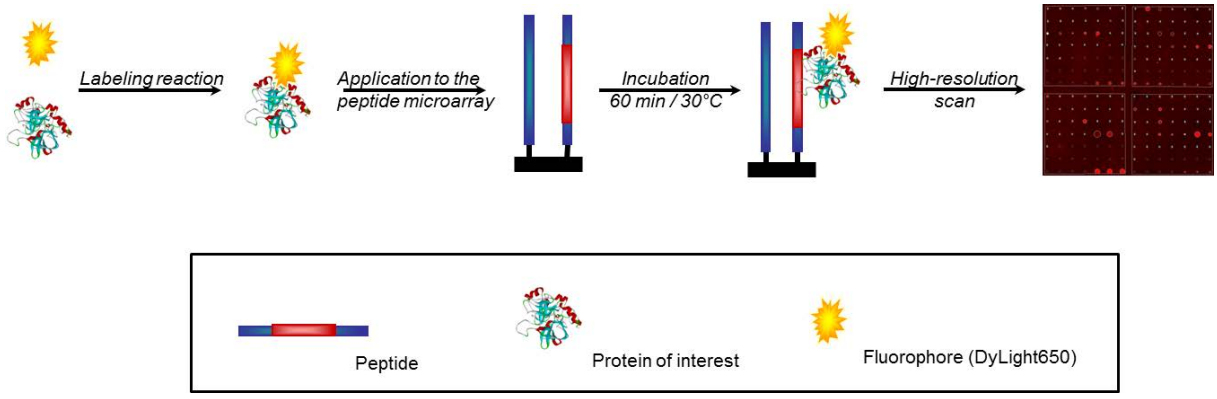


Figure 70: Assay principle. JPT's Peptide Microarrays were designed to elucidate protein-protein interactions. Each spot in the microarray represents a single individual peptide. After a direct labelling of the protein of interest and its incubation on the peptide microarray, the latter can directly be detected by measuring fluorescent signals. Resulting binding signatures represent the unique insights into the individual binding events.

from JPT. The assay principle used in this protein-peptide binding study is schematically shown in Figure 70.

118 peptides (15 aa length, 3 aa sequence overlap) covering the GRA16_TGGT1 sequence (aa 24-505) were synthesized and immobilized on a microarray slide. FL recombinant human USP7 (100µg, lyophilized) was sent to the company. Direct protein labelling was performed using DyLight Microscale Antibody Labelling Kit (Thermo Scientific 84536) according to manufacturer's specifications. Briefly, the sample was diluted in Labelling Buffer (0.05 M Borate buffer in PBS).

The final concentration of the diluted samples was 0.5 mg/ml. 0.1 ml of the sample solutions were added to vials containing DyLight Reagent, mixed gently and incubated for 60 min at room temperature in the dark. Labelled samples were purified using spin columns filled with purification resin. After labelling, serial dilutions of the labelled proteins prepared in Blocking Buffer were incubated on the corresponding peptide microarrays for 1 hour at 30°C. After washing and drying, the slides were scanned with a high-resolution laser scanner at 635 nm to obtain fluorescence intensity profiles. Images were quantified to yield a mean pixel value for each peptide.

To visualize obtained results, heatmap diagrams were computed showing fluorescence intensities in a colour-coded manner from white (no binding) to red (strong binding). For all evaluations the MMC2-value for each peptide was calculated.

ITC

The ITC runs were performed using an iTC200 (Malvern Panalytical), using 300 µM peptide (syringe), and 30 µM of USP7 (N-ter/Catalytical domain/C-ter), T=20°C, 16 x 2.5 µL injections, stirring speed of 750rpm, feedback high. The experiments were performed in triplicate when an interaction was detected. The buffer vs buffer, buffer vs peptide, and protein vs buffer runs were performed and used as compiled controls. Data were analysed using the Microcal PEAQ-ITC Analysis software in the single-binding site mode with buffer-buffer and ligand-buffer controls.

X-ray crystallography

Crystallization optimization

Samples were prepared by mixing a 2:1 molar ratio of GRA16 peptide with USP7 N-ter domain (at a concentration of ~50 mg/ml) and incubating at room temperature for 30 minutes.

Sparse matrix crystallization screens were carried out at the High-Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble (Dimasi *et al.*, 2007; Dupeux *et al.*, 2011). Crystallization plates were setup using the sitting drop method, where drops consisted of 100 nl protein sample and 100 nl reservoir solution. Plates were kept at 20°C and visualised according to an imaging schedule with visible and UV light.

Promising conditions were refined. Refinement plates were prepared using a Formulator (Formulatrix) and sent to the HTX lab for set up.

Crystals were harvested using CrystalDirect technology (Zander *et al.*, 2016).

Data collection

Data were collected at beamlines IO3 at Diamond (UK) and MASSIF1 at ESRF (France) (Bowler *et al.*, 2015) using automatic protocols for the location and optimal centring of crystals (Svensson *et al.*, 2018). Strategy calculations accounted for flux and crystal volume in the parameter prediction for complete datasets.

Structure determination and refinement

Data were processed with XDS (Kabsch, 2010) and programs from the CCP4 suite (Collaborative Computational Project Number 4) (Winn *et al.*, 2011). The structure was determined by molecular replacement using MolRep. The crystal structure of the N-terminal domain of USP7 (aa 54-205) bound to an EBNA1 peptide (PDB 1YY6) was used as search model with the peptide residues and all water molecules removed. Refinement was carried out alternately using Phenix (Afonine *et al.*, 2012) and by manual rebuilding with COOT (Emsley and Cowtan, 2004). Tight NCS restraints between monomers in the asymmetric unit were used during refinement. Models were validated using Molprobity (Chen *et al.*, 2010) and figures were produced with PyMol (Schrodinger LLC, 2015).

Cellular assays

Cells

HEK293 T-REx epithelial cells (RRID:CVCL_D585) were cultured in DMEM (Invitrogen) supplemented with 10% heat inactivated FBS (Invitrogen), 10 mM Hepes buffer pH 7.2, 2 mM L-glutamine, and 50 µg/ml penicillin and streptomycin (Invitrogen). Cells were incubated at 37°C in 5% CO2.

293-T-REx transfection

Four GRA16 constructs were designed (GRA16^{FL}, GRA16$^{\Delta 100\text{-}118}$, GRA16$^{\Delta 196\text{-}206}$, and GRA16$^{\Delta 224\text{-}250}$) fused to HA-Flag tag and cloned into pcDNA T-REx inducible expression plasmid.

Flag-fusion-protein-expressing plasmids and a selectable marker for puromycin resistance were cotransfected in HEK293 T-Rex cells with Lipofectamine Reagent (Invitrogen) according to manufacturer's instructions. Transfected cells were grown in presence of 2 µg/ml puromycin for selection. Individual drug-resistant clones were expanded and tested for tetracycline-inducible gene expression.

Immuno-precipitation assay

Cell extracts were incubated with 50 µL anti-FLAG M2 Magnetic Beads (Sigma) for 1 hour. Beads were washed with 20 column volumes (CV) Lysis buffer (20 mM Tris pH 7.5, 0.5 M KCl, 20% glycerol, 0.2 mM EDTA, 0.5 mM DTT, 0.05% NP40, and protease inhibitors). Bound poly-peptides were eluted with 250 µg/ml Flag peptide diluted in BC100 buffer (20 mM Tris-HCl, pH 8.0, 0.1 M KCl, 20% Glycerol).

Western blots

Proteins were separated by SDS-PAGE, transferred to a PVDF membrane (Immobilon-P; Millipore) by liquid transfer and Western blots were probed using appropriate primary antibodies (anti-USP7 and anti-HA) followed by phosphatase-conjugated goat secondary antibodies (Promega). Signals were detected using NBT-BCIP (Amresco).

Results

GRA16 is found in a high-molecular-weight complex involving USP7 and PP2A-B55 (Bougdour *et al.*, 2013). We proceeded to characterize the interaction between the *T. gondii* effector and the host deubiquitinase USP7.

SLIM identification

We started with a sequence analysis of GRA16 using the Eukaryotic Linear Motif (ELM) resource (Kumar *et al.*, 2020). Several USP7 binding motifs were found (Figure 71 and Figure 73):

- MATH1 motif, which is found in USP7 classical substrates p53 and MDM2. It recognises a surface groove in the USP7 MATH domain. This motif will have frequent matches in proteins due to its simplicity, but it is not clear how numerous are the USP7 substrates.
- MATH2 motif, which is found in the EBNA1 protein from Epstein Barr virus as well as in two proteins of the Kaposi Sarcoma Herpesvirus (KSHV). It recognises the same site as the MATH1 motif with higher affinity.
- UBL2_3 motif, which is a targeting motif found in USP7 substrates. It docks to the C-terminal Ubl2 domain of the protein. Its description leads to poor predictability, but the binding specificity might be obtained by cooperating binding to the MATH domain.

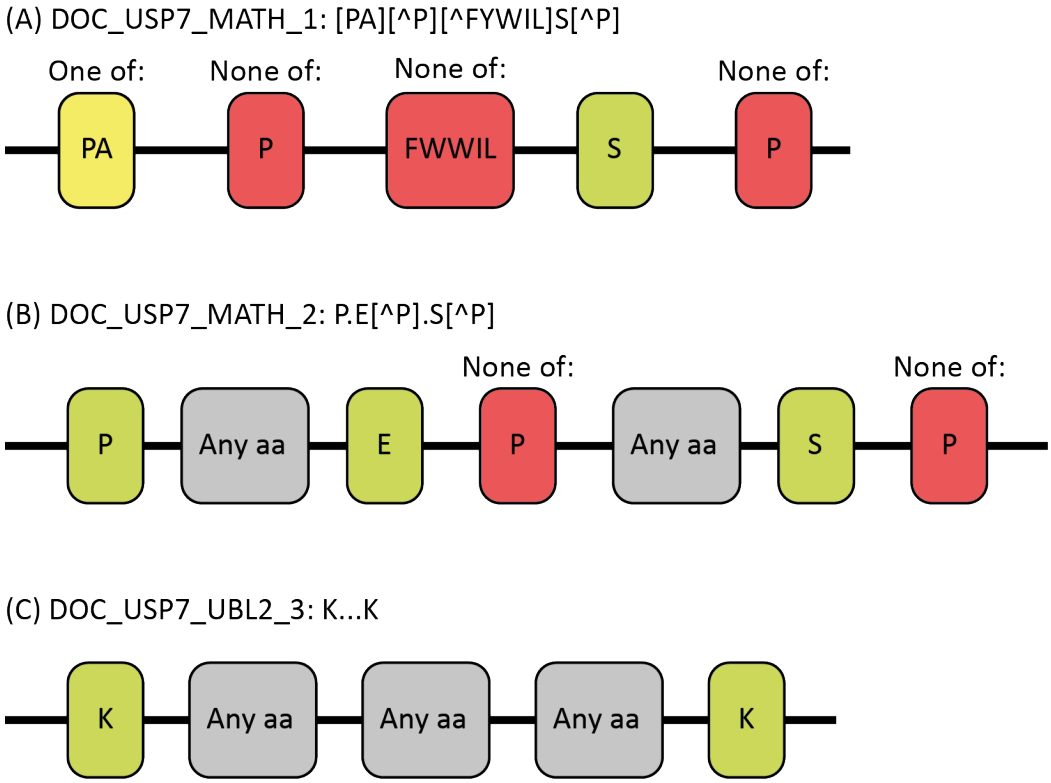


Figure 71: Linear docking motifs described to bind to USP7.

These motifs are not highly selective and have frequent matches in proteins. They are not sufficient on their own to make hypotheses on binding regions, but they are a good tool to validate hits from other methods and to understand the mechanism of the identified interactions.

Identification of the potential binding regions of GRA16 to USP7

In parallel, to experimentally identify regions of GRA16 interacting with USP7, JPT Peptide Technologies' PepStar™ peptide microarrays were performed by the company. In brief, the sequence of GRA16 from TGGT1 strain was used to generate a peptide library that was immobilized on the micro-array. The recombinant FL human USP7 protein was labelled and incubated on the micro-array. Binding between USP7 and GRA16 peptides was detected by fluorescence.

The interaction profiles of FL human USP7 with GRA16 peptides are shown on Figure 72. Four potential binding regions, referred to as binding regions 1-4, were detected in GRA16 (Figure 73).

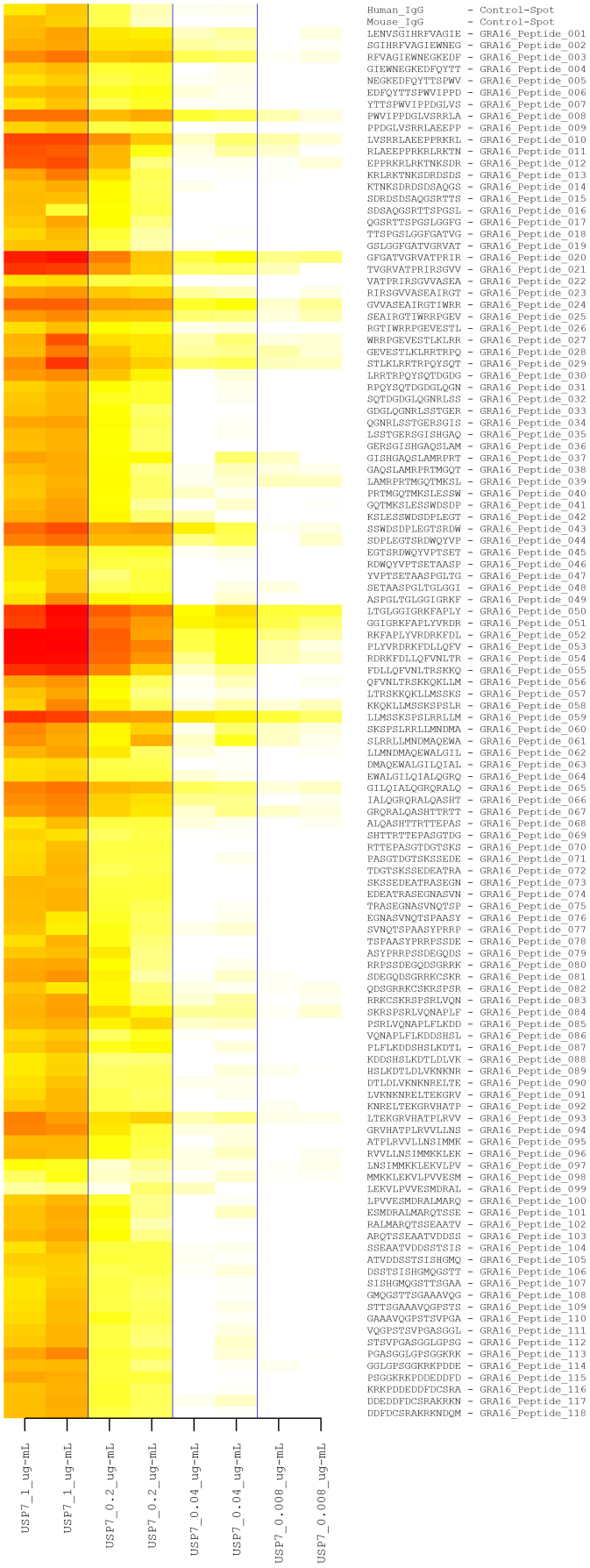


Figure 72: Heatmap diagram (log) showing all incubations of labelled Human USP7 and all displayed GRA16 peptides on slide. The MMC2 values are shown as colours; y-axis represents peptide sequences in the library, x-axis shows samples applied. Colour coding ranges from white (0 or low intensity) over yellow (middle intensity) to red (high intensity), blue lines separate serial dilutions of each sample incubated twice. All signals were normalized towards the highest signal measured.

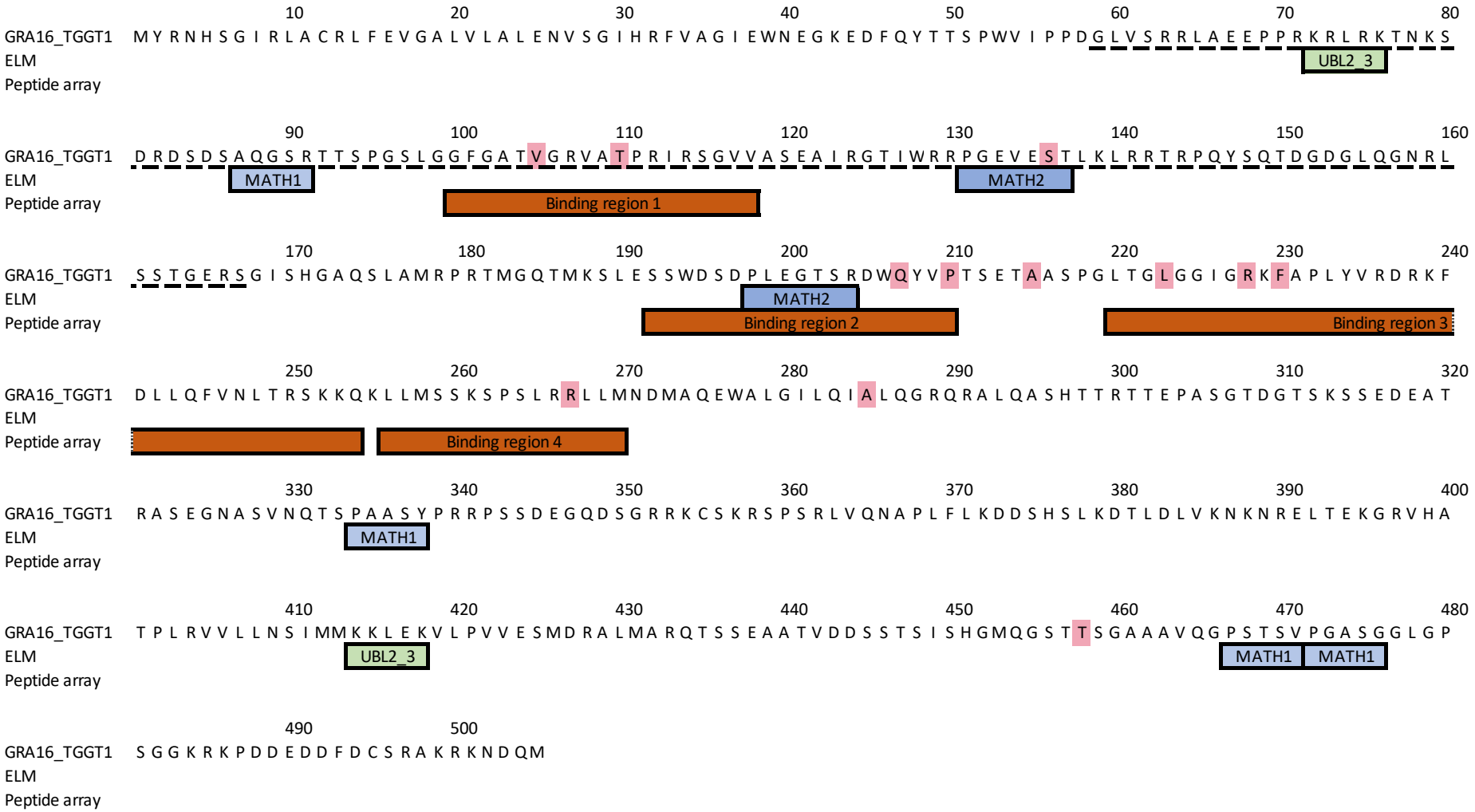


Figure 73: GRA16_TGGT1 sequence with potential linear motifs identified by elm.eu.org (in green and blue) and potential binding regions identified by peptide microarrays (in orange). Underlined in dash is the sequence (58-168) necessary for USP7 binding identified by Bougdour et al. (2013).

Only one region (binding region 1) was overlapping with the region necessary for USP7 binding identified by (Bougdour *et al.*, 2013).

One of the identified regions (binding region 2) contains a MATH2 motif. This motif was identified in the *Epstein Barr virus* EBNA1 protein (Saridakis *et al.*, 2005) and recognizes the same surface groove in USP7 as the p53 and MDM2 MATH1 motifs but requires a different description. Other occurrences of this motif appear in some Kaposi sarcoma herpesvirus (KSHV) proteins, vIRF1 and vIRF4 (Lee *et al.*, 2011; Chavoshi *et al.*, 2016). EBNA1 blocks USP7 activity and promotes p53 elimination, therefore protecting cells from apoptosis and might contribute to the survival of Epstein-Barr virus-infected cells. The mechanism of action of GRA16 could be similar.

Another MATH2 motif was found in GRA16_TGGT1 sequence (aa 131-138), but did not overlap with PepStar peptide array hits. However, it has been described to belong to the region essential for USP7 binding (Bougdour *et al.*, 2013). The peptide array signal of this region was weaker than the defined hit threshold but still presented some signal. Therefore, we decided to include it in the continuation of the study.

Interestingly, most of the identified potential binding regions in GRA16_TGGT1 were not entirely conserved in GRA16_TGME49 and GRA16_TGVEG, indicating that these interactions could be strain specific (Figure 73 and Appendix 5, p. 216).

The UBL2_3 motif and several MATH1 motifs detected in the sequence were not overlapping with binding regions identified in the assay.

Based on these results, we ordered different peptides to test their affinity for USP7 domains (Table 13).

Expression and purification of USP7 domains

Based on the literature, three domains of USP7 were identified to be functionally relevant and compatible with *E. coli* recombinant protein expression and crystallization:

- USP7 N-ter domain (53-206)
- USP7 catalytic domain (208-560)
- USP7 C-ter domain (560-1084)

These three domains were successfully expressed in *E. coli* and purified (Figure 74). The resulting protein samples were sent for MS analysis that confirmed the purity and the correct expected mass of the expressed constructs.

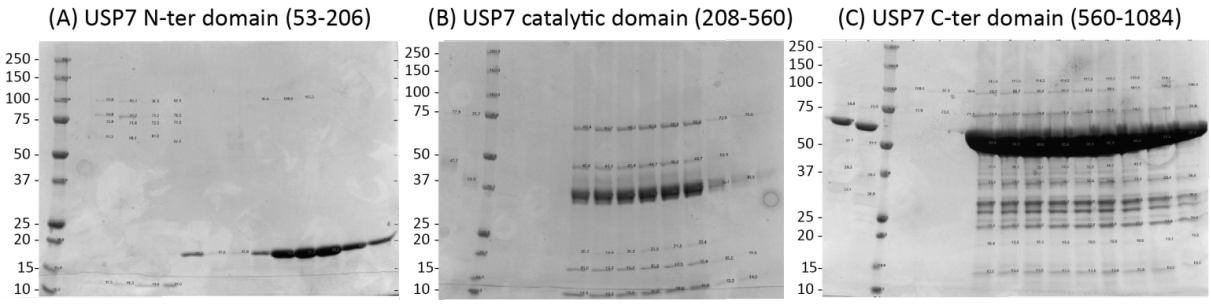


Figure 74: SDS PAGE gels of purified USP7 domains

GRA16 peptide containing MATH2 docking motif has high affinity for USP7 C-terminal domain

The affinity of the identified peptides of GRA16 for the three USP7 domains was measured by ITC. Results are shown in Table 14. One of the GRA16 regions identified in the peptide array assay was confirmed to bind to USP7 N-terminal domain.

The peptides carrying the MATH2 binding motif from the peptide array binding region 2 (residues 192-210) showed high affinity for USP7 N-terminal domain (Table 14 and Figure 75). The shorter peptide tested displayed an affinity of ~350 nM for USP7 N-ter domain.

This interaction is consistent with the described mechanism of docking of USP7 binding partners, where MATH binding motifs bind to a surface groove located in the TRAF-like domain at the N-terminal of USP7. In addition to the sequence homology of this GRA16 region with the EBNA1 region binding to USP7, these results support the relevance of this newly described interaction.

We tested different lengths of peptides covering this region, as well as other strain variants, to analyse the contribution of the different amino acids to the interaction. Interestingly, the longer peptides displayed higher affinity for USP7 N-ter domain than the shorter ones, about 10-fold higher for the binding of GRA16_TGGT1_192-210 (K_D of ~30 nM). The corresponding peptide from TGME49/TGVEG strains also displayed high affinity for USP7 (K_D of ~60 nM). This higher affinity might indicate additional bindings in complement to what is described for the MATH2 motif.

The other region containing a MATH2 binding motif, corresponding to the GRA16_TGGT1_129-138 peptide, did not bind to USP7 N-ter domain nor the catalytic and C-ter domain.

None of the tested peptides from the peptide array assay binding regions 1, 3 and 4 showed any significant binding for any of the USP7 domains.

Further experiments will need to be performed in order to understand if the regions identified in the PepStarTM peptide microarrays are relevant. One of the limitations of testing interactions on individual domains is that it does not reflect potential multi-domain interactions. ITC measurements of the interaction between GRA16 peptides and FL USP7 would solve this issue.

Table 14: ITC results of the interaction between GRA16 peptides and USP7 domains. Ø = no binding, / = not tested, rep = number of replicates. When affinity was detected, results present the mean and standard deviation of the K_D.

Peptide	**Sequence**	**Interaction with USP7 N-ter domain (53-206)**			**Interaction with USP7 catalytical domain (208-560)**			**Interaction with USP7 C-ter domain (560-1084)**		
		K_D (nM)	Sd (nM)	rep	K_D (nM)	Sd (nM)	rep	K_D (nM)	Sd (nM)	rep
GRA16_TGGT1_100-118	GFGATVGRVATPRIRSGVV	Ø		1	Ø		2	Ø		4
GRA16_TGGT1_104-114	TVGRVATPRIR	/			/			Ø		2
GRA16_TGGT1_129-138	RRPGEVESTL	Ø		2	/			/		
GRA16_192-206	SSWDSDPLEGTSRDW	294.0	197.1	6	/					
GRA16_196-206	SDPLEGTSRDW	352.2	62.0	3	/					
GRA16_TGGT1_192-210	SSWDSDPLEGTSRDWQYVP	31.2	15.3	6	Ø		1	Ø		4
GRA16_TGME49/TGVEG_192-210	SSWDSDPLEGTSRDWLYVT	58.6	33.9	5	/					
GRA16_TGGT1_196-210	SDPLEGTSRDWQYVP	119.6	23.5	6	/					
GRA16_TGGT1_220-254	LTGLGGIGRKFAPLYVRDRKFDLLQFVNLTRSKKQ	Ø		2	Ø		5	Ø		4
GRA16_TGGT1_256-270	LLMSSKSPSLRRLLM	Ø		2	Ø		4	Ø		4

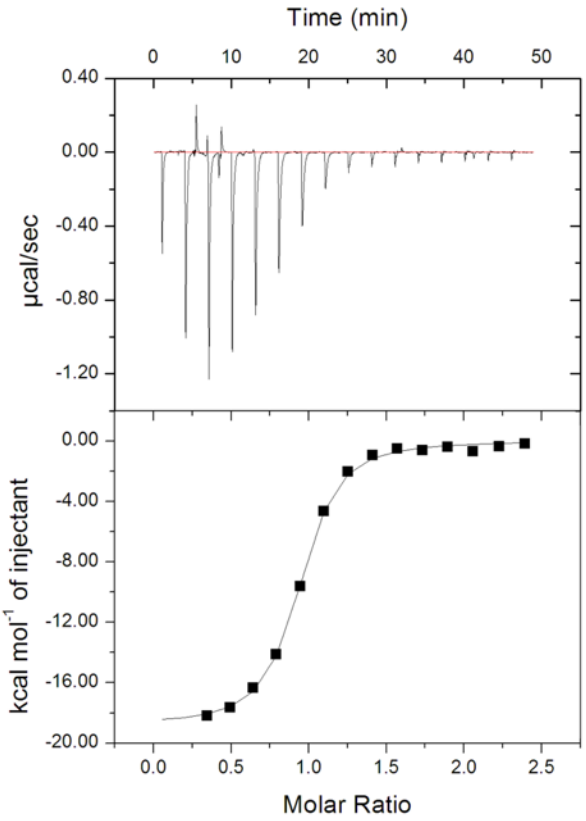


Figure 75: Typical ITC result of the binding between GRA16_196-206 and USP7 N-ter domain.

We pursued the project with the structural study of the interaction of the GRA16 peptides containing a MATH2 motif with USP7 N-terminal domain.

Crystal structure of GRA16 peptide containing a MATH2 motif to USP7 N-terminal domain

Crystallization optimization

Sparse matrix crystallization screens were performed at the HTX facility to try co-crystallize the GRA16_196-206 peptide with USP7 N-ter domain (53-206). Crystalline objects appeared 24 hours after setup (Figure 76).

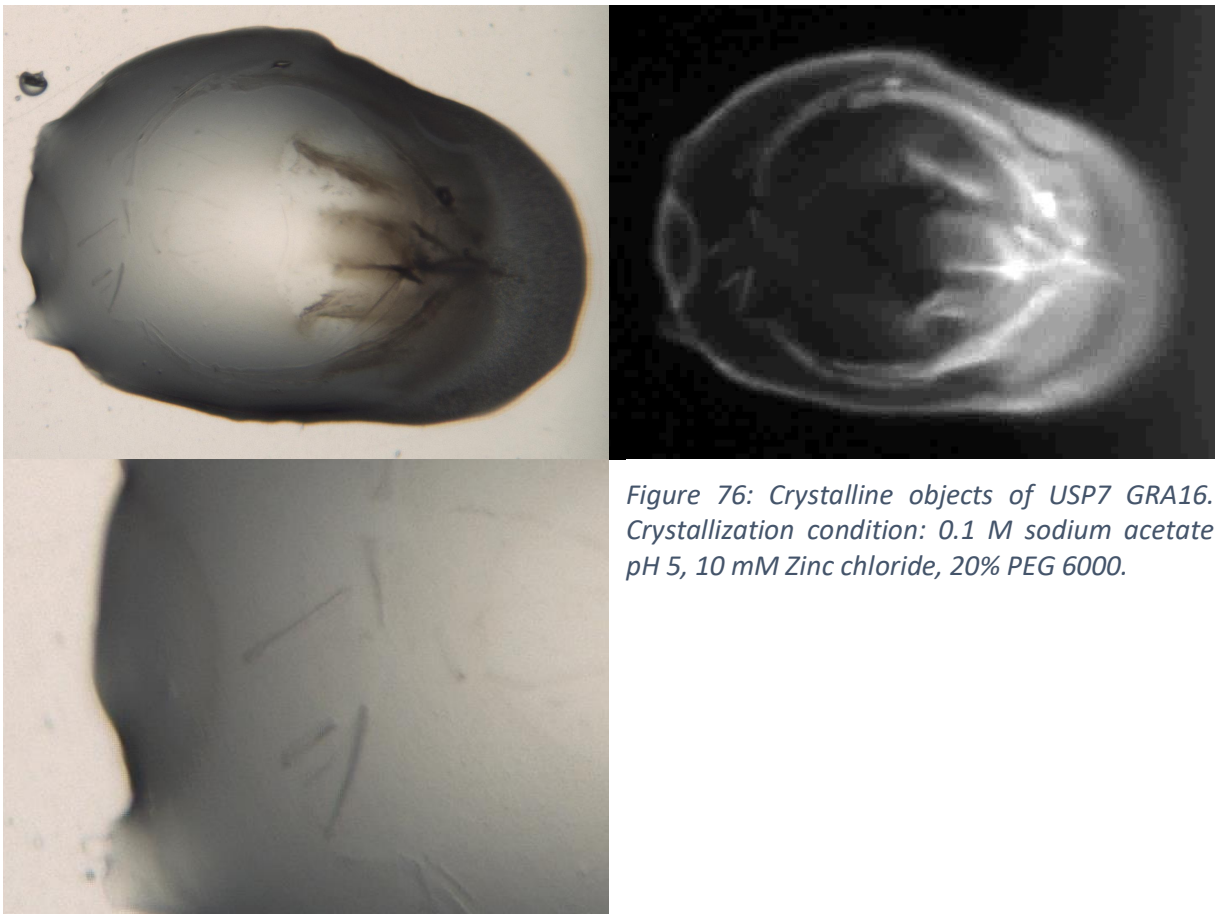


Figure 76: Crystalline objects of USP7 GRA16. Crystallization condition: 0.1 M sodium acetate pH 5, 10 mM Zinc chloride, 20% PEG 6000.

Based on this promising condition, refinement screens were performed and crystals were obtained in several drops, appearing after 1-2 weeks while the initial precipitate disappeared (Figure 77).

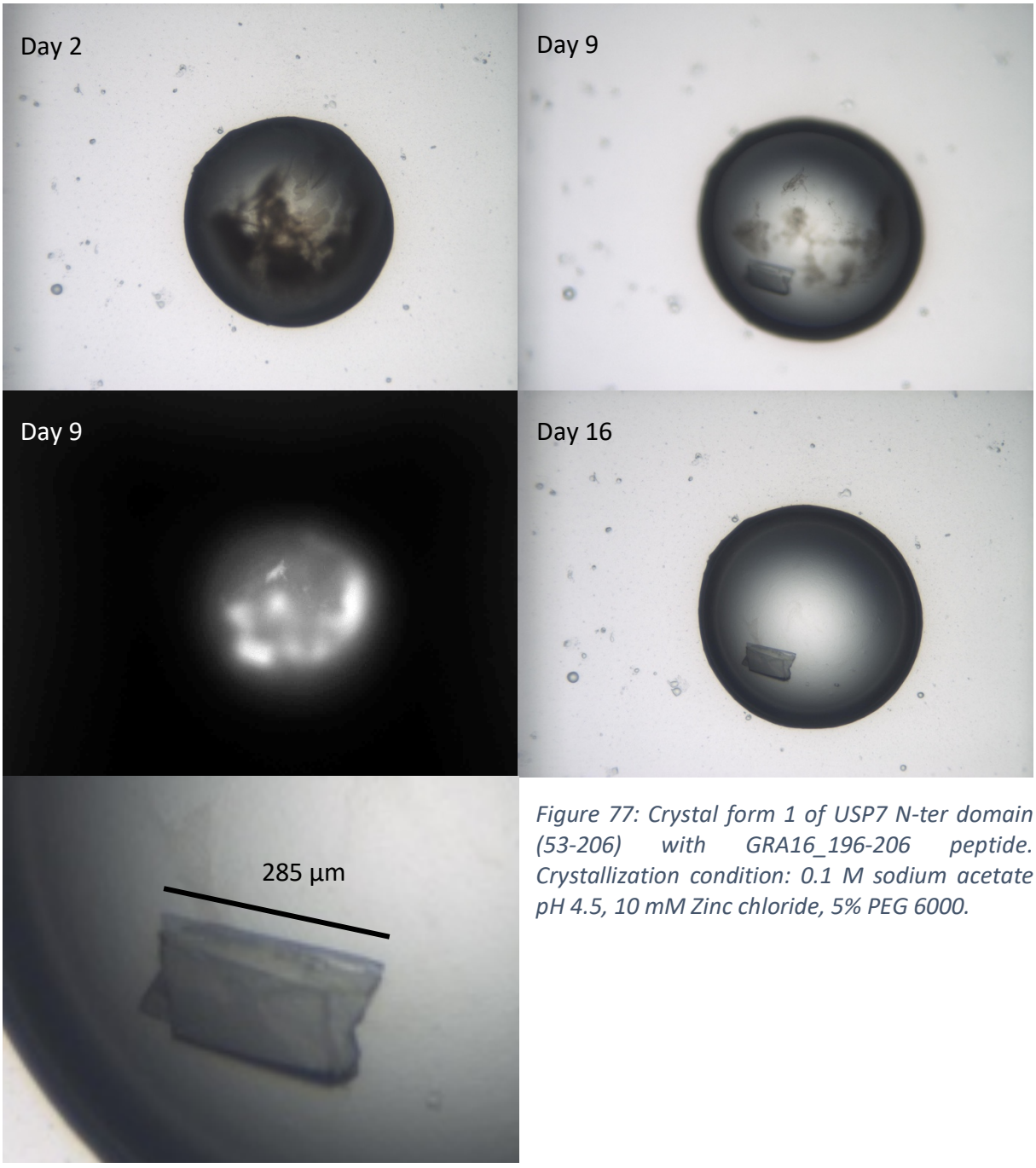


Figure 77: Crystal form 1 of USP7 N-ter domain (53-206) with GRA16_196-206 peptide. Crystallization condition: 0.1 M sodium acetate pH 4.5, 10 mM Zinc chloride, 5% PEG 6000.

A second crystal form appeared after 3 months, and was harvested 5 months later (Figure 78).

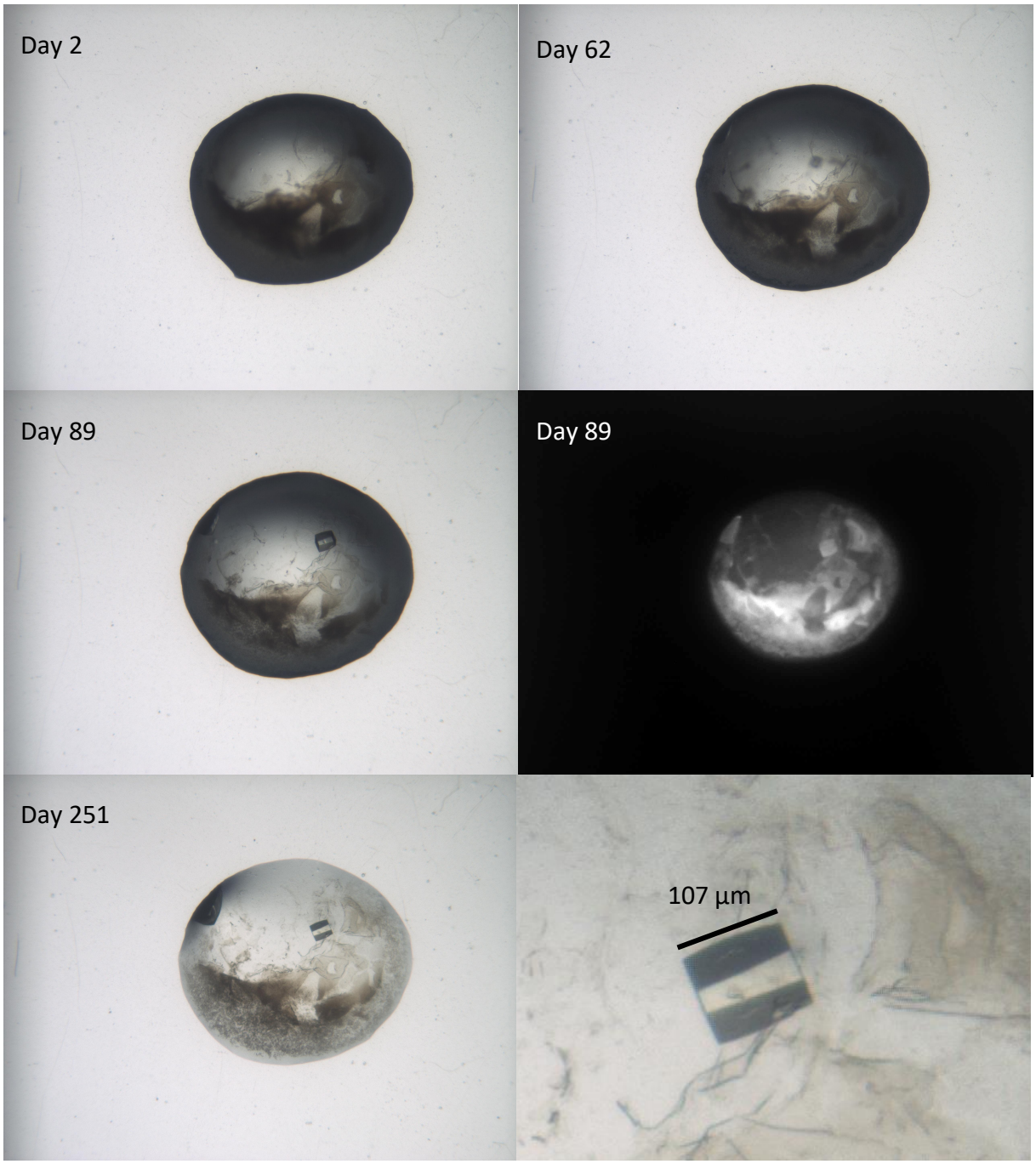


Figure 78: Crystal form 2 of USP7 N-ter domain (53-206) with GRA16_196-206 peptide. Crystallization condition: 0.1M sodium acetate pH 5, 10mM Zinc chloride, 5% PEG 6000.

Data collection and structure determination

Initially, two crystals were sent to beamline IO3 at Diamond (UK) for data collection (including crystal form 1 from Figure 77). Several positions were tested and most of them diffracted but only one dataset was successfully processed, leading to the crystal form 1 structure (Figure 79 and Table 15). Diffraction was to high resolution but many spots were split and multiple lattices could be observed even if the crystal appeared to be single. These problems were not observed in crystal form 2 (from Figure 78) that was measured at the MASSIF-1 beamline at ESRF (France), and resulted in a dataset of better quality (Figure 79 and Table 15).

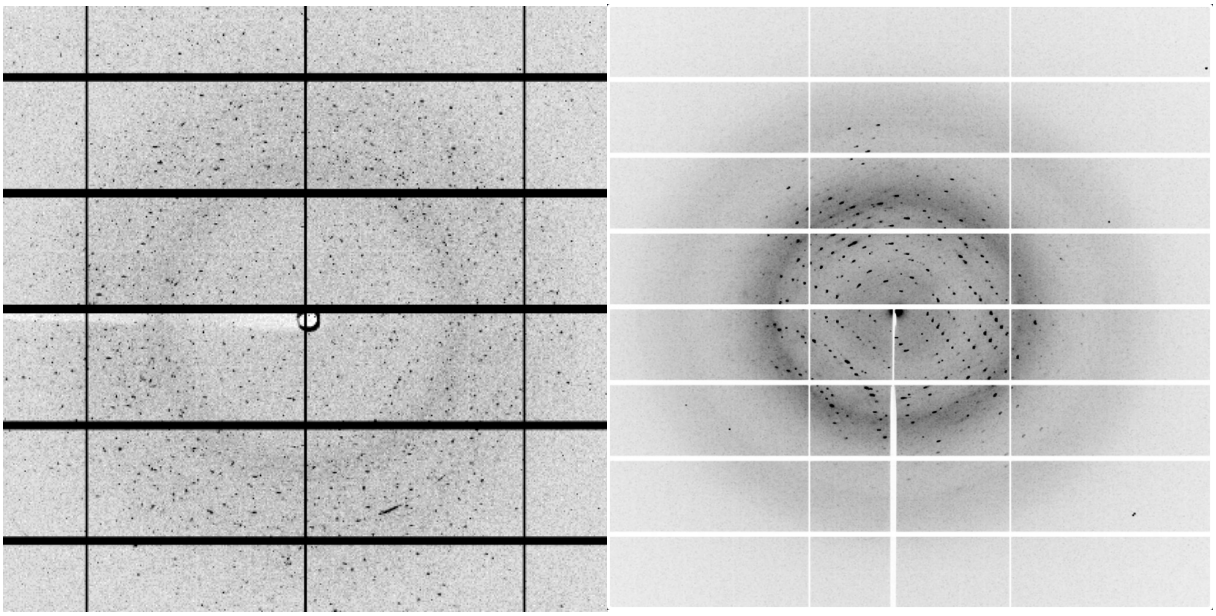


Figure 79: X-ray diffraction images from crystal form 1 (left) and crystal form 2 (right).

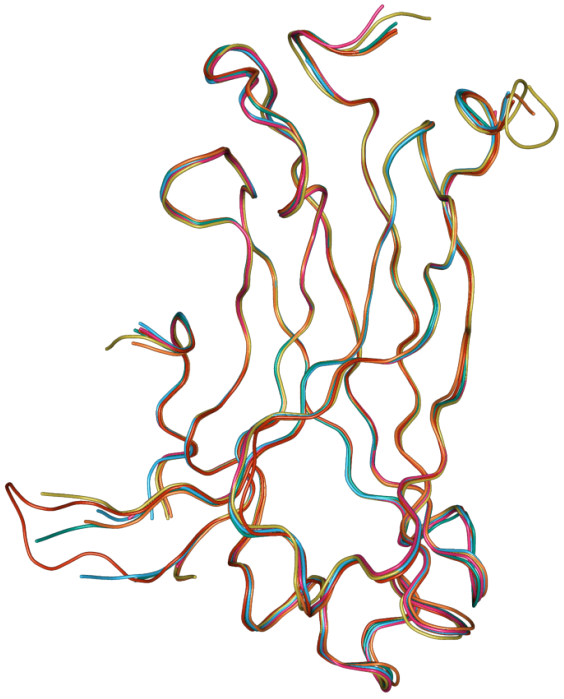
Both crystal structures are presented, and the information of the crystal form and chains will be specified for each figure.

Table 15: Data collection and refinement statistics

Model	**Crystal form 1**	**Crystal form 2**
Crystallization and data collection		
Stock solution	50 mg/ml, 3:1 ratio of peptide:protein	50 mg/ml, 3:1 ratio of peptide:protein
Crystallisation conditions	0.1M sodium acetate pH 4.5, 10mM Zinc chloride, 5% PEG 6000	0.1M sodium acetate pH 5, 10mM Zinc chloride, 5% PEG 6000
Beamline	IO3 (Diamond, UK)	MASSIF-1 (ESRF)
Wavelength (Å)	0.976	0.9655
Crystal dimensions (x,y,z, mm)	-	0.125 x 0.108 x 0.091
Space group	$P2_1$	$P2_1$
a, b, c (Å)	60.98, 59.88, 80.02	48.591, 60.260, 51.385
α, β, γ (°)	90, 94.1, 90	90, 104.42, 90
Unit-cell volume ($Å^3$)		
Wilson B factor ($Å^2$)	28.52	27.15
Resolution (Å)	28.44-1.69	47.06 - 1.626
Anisotropic (a*,b*,c*,Å)	-	1.6, 2.1, 1.8
Completeness (%)	96.4	65.5
Anisotropic (%)	-	92.7
Observed reflections	62279	62914
Unique reflections	184885	21465
Average multiplicity	3.0	3.0
⟨I/σ(I)⟩	10.1	12.2
R_{meas}	0.174 (0.648)	0.035 (0.394)
R_{merge}	0.124 (0.484)	0.028 (0.323)
$CC_{1/2}$	0.97 (0.36)	0.999 (0.934)
Refinement		
Resolution (Å)	1.69	1.63
R_{work},, R_{free} (5% set)	24.2, 29.9	0.1817, 0.2138
TLS groups		4
Geometry		
R.m.s.d., bond lengths (Å)	0.009	0.008
R.m.s.d., angles (°)	1.552	1.2272
Ramachandran statistics		
Preferred	95.0	97.5
Allowed		2.1
Outliers	1.3	0.0

The crystal form 1 structure of the USP7 N-terminal domain (53-206) bound to the GRA16_196-206 peptide was determined at 1.69 Å, the crystal form 2 structure was determined at 1.6 Å (Table 15). For both, phases were obtained by molecular replacement with Phaser using the USP7 N-terminal domain structure from PDB 1yy6 (deleting the EBNA1 peptide chain) and refined with a combination of CCP4, Phenix, and Coot. The complex crystallized in the monoclinic space group $P2_1$ in both crystal forms, but the unit cell of the second crystal form was smaller and contained only two pairs of USP7 protein – GRA16 peptide, when the asymmetric unit from crystal form 1 comprises 4 pairs of protein – peptide. All the pairs of USP7 protein – GRA16 peptide were in a similar conformation, with a few additional amino-acids built on some chains, stabilized by crystal contacts (Figure 80 and Figure 82).

The GRA16 peptide was located in the Fo-Fc difference electron density map after the initial round of refinement with the USP7 molecule alone (Figure 81B). Excellent density and similarity with the EBNA1 peptide allowed us to build in the GRA16 peptide unambiguously. Along with refinement, several alternative conformations of some amino acids clearly appeared.



Crystal form	Chain		Built residues
1	A	USP7	62-204
	B	GRA16	196-204
	C	USP7	61-106 112-204
	D	GRA16	196-204
	E	USP7	65-105 113-204
	F	GRA16	196-204
	G	USP7	58-107 112-206
	H	GRA16	196-204
2	A	USP7	62-106 11-201
	B	GRA16	196-204
	C	USP7	63-107 112-204
	D	GRA16	196-204

Figure 80: Alignment of the pairs of USP7 protein - GRA16 peptide from crystal form 1 and 2.

The GRA16 peptide binds to the classical peptide binding site of TRAF like domains, similar to the EBNA1 peptide (Figure 81D), in an extended conformation bound to the β-strand 7 in an antiparallel manner, adding a strand to the sheet composed of USP7 β-strands 3 to 7 (Figure 81A). A combination of hydrophobic and hydrophilic interactions are formed (Figure 81C). The GRA16 serine residue Ser203 is crucial for the interaction, interacting closely through several H-bonds to USP7 Arg104 and Asp164. Other H-bonds are formed between GRA16 residues 198 to 201 and USP7 residues. The interaction is also stabilized by hydrophobic interactions involving USP7 Met100, Met102 and Phe118 residues with the C-terminal of GRA16 peptide on one part, and USP7 Ile 154, Pe1667 and Arg152 residues with the N-terminal of GRA16

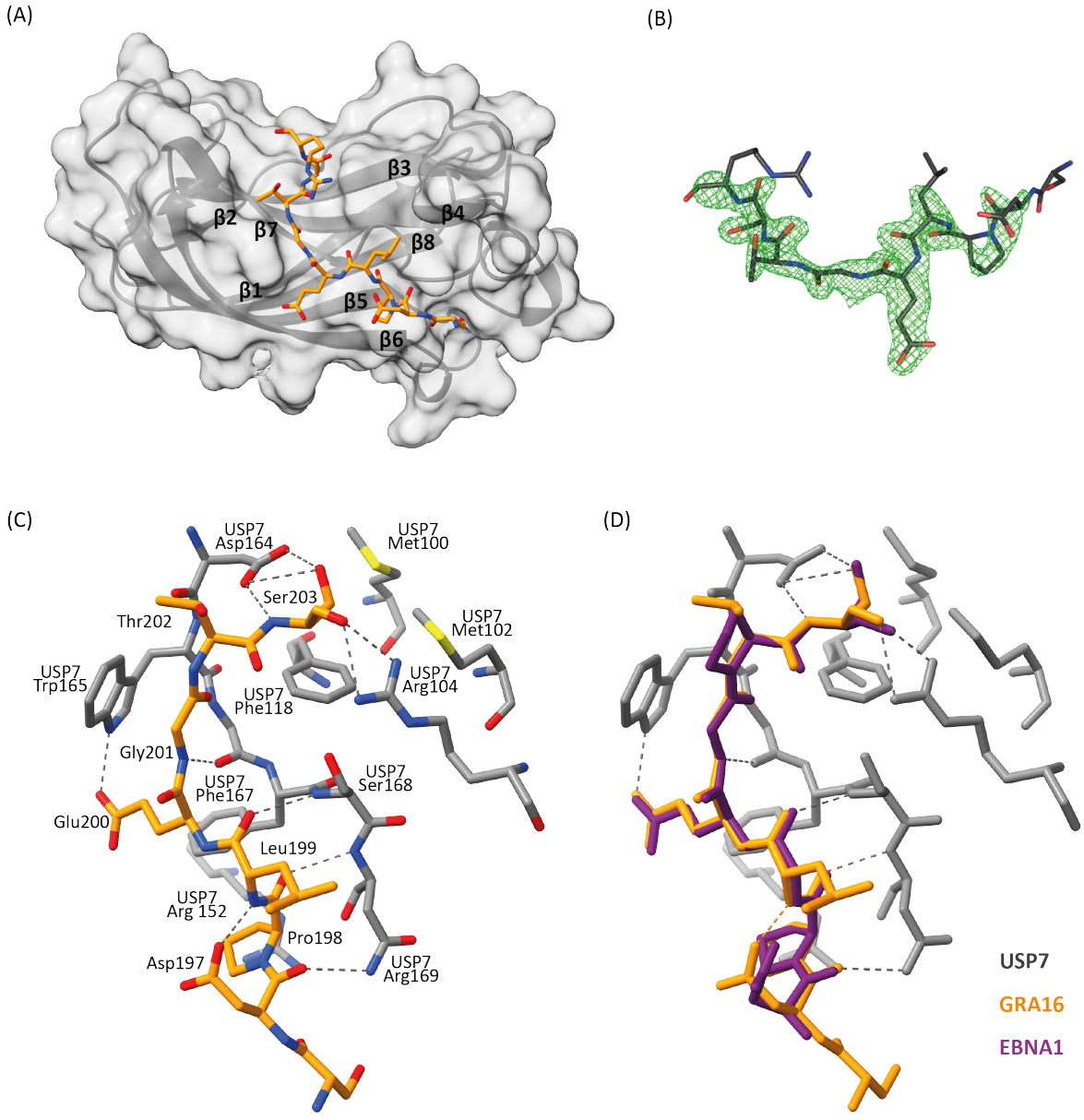


Figure 81: Crystal structure of USP7 N-terminal domain (53-206) bound to the GRA16_196-206 peptide (crystal form 2, chains A and B). (A) Interactions between GRA16 peptide (as sticks in yellow) and USP7 (as cartoon and surface in grey) (B) Omit map electron density (in green) in which the GRA16 peptide was built. (C) Detailed interactions between the GRA16_196-206 peptide (in yellow) and USP7 (in grey). (D) Comparison with EBNA1 peptide (in purple) binding to USP7 (PDB 1yy6).

peptide on another one. The docking of the GRA16 peptides seemed to induce no significant conformational change to USP7, similar to the other peptide bound structures of the USP7 N-terminal domain.

Despite a higher affinity for the longer GRA16 peptide, we did not manage to co-crystalize it with the USP7 N-terminal domain. Looking at the crystal packing of the structure gives us an explanation. Interactions between GRA16 peptides and symmetry and non-symmetry related USP7 molecules are involved in the packing (Figure 82).

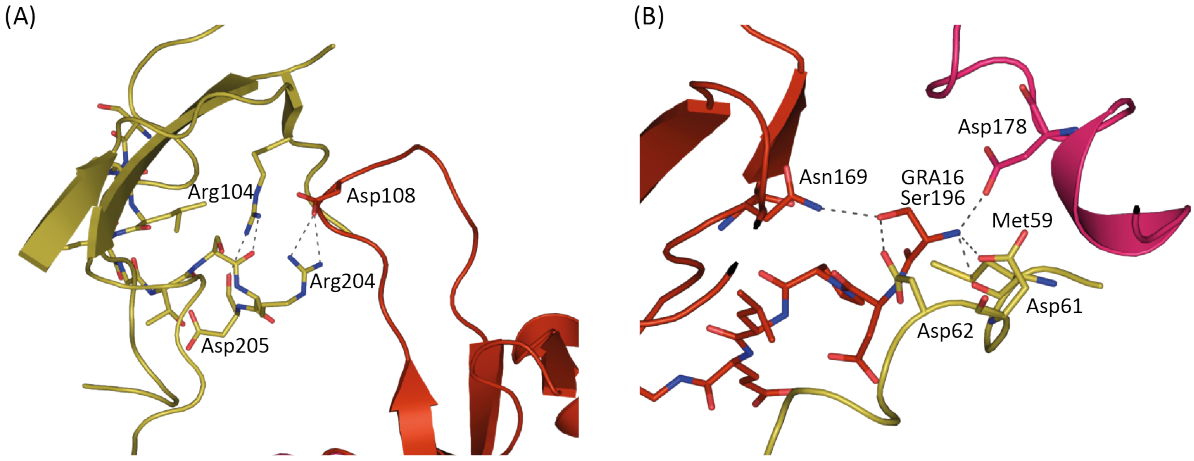


Figure 82: Interaction between GRA16 peptide and symmetry and non-symmetry related USP7 N-terminal domain in crystal form 1. (A) GRA16 Arg204 from chain H stabilizes USP7 loop (aa 106-112) from non-symmetry related chain A through H-bond with Asp108 (B) GRA16 Ser196 coordinate symmetry related USP7 chain A and non-symmetry related USP7 chains E and G at the crystal interface.

Crystal forms 1 and 2 appeared in the same plate, with just a difference of pH of 0,5 units. Crystal form 2 appeared long after crystal form 1 (3 months vs 9 days), and was harvested later (5 months). This led to different unit cell dimensions within the same space group. The early crystal form 1 was composed of several layers, and the diffraction was not homogeneous. The structure contains 4 pairs of protein – peptide in its asymmetric unit. The late crystal form 2 was more symmetric and homogeneous, and presents only 2 pairs of protein – peptide in its asymmetric unit. Alignment of the crystal packing shows a clear transition between form 1 and form 2, with a slight rotation and the molecules getting closer in form 2 resulting in an additional symmetry plane (Figure 83). This could be explained by the dehydration of the drop and crystal, allowing the molecules to pack closer together. Interestingly, the crystal contacts observed in crystal form 1 (Figure 82) were not observed anymore in crystal form 2.

Comparison of the binding between different peptides and USP7 N-terminal domain

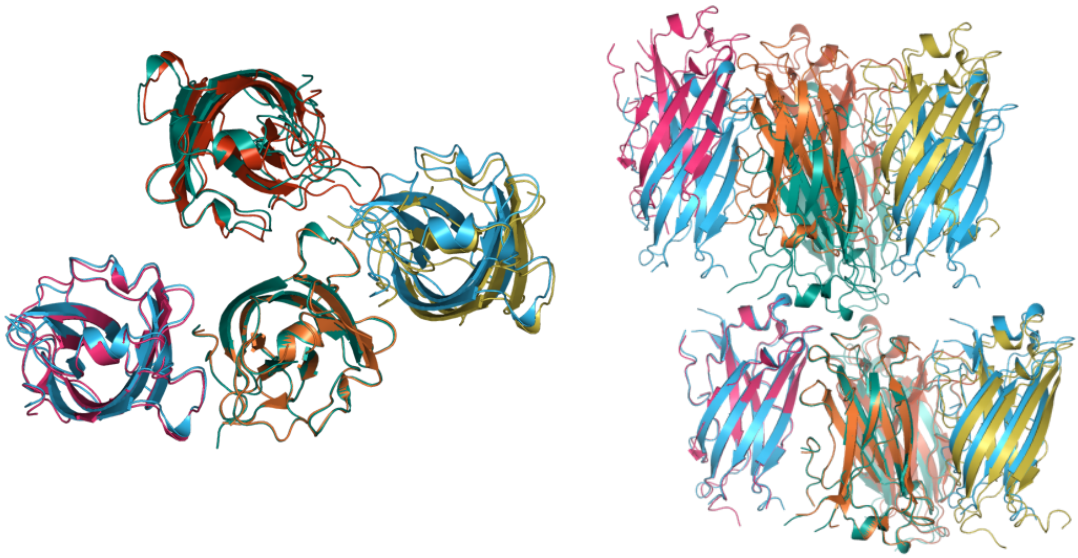


Figure 83: Comparison of the crystal packing of the two crystal forms (crystal form 1 in red, orange, magenta and yellow; crystal form 2 in cyan and teal).

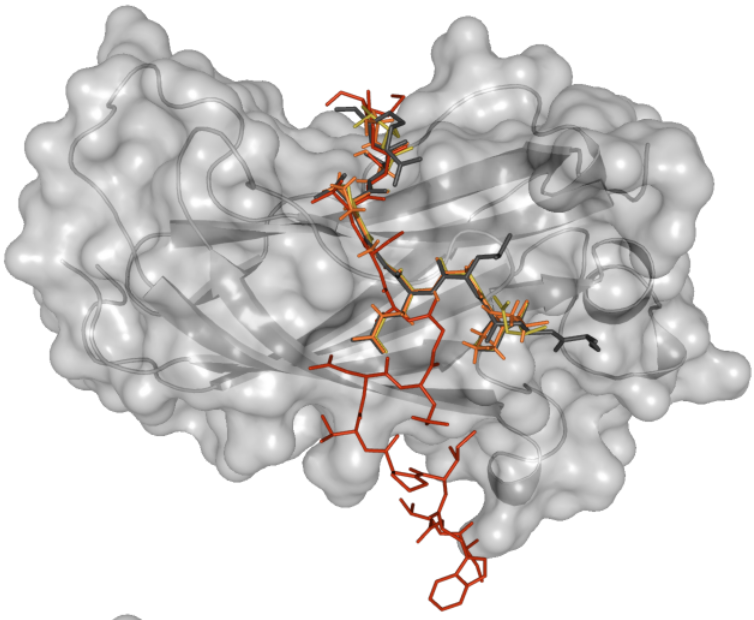
USP7 N-terminal domain adopts a structure similar to TRAF-C domains from TRAF proteins (Tumor necrosis factor (TNF) receptor-associated factors). It forms an eight-stranded antiparallel β-sandwich. The β7-strand forms a β-bulge that is conserved in all TRAF domain structures, and is important for peptide binding.

The function of the TRAF-like domain is to bind to USP7 substrates. USP7 substrates contain similar motifs: the MATH1 motif is found in human binding proteins such as MDM2 and p53, or the MATH2 motif that has been described in viral proteins and is found in *T. gondii* GRA16. The substrates compete for USP7 binding.

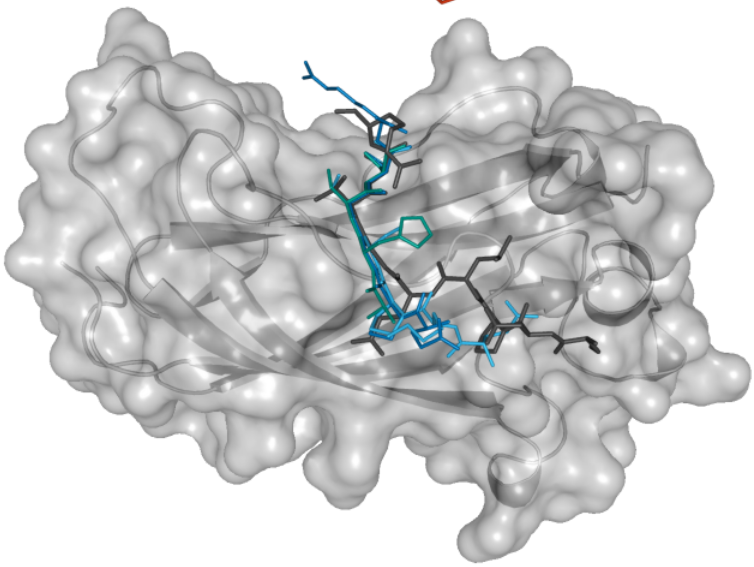
ITC results showed that GRA16 peptide has a higher affinity for USP7 compared to the EBNA1 peptide (around 0.9 µM (Saridakis *et al.*, 2005)). Comparison between the two structures (PDB 1yy6 and ours) shows a similar binding mode and that the most crucial interactors are conserved (Figure 81D).

Alignment with the human interactors of USP7, p53 and MDM2, showed that only the C-terminal interaction of the peptides with USP7 is conserved, which comprised the only conserved amino acid between all the interactor: Ser203 in GRA16, Ser447 in EBNA1, Ser367 in p53, and Ser147 in MDM2 (Figure 84B). As for the EBNA1 protein, extended interaction in the N-terminal region of the GRA16 peptide explains the significantly higher affinity for USP7 in comparison to human interactors.

(A)



(B)



(C)

GRA16 TGGT1 192-210	SSWDSDPLEGTSRDWQYVP	
GRA16 TGME49/TGVEG 192-210	SSWDSDPLEGTSRDWLYVT	
EBNA1 436-454	QGPADDPGEGPSTGPRGQG	1YY6
KSHV-vIRF1 39-57	PSRSESPGEGPSGTGGSAA	4YSI
KSHV-vIRF4 203-221	VWIPVNEGASTSGMGSSGT	2XXN
MDM2 139-157	VQELQEEKPSSSHLVSRPS	2FOP
MDM2 389-407	QESEDYSQPSTSSSIIYSS	3MQS
p53 351-369	KDAQAGKEPGGSRAHSSHL	2FOO
p53 356-374 S362A	GKEPGGARAHSSHLKSKKG	2FOJ
HDMX 390-408	NSVEFLDLAHSSESQETIS	3MQR

Figure 84: Comparison of the binding mode of MATH binding peptides (in stick representation, coloured according to table C) to the USP7 N-terminal domain (in cartoon and surface representation in grey) (A) GRA16 and viral interactors EBNA1 (PDB 1YY6), vIRF1 (PDB 4YSI), and vIRF4 (PDB 2XXN) (B) GRA16 and human interactors p53 (PDB 2FOO), MDM2 (PDB 2FOP), and MDMX (PDB 3MQR) (C) Alignment of MATH binding peptide sequence of GRA16, EBNA1, vIRF1, vIRF4, p53, MDM2 and MDMX.

Deletion of GRA16 MATH2 motif abolishes its interaction with USP7

In collaboration with Ali Hakimi's group (IAB, Grenoble), we studied the effects of the deletion of the identified potential GRA16 binding region 1 (res 100-118), region 2 (res 196-206) and region 3 (res 224-250) detected by the peptide array experiment on the interaction with USP7 (Figure 85).

As previously described in (Bougdour *et al.*, 2013), USP7 co-precipitates with GRA16^{FL}. The deletion of region 1 or 3 (GRA16$^{\Delta 100-118}$ and GRA16$^{\Delta 224-250}$ respectively) does not seem to affect the binding with USP7. However, the deletion of region 2, containing the MATH2 motif, (GRA16$^{\Delta 196-206}$) abolishes almost completely the co-precipitation with USP7.

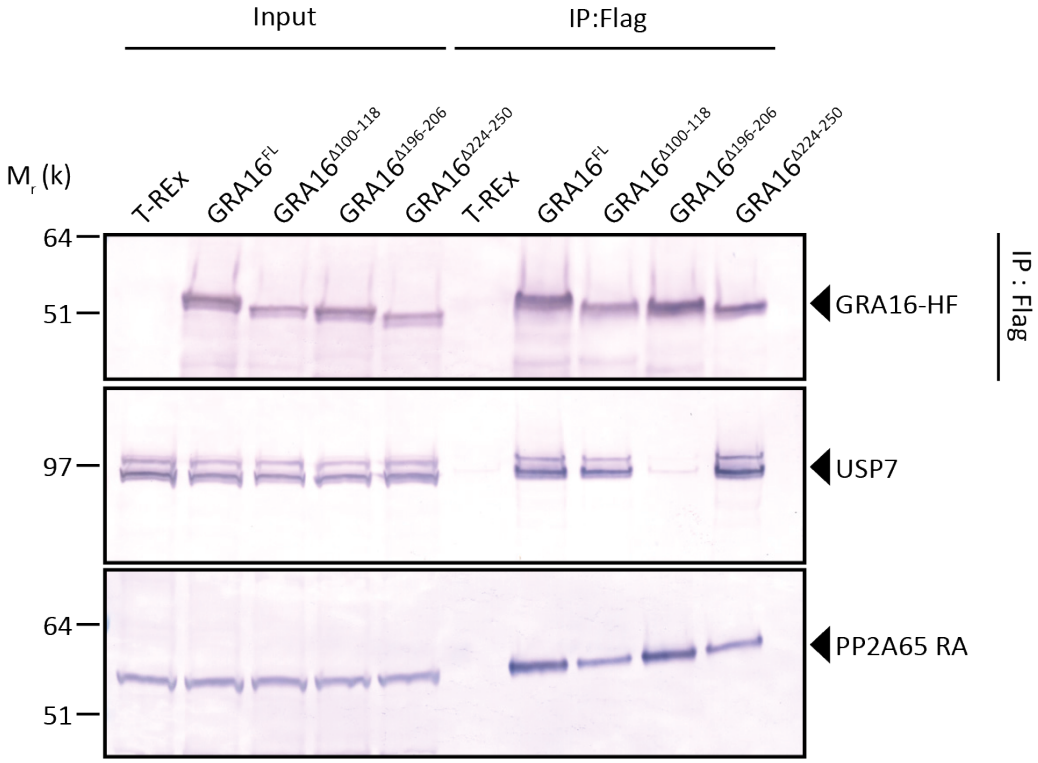


Figure 85: The MATH2 motif of GRA16 is necessary for binding with USP7. T-REx cells were transfected with different GRA16 constructs. GRA16-HAFlag was immunoprecipitated from T-REx cells with anti-Flag antibodies and analysed by western blotting.

None of the deletions seemed to affect binding to PP2A as the PP265 RA scaffolding unit co-precipitated with all the GRA16 constructs.

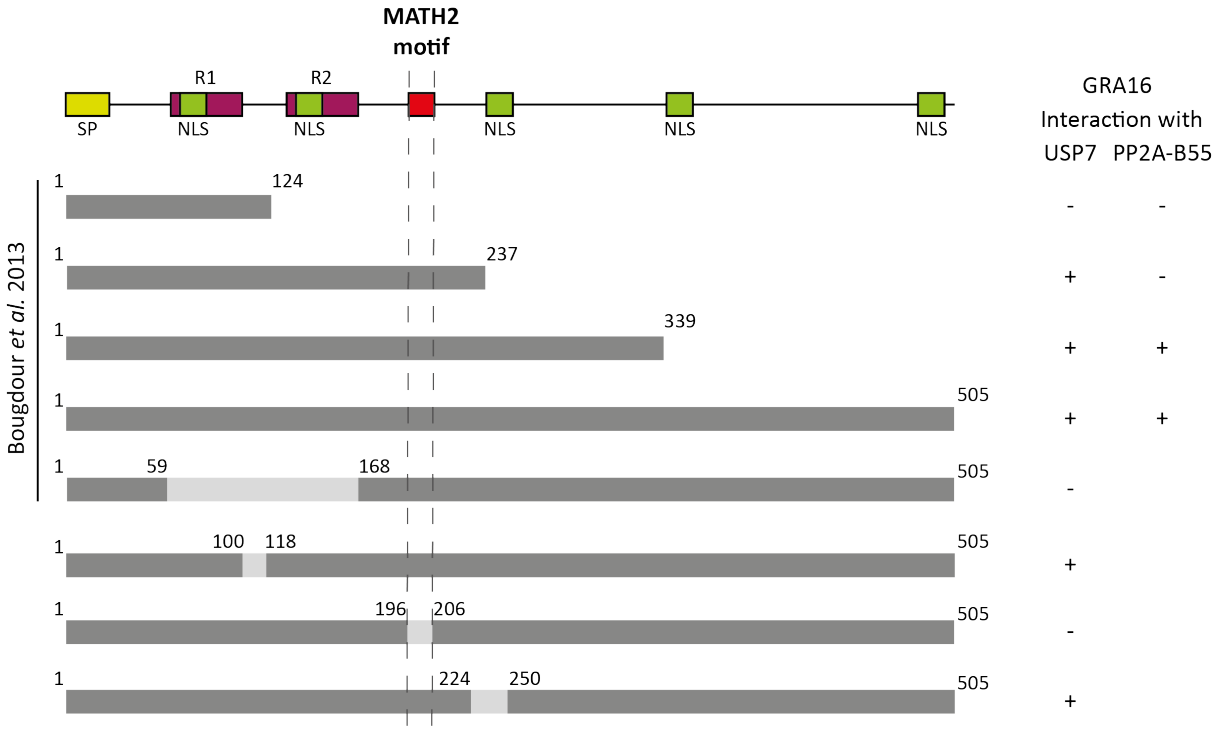


Figure 86: Diagram comparing our results with the results published in Bougdour et al. 2013.

This cellular assay confirms that the region of GRA16 containing the MATH2 motif is the major determinant of the interaction with USP7. This supports the *in vitro* observation of the high affinity between GRA16 peptides and USP7 N-terminal domain. GRA16 binds to the same USP7 docking site as the host p53 and MDM2 proteins. It adopts the same binding mode as the one described for the viral EBNA1 protein (Saridakis *et al.*, 2005).

Discussion and perspectives

Necessity to complement *in silico* and *in vitro* interaction studies with cellular assays

SLiMs are a great tool to predict and understand protein-protein interactions. However, most of them have a loose description resulting in a number of false positives. These motifs are useful to identify potential interacting regions and to understand the mechanism of interaction, but are not reliable on their own. Combining their prediction with peptide array assays is a good strategy as the peptide array assays look at interactions between short peptides and FL protein. The comparison between the hits of the two techniques already narrows down the selection.

For example, in the GRA16 sequence, four MATH1 and two MATH2 motifs were found. Only one of them, the MATH2 found in res 196-210, overlapped with the peptide array region. After validating the interaction with ITC and determining the crystal structure, we tested the affinity for USP7 of the second occurrence of the MATH2 motif present in GRA16 sequence, and no binding was detected.

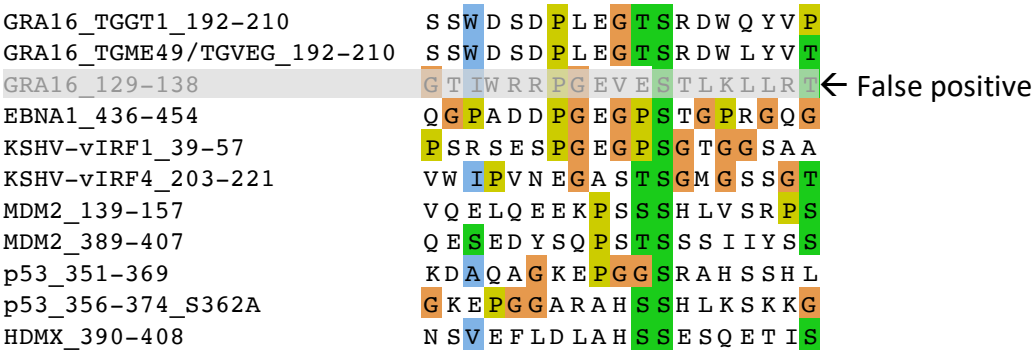


Figure 87: Alignment of MATH binding motifs.

We did not observe binding for 3 out of 4 potential binding regions detected in Peptide array. Some limitations of our strategy originate from the decision to measure the affinity between peptides and USP7 domains, when the peptide array experiment was conducted with the USP7 FL. The possible interactions mediated between different domains, or disordered parts in linker regions could not be detected in this setup.

The deletion of potential binding regions 1 and 3 and subsequent IP assays indicate that they do not affect the interaction with USP7. They might be false positives from the peptide array measurement.

However, (Bougdour *et al.*, 2013) published that GRA16$^{\Delta 59\text{-}168}$ lost its ability to bind to USP7 when expressed in a stable inducible T-Rex-293 cell line (but would still bind to PP2A). This could indicate that the motif GRA16$^{192\text{-}210}$ might not be the only region involved in the interaction, and might indicate the presence of secondary binding regions. Structural studies

in the context of FL GRA16 could bring answers on the extent of the interaction between the parasite effector and the host deubiquitinase.

GRA16 interacts with USP7 N-terminal domain through a MATH2 motif

We demonstrated that GRA16 interacts with the USP7 N-terminal domain through a MATH2 motif located after the repeats. This interacting region has been detected by a combination of *in silico* prediction using the ELM resource (Kumar *et al.*, 2020) and *in vitro* peptide array assay. It has been confirmed by ITC measurement of the interaction and the resolution of the crystal structure of USP7 N-terminal domain bound to GRA16 peptide.

The relevance of this interacting region was validated by immune-precipitation studies in which the deletion of the MATH2 motif from GRA16 sequence abolished the co-precipitation with USP7 *in vivo*.

Interestingly, the affinity was higher for peptides longer than just the MATH2 motif. There is probably some interaction between USP7 and GRA16 residues located beyond the motif description. However, we could not obtain crystals with the longer GRA16 peptides despite their higher affinity for USP7, probably because of necessary crystal contact interactions in the extended peptide binding region.

GRA16 is the first non-viral protein containing a MATH2 motif to be described. So far, it has only been observed in viral proteins, and three structures of MATH2 viral peptide with USP7 N-terminal domain are available, including EBNA1 protein from Epstein Barr virus (Saridakis *et al.*, 2005) and in two vIRF proteins from Kaposi sarcoma herpesvirus (KSHV) (Lee *et al.*, 2011; Chavoshi *et al.*, 2016). Many more viral proteins are interacting with USP7 and have putative or confirmed MATH2 motifs (Bojagora and Saridakis, 2020). GRA16 MATH2 peptide bind to USP7 the exact same way as EBNA1 peptide does. It is possible that GRA16 sequence could have been integrated in *T. gondii* genome through viral infection of the parasite.

A tool to study the effects of GRA16 during infection

A crucial question remains: what is the effect of GRA16 on USP7? It could act as an inhibitor by binding competition for the MATH docking site present in the N-terminal domain of USP7 (TRAF-like domain). It binds to the same site as other interactors, including MDM2 and p53. This effect has been proposed for the MATH2-containing viral proteins (Lee *et al.*, 2011; Chavoshi *et al.*, 2016; Bojagora and Saridakis, 2020).

It could inhibit the deubiquitination of p53 by USP7, promoting its degradation by proteolysis, resulting in the promotion of cell survival (Figure 88). However, p53 is involved in a multitude of pathways and is regulated tightly by a complex equilibrium. p53 is a tumour suppressor that is dysregulated in many cancers, and consequently in many immortalized cell-lines, making its study more difficult in this context.

We need to identify other USP7 substrates, potentially independent of p53, and test USP7 activity.

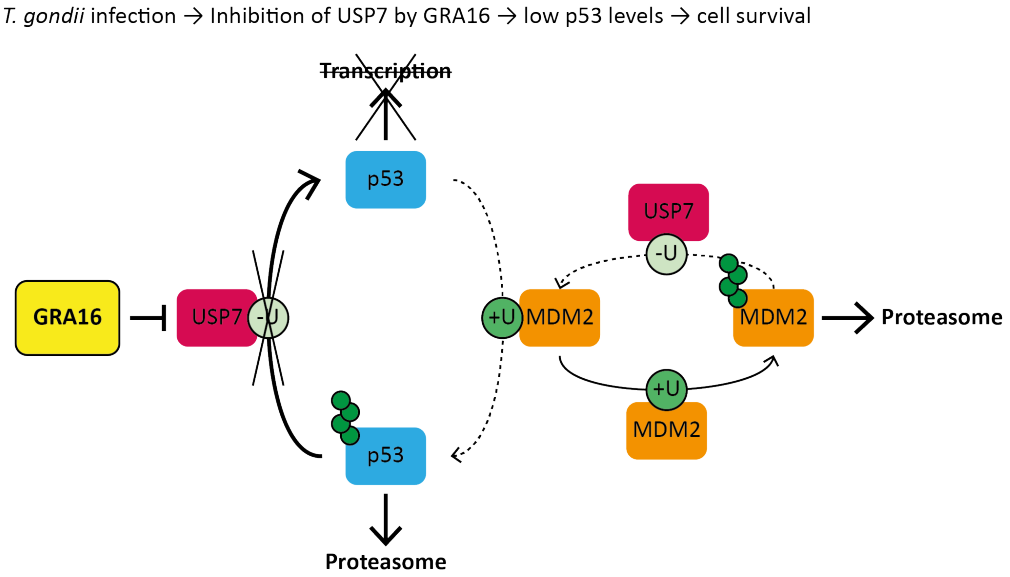


Figure 88: Possible mechanism of inhibition of USP7 by GRA16. USP7 inhibition could maintain low levels of p53 in infected cells despite the stress induced by infection and promote host cell survival.

The identification of the key regions and residues mediating the interaction between GRA16 and USP7 allows us to design point mutations in GRA16 sequence that would abolish USP7 binding. These mutations will be first evaluated *in vitro* at the peptide level by ITC, before being inserted into the *T. gondii* genome. It will constitute a valuable tool to study the resulting effects of the interaction between GRA16 and USP7 during parasite infection with transcriptomic studies. The comparison of the outcomes of over-expression of GRA16 WT and mutated in human cells could also bring us important insights on the mechanism of action of GRA16.

It would also be interesting to study the effects of USP7 inhibitors on the infection by *T. gondii* process.

Toward the study of the GRA16-USP7-PP2A complex

GRA16 co-elutes with a high MW complex containing USP7 and PP2A-B55 (Bougdour *et al.*, 2013). It does not affect the levels of PP2A-B55 but rather its localization, the whole complex shuttles and accumulates into the host cell nucleus. The interaction with PP2A is mediated by a different region in GRA16 than the one interacting with USP7, comprised between residues 237 and 339, and the deletion of the USP7 binding region does not abolishes binding to PP2A (Figure 85, p. 178). GRA16 brings together the USP7 deubiquitinase and the PP2A phosphatase. Whether GRA16 exerts its effect on both independently or by bringing them together remains to elucidate.

PP2A is a heterotrimeric protein phosphatase composed of a structural subunit A, a regulatory subunit B and a catalytic subunit C. The regulatory subunit B modulates both the substrate specificity and the cellular localization of the holoenzyme. Each subunit comprises several isoforms, resulting in over 90 different possible holoenzyme assemblies. GRA16 binds to a

specific composition of the holoenzyme involving the catalytic subunit PP2Ac, the protein scaffold A and the regulatory subunit B55. Active PP2A-B55α resides primarily into the cytoplasm. PP2A is involved in the repression of cellular proliferation, blocking the entry into M phase.

It was recently described that USP7 interacts with PP2A and supports its active localization in the cytoplasm (Galarreta *et al.*, 2021). They propose that during interphase in normal cells, USP7 mediates the inhibition of CDK1 by PP2A-B55. CDK1 drives the entry into mitosis, and premature activation of this mitotic kinase during interphase generates DNA damage.

USP7 inhibition by small molecules stops DNA replication. The proposed mechanism of action is that they contribute to p53 stabilization due to the degradation of MDM2 (Li *et al.*, 2002). However, USP7 inhibitors also induce DNA damage independently of p53 (Lecona *et al.*, 2016). (Galarreta *et al.*, 2021) showed that USP7 inhibitors suppresses PP2A activity which in turn drives the activation of CDK1 leading to DNA damage and toxicity. USP7 inhibition does not affect levels of PP2A members but triggers the translocation of PP2A-B55 to the nucleus. This change of localization, as observed in *T. gondii* infection, could underlie the regulation mechanism. These results open a new scope on the effects of USP7 inhibition independent of p53, but the mechanism of interaction between USP7 and PP2A-B55 remains to be elucidated.

GRA16 could mediate USP7 inhibition by direct binding competition with cellular partners through its MATH2 motif; it could also modulate PP2A-B55 phosphatase activity by stabilizing its interaction with USP7; and it could affect the USP7/PP2A-B55 activity by translocation of the complex to the host cell nucleus.

To study the direct interaction between GRA16 and PP2A B55 subunit, it will be interesting to adopt a similar *in vitro* approach as with USP7, consisting in a peptide array assay, followed by ITC validation of the hits and crystallographic structural studies.

We also plan on studying the full USP7-PP2A-GRA16 complex with electron microscopy. The sample will consist of the tagged recombinant GRA16 and endogenous USP7 and PP2A subunits co-purified along. The study of this complex stabilized by the *T. gondii* effector could bring insights to this newly described interaction between USP7 and PP2A.

Conclusions

Protein-protein interactions are key in cell signalling and responses to stress. This thesis focused on unravelling the details of a canonical interaction, discovered a non-canonical mechanism, and explored the hijacking strategies of cellular signalling networks by a pathogen.

- We presented the first structural model of the interaction between a MAPK and upstream MAP2K trapped in a putative transition state.
    - We obtained a MKK6/p38α model built from a cryoEM density map at 7.9 Å resolution.
    - We presented the global architecture of the dimer that allows the positioning of the p38α activation loop at the centre of the two kinases with relatively little interaction with the MKK6 active site, therefore enabling the presentation of both phospho-sites of p38α to MKK6, as well as the correct positioning of MKK6 catalytic residues for phosphorylation.
    - We described a novel interaction interface between the MKK6 αG-helix and the p38α hydrophobic pocket involving residues highly conserved amongst MAPK and MAP2K families.
    - We proposed the possible formation of a regulatory α-helix at the N-terminus of MKK6 that would interact with the αC-helix and participate to the MKK6 active conformation.
- Eukaryotic protein kinases use ATP to phosphorylate their targets. We described a novel non-canonical mechanism of phosphorylation of p38α by MKK6 using ADP as a source of phosphate.
    - We demonstrated that MKK6 can hydrolyse the β-phosphate of ADP to phosphorylate its substrate p38α.
    - We proposed that it could confer an adaptive advantage to maintain signalling active in stressed environment with limited ATP.
- The parasite *Toxoplasma gondii* secretes effector proteins to remodel its host cellular responses.
    - We showed that the GRA24 effector interacts with the MAPK ERK1/2 through a KIM motif, as it does with the other MAPK p38α.
    - We proposed that its mode of action on ERK1/2 could be downstream of ERK1/2 phosphorylation.
    - We presented that the GRA16 effector binds to the human USP7 TRAF-like domain through a MATH motif.

All these results participate in the better understanding of the control, use and misuse of cellular signalling networks.

Presentation of the results

The results obtained during this PhD have been presented to several local and international events, including:

- The Protein Kinases and Protein Phosphorylation Conference: Mechanisms to Therapeutics (Palm Spring, USA – 2019): poster and flash presentation, FASEB poster award
- PSB symposium “Macromolecules in action” (Grenoble, France – 2019): poster presentation
- Instruct Biennial Structural Biology Conference (Madrid, Spain – 2019): poster presentation
- Cryo-Electron Microscopy and 3D Image Processing (CEM3DIP 2018) EMBO course (Delhi, India – 2018): oral and poster presentations, and travel fellowship
- PSB student day (Grenoble, France - 2018): clip presentation award

I also presented my work in several outreach events including:

- Pint of Science festival (2018)
- Dr D. bar conference cycle (2018)
- Grenoble science festival “Parvis des sciences" (2018 and 2019)
- Primary school science day with the European Learning Laboratory for the Life Sciences ELLS (2019)

Additional work, not presented in this manuscript, led to the following publication:

> Naschberger Andreas, **Pauline Juyoux**, Jill von Velsen, Bernhard Rupp, and Matthew W. Bowler, ‘Controlled Dehydration, Structural Flexibility and Gadolinium MRI Contrast Compound Binding in the Human Plasma Glycoprotein Afamin’, *Acta Crystallographica Section D Structural Biology*, 75.12 (2019), 1071–83 <https://doi.org/10.1107/S2059798319013500>

Key resources table

	Origin	Used in chapter
Recombinant proteins (produced)		
ERK1 (FL 1-379)	plasmid from Genscript	III
human MKK6 S207D T211D GRA24KIM1 (referred to as MKK6DDGRA)	plasmid from Genscript	I
JNK1 (5-364)	plasmid from Genscript	III
Lambda phosphatase	Addgene plasmid #79748	I, II, III
MKK6DD	plasmid from Genscript	II
mouse MKK6 S207D T211D GRA24KIM1 (referred to as mMKK6DDGRA)	plasmid from Genscript	I
p38α (1-360)	plasmid from Genscript	I, II
p38α (10-360)	plasmid from Genscript	I, III
p38α K53R (1-360)	plasmid from Genscript	II
p38α T180V (1-360)	plasmid from Genscript	I
p38α T180V (10-360)	plasmid from Genscript	I
p38α Y182F (1-360)	plasmid from Genscript	I
p38α Y182F (10-360)	plasmid from Genscript	I
USP7 C-ter domain (560-1084)	plasmid from Genscript	IV
USP7 catalytical domain (208-560)	plasmid from Genscript	IV
USP7 N-ter domain (53-206)	plasmid from Genscript	IV
Recombinant proteins (purchased)		
FL recombinant human USP7	Sino Biologicals (ref 11681-H20B1)	IV
Peptides		
GRA16_192-206	LifeTein	IV
GRA16_196-206	LifeTein	IV
GRA16_TGGT1_100-118	LifeTein	IV
GRA16_TGGT1_104-114	LifeTein	IV
GRA16_TGGT1_129-138	LifeTein	IV
GRA16_TGGT1_192-210	LifeTein	IV
GRA16_TGGT1_196-210	LifeTein	IV
GRA16_TGGT1_220-254	LifeTein	IV
GRA16_TGGT1_256-270	LifeTein	IV
GRA16_TGME49/TGVEG_192-210	LifeTein	IV
GRA24 KIM1	LifeTein	III
MKK6 KIM	LifeTein	III
Reagents		
ADP	Sigma (ref A2754)	I, II
ADP β-^{32}P	Hartmann Analytic GmbH	II
$AlCl_3$	Sigma (ref 229393)	I

Alkaline phosphatase-conjugated goat secondary antibodies	Promega	III
Amicon membrane filters concentrators	Millipore	I, II, III, IV
AMP	Sigma (ref 01930)	II
AMP-CP	Sigma (ref M8386)	II
AMP-CPP	Sigma (ref M6517)	II
AMP-PCP	Sigma (ref M7510)	I, II
Ampicillin	Euromedex (ref EU0400-D)	I, II, III
Anti-FLAG M2 Magnetic Beads	Sigma	
ATP	Sigma (ref A7699)	I, II
ATP β-^{32}P	Hartmann Analytic GmbH	II
ATP γ-^{32}P	Perkin Elmer	II
Biotin	IBA (ref 2-1016-002)	I, II, III, IV
Carbon coated copper grid	Electron Microscopy Science (ref CF300-Cu-50)	I
Chloramphenicol	Euromedex (ref 3886-A)	I, II, III, IV
DMEM	Invitrogen	
DNaseI	Sigma (ref 10104159001)	I, II, III, IV
EGR1 antibody	Cell Signaling Technology	III
FBS	Invitrogen	
Galactose	Sigma (ref L3625)	I
Glutaraldehyde	Sigma (ref G5882)	I
H4 Acetylated antibody	EMD Millipore	III
HEK293 T-REx epithelial cells	RRID:CVCL_D585	
HisTrap column	GE Healthcare/Cytiva (ref 11773209)	I, II, III
InstantBlue	Abcam (ref ab119211)	I, II, III, IV
Kanamycin	Euromedex (ref EU0420)	I, II, III, IV
Lipofectamine Reagent	Invitrogen	
NBT-BCIP	Amresco	III
NH_4F	Sigma (ref 338869)	I
NucleoSpin Plasmid, Mini kit for plasmid DNA	Macherey-Nagel (ref 740588.50)	I, II, III, IV
Phospho-ERK MAPK antibody	Cell Signaling Technology	III
Phospho-p38 MAPK antibody	Cell Signaling Technology	III
Pierce protease inhibitor EDTA-free tablet	ThermoFisher (ref A32965)	I, II, III, IV
Pre-cast 4-20% gradient Tris-Glycine	ThermoFisher Scientific (ref XP04205BOX)	I, II, III, IV
Precision Plus Protein Dual Xtra standards	Biorad (ref 64300367)	I, II, III, IV
PVDF membrane	Immobilon-P; Millipore	II, III, IV
Rosetta™(DE3)pLysS competent cells	Novagen (ref 70956)	I, II, III, IV
Spectinomycin dihydrochloride pentahydrate	Sigma (ref S4014)	I, II, III
StrepTactin XT column	IBA (ref 2-5024-001)	I, II, III, IV
Superdex 75 10/300 GL column	GE Healthcare	I, II, III, IV
TCEP	Hampton (ref HR2-801)	I, II, III, IV
TgQRS antibody	van Rooyen *et al.*, 2014	III
Ultra-competent XL10 cells	Agilent Technologies (ref 200314)	I, II, III, IV
Ultra-Pure ADP	Cell Technology (ref ADP100-2)	II
UltraAufoil 1.2/1.3 grid	EMS (ref Q350AR13A)	I
Zebra spin desalting columns	Thermo Scientific (ref 898882)	I

Table of figures

Figure 1: The human kinome phylogenetic tree (adapted from Cell Signaling Technology poster). 22

Figure 2: The MAPK pathways. The canonical pathway starts with the integration of specific stimuli by the MAP3K by activator proteins and triggers the MAP3K activation by (auto)phosphorylation. The MAP3K phosphorylates its downstream MAP2K, that in turn phosphorylates the MAPK. The activated MAPK phosphorylates a number of substrates to induce the appropriate response to the initial stimulus. The main MAPK families, ERK1/2, p38s, JNKs and ERK5, are represented with their upstream activating MAP2Ks and MAP3Ks, as well as with a selection of stimuli, pathway activators and substrates. 23

Figure 3: Evolutionary tree of human MAPKs (A) and MAP2Ks (B). Divergent branch corresponding to the main yeast homologues are indicated. The sequence alignment was performed using Clustal Ω (Sievers and Higgins, 2018). p38α and MKK6, the main objects of study of this work, are coloured. 24

Figure 4: Diversity of p38α roles in health and disease, from (Canovas and Nebreda, 2021). 26

Figure 5: Structure-based sequence alignment of the kinases of interest in this work (ERK1, p38α and MKK6) relative to the reference kinase PKA. The alignment was published in Modi and Dunbrack, 2019a. Common kinase secondary structure with the kinase nomenclature are shown. Key conserved motifs and residues are highlighted in grey. Specific features of MAP2K and MAPK are indicated in teal. 27

Figure 6: Kinase structural elements. The double phosphorylated p38γ structure (PDB 1cm8) is coloured according to conserved structural features. The β-sheet containing N-lobe and the α-helical C-lobe are connected through a hinge (dark grey). In the N-lobe, the P-loop (pink) and the αC-helix (purple) are essential for ATP binding and active kinase conformation. In the C-lobe, the activation loop (red) is phosphorylated on both residues of the conserved MAPK TxY motif. It is preceded by the DFG motif, important for ATP binding. The MAPK insert (MKI in blue) is specific to the MAPK family. 28

Figure 7: The ATP binding site. The biphosphorylated p38γ structure bound to the ATP analogue AMP-PNP (PDB 1cm8) is coloured according to conserved structural features. The numbering is based on p38α sequence. Lys53 from the β3-strand and Glu71 from the αC-helix (in purple) form a salt-bridge. The AMP-PNP nucleotide (in light grey) binds to the ATP binding site, forming H-bonds with the hinge (grey). The phosphates are coordinated by two Mg^{2+} ions and several key residues: Lys53, Asp168 from the DFG motif (yellow), Asn155. The catalytic aspartate Asp153 from the HRD motif and the Lys155 are important for the binding of the phospho-acceptor substrate and the catalysis of the reaction. The P-loop is disordered in this structure (pink). The two phosphorylated residues Thr180 and Tyr182 in the activation loop (in red) interact with positively charged residues in both lobes, maintaining the close conformation. 29

Figure 8: Structure of the MKK6 KIM peptide bound to the docking site of p38α (PDB 2y8o). MKK6 KIM peptide is represented as cartoon with balls and sticks side chains. Left: The surface

of p38α is displayed coloured by its electrostatic potential (from -10 (red) to 10 kTe^{-1} (blue)). Right: p38α is represented as cartoon with key residues as sticks. 32

Figure 9: Life cycle of T. gondii adapted from (Hunter and Sibley, 2012). T. gondii presents different developmental stages along its lifecycle: non-sexual forms that can infect cells (sporozoite, tachyzoite, and bradyzoite), and sexual forms (gametes and oocyst). T. gondii definitive hosts are from the Felidae family (which includes the domestic cat): they are the only hosts in which the sexual stages of the parasite can occur. Most warm-blood animals can be infected though their environment and are considered as secondary or accidental hosts. Human infection usually occurs through the consumption of food or water contaminated by oocysts, or tissue cysts in undercooked meat of infected animals. The foetus can be contaminated by tachyzoites if the parent gets primarily infected during pregnancy. 39

Figure 10: Model of T. gondii tachyzoites infecting a cell from (Rastogi, Cygan and Boothroyd, 2019). T. gondii tachyzoites contain three types of secretory organelles: micronemes, rhoptries and dense granules that secrete the effector proteins MIC, RONs, ROP and GRA to invade the host cell cytosol within a parasitophorous vacuole and remodel the host cell transcriptional machinery.....40

Figure 11: MKK6 crystal structures. (A) The crystal structure of non-phosphorylated FL MKK6 (PDB 3vn9) reveals of a potential auto-inhibition conformation (Matsumoto et al., 2012). The activation loop is structured with three short α-helices referred to as activation helices 1-3 (AH1-3) and contribute to lock the γ-phosphate of the ATP analogue inside its active site. AH1 displace the αC-helix and position it away from the active site, preventing the formation of the salt bridge with the β3-strand that is essential for the active conformation of the kinase. (B) The crystal structure of ΔN44 MKK6DD (PDB 3enm), a constitutively active mutant (Ser2017 and Thr211 of the activation loop are mutated to aspartate to mimic phosphorylation) proposes another auto-inhibition mechanism (Min et al., 2009). Two MKK6 molecules form an elongated dimer bringing the two N-termini together. At the interface, the Glycine rich loop forms a distorted Ω loop that interacts with the other MKK6 monomer in its active site. (C) In the crystal structure of ΔN47 MKK6DD with the ATP-competitive inhibitor staurosporine (PDB 3fme) MKK6DD does not form a dimer. It is not in an active conformation either. The DFG motif seems to be in DFG-in conformation (the Gly is not visible in the model and the side-chains of the Asp and Phe are not modelled, but they adopt a similar conformation to the DFG-in in PDB 3enm). 48

Figure 12: Selection of p38 crystal structures (A) Apo, inactive and unphosphorylated p38α (PDB 5etc). (B) Unphosphorylated p38α bound to MKK6 KIM peptide (PDB 2y8o). (C) Unphosphorylated p38α bound to MKK3 KIM peptide (PDB 3p4y). (D) Unphosphorylated p38α bound to T. gondii GRA24 KIM peptide (PDB 5eta). (E) Biphosphorylated mouse p38α (PDB 3py3). (F) Biphosphorylated p38γ bound to AMP-PNP (PDB 1cm8). (G) Unphosphorylated salmon p38α dimer (PDB 3oht). 49

Figure 13: Final recombinant protein constructs designed for biochemical and structural studies. p38α constructs are tagged with an N-terminal His6 tag fused to a TEV or 3C protease cleavage site. Single point mutations (T180V and Y182F) were introduced. MKK6DDGRA constructs are composed of the GRA24 KIM1 sequence fused to the MKK6 sequence with phosphomimetic mutation on the activation loop (S207D T211D). The constructs are tagged with a C-terminal Twin StrepII tag linked attached to a 3C protease cleavage site. 61

Figure 14: SDS-PAGE gel of the mMKK6DDGRA/p38α complex obtained with the initial expression and purification strategy.....62

Figure 15: MKK6DDGRA/p38α complex purification. (A) MKK6DDGRA affinity purification with Streptactin. (B) p38α affinity purification with Ni IMAC. (C) MKK6DDGRA/p38α purification by size exclusion chromatography.....64

Figure 16: To trap the sample in an intermediate phosphoryl transfer conformation, we evaluated AMP-PCP non-cleavable ATP analogue, and ADP in combination with AlF_4^-. 65

Figure 17: native ESI-QTOF MS spectra of mMKK6DDGRA with p38α (A) WT, (B) T180V, and (C) Y182F samples incubated on ice with 10 mM AMP-PCP, ADP + AlF_4^-, ATP or without nucleotide, or incubated at RT with 10 mM ATP. For each sample, peak for each phosphorylation form (increased mass of 80 Da for each phosphorylation) is highlighted.....67

Figure 18: Phosphorylation state of p38α Thr180 and Tyr182 residues from the LC-MS/MS analysis of mMKK6DDGRA with p38α WT (A), T180V (B), and Y182F (C) samples incubated on ice with 10 mM AMP-PCP, ADP + AlF_4^-, ATP or without nucleotide, or incubated at RT with 10 mM ATP. For each analysed sample, the results of phosphorylation state of Thr180 (top) and Tyr182 (bottom) residues of p38α are represented in % of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (MASCOT phospho-score bellow 32) (orange) and phosphorylated (MASCOT phospho-score higher than 32) (red) by MASCOT analysis. 69

Figure 19: Ab initio model of the mMKK6DDGRA/p38α WT complex obtained from SAXS. Ab initio model (grey mesh) is represented with p38α (PDB 5eta, teal), MKK6 activated mutant (PDB 3fme, orange) and GRA24 KIM1 peptide (from PDB 5eta, red).....72

Figure 20: Example of one NS micrograph of the crosslinked mMKK6DDGRA/p38α complex sample taken on a FEI Tecnai T12 electron microscope operating at 120 kV at 49 kx magnification, equipped with a Ceta 16M camera.....74

Figure 21: 2D classification of NS EM particles of crosslinked mMKK6DDGRA/p38α complex. The number of particles per class is indicated (box size = 20,6 nm). 74

Figure 22: Different views of the 3D initial model of the crosslinked mMKK6DDGRA/p38α complex, fitting crystal structures of p38α (PDB 5ETA) in blue, GRA24 KIM peptide (PDB 5ETA) in red and MKK6 (PDB 3FME) in orange.....75

Figure 23: SAXS analysis of the MKK6/p38α complex. Comparison of combination with p38α WT or mutants and with different nucleotides analogues (AlF_4^- + ADP and AMP-PCP).....77

Figure 24: The MKK6/p38α T180 complex with ADP + AlF_4^- from the Glacios dataset (A) 2D classes. The two kinases in a face-to-face conformation can be observed in the class highlighted in red. The resolution was limited to 8 Å for alignment. (B) Projections through the four 3D classes generated by ab initio modelling from the selected 2D classes in (A). The orientation highlighted in red corresponds to the hetero-kinase dimer. (C) 3D density map obtained after the refinement of the class selected in (B) The crystal structures of p38α (teal) with GRA24 KIM peptide (orange) (PDB 5eta) and MKK6DD (orange) (PDB 3fme) were automatically fitted as rigid bodies into the density map using ChimeraX. 79

Figure 25: (A) Atlas of the grid indicating the region selected for data collection (B) CTF estimation (C) Micrograph with an estimated defocus of 1.5 µm (D) WARP picked 391 particles on the micrograph.....80

Figure 26: Examples of micrographs from the Krios data collection. Noticeable ice contamination can be observed in comparison to the micrographs from the same grid collected during the Glacios screening session (Figure 25).....81

Figure 27: Statistics of the final 3D model of the MKK6DDGRA/p38α T180V complex with ADP + AlF_4^- (A) 2D classes (B) FSC corrected resolution estimation is 7.93 Å based on gold standard calculation (C) Viewing direction distribution. 82

Figure 28: 3D model of the MKK6DDGRA/p38α T180V dimer. (A) The density map sharpened with DeepEMhancer reveals the position of secondary structure features. The crystal structures of p38α (in teal) with GRA24 KIM peptide (in orange) (PDB 5eta) and MKK6DD (in orange) (PDB 3fme) were refined in Phenix into the density map. The density map contour level is set to a low value and the α-helices are represented as cylinders for clarity. (B) The density map contour level is set to enclose 120 nm^3 (Porod volume from SAXS measurement). Two side views, top and bottom views are represented. 83

Figure 29: Novel interaction interface between MKK6 αG-helix and p38α hydrophobic pocket. (A) The density map contour level is lowered to a minimum, the density of the interaction interface between the two C-lobes, and more specifically around MKK6 Phe264 is persistent. (B) MKK6 αG-helix (in orange) forms a combination of hydrophobic interactions and hydrogen bonds with the p38α hydrophobic pocket (in teal). The main interactions are represented with thicker sticks. 84

Figure 30: MKK6 N-terminus potential conformations. The density map contour level is set to enclose 120 nm^3 (Porod volume from SAXS measurement). (A) Different possible conformations of the αC-helix of MKK6 observed in crystal structures are represented. Only the αC-helices and β3-strands are represented in colour (PDB 3enm in orange, PDB 3fme in salmon and PDB 3vn9 in yellow). (B) Superimposition between MKK6 (PDB 3enm) and MKK7 structures (PDB 6yg1). The GRA24KIM peptide (PDB 5eta) is represented. Only the αC-helices, the β3-strands, the regulatory α-helix and the KIM peptide are represented in colour (MKK6 and GRA24KIM in orange, MKK7 in green) The αC helices are overlapping, and the MKK7 regulatory helix fits into the density, weighting in favour for a structured N-terminal extension in our sample. (C) Generated model of MKK6 (res 31-338) based on MKK7 structure (PDB 6yg1) (in orange the kinase core and in green the newly modelled N-terminus). The green dashed line represent the 15 residues missing from the model. The conserved salt bridge between the αC-helix and β3-strand is displayed in grey. (D) Close view with low contour level of the cryoEM density map. The arrow points to the density supporting the presence of a N-terminal regulatory α-helix in MKK6. 86

Figure 31: p38α activation loop is ordered and oriented towards the active site of MKK6. The contour level of the cryoEM density map are adjusted to visualise the activation loop (A) Activation loop of the non-phosphorylated p38α (PDB 5etc). (B)) Activation loop of the biphosphorylated mouse p38α (PDB 3py3). (C) Non-phosphorylated p38α activation loop refined into the cryoEM density map. The valine is shown as a threonine to illustrate the WT situation. 87

Figure 32: MKK6 active site. The MKK6 homology model based on the crystal structure of active MKK7 (PDB 6yg1) is represented as cartoon and sticks in orange. The p38α activation loop from the refined model is represented in teal. 88

Figure 33: Comparison of kinase-substrate binding to the kinase active site. (A) CryoEM structure of MKK6DDGRA (in orange) and p38α T180V activation loop (in teal). (B) crystal structure of CDK2 and an optimal peptide substrate (PDB 1gy3). (C) Crystal structure of PKA and PKI-α peptide (PDB 4iaz). 89

Figure 34: Comparison of the MKK6-p38α dimer with other kinase dimers. For the asymmetric homodimers (A, B and C), the "enzyme" kinase is represented in dark color and the "substrate" kinase in light color. The αG helix of the "enzyme" kinases are indicated by an arrow. The right panel shows the overlap of the published kinase dimer structures with our structure of the MKK6 (in orange) – p38α (in cyan) dimer. 91

Figure 35: Comparison of the density maps of the MKK6DDGRA/p38α complex successively obtained by SAXS, negative stain electron microscopy and cryoEM. The density map contour levels are set to enclose 120 nm^3 (Porod volume from SAXS measurement). 92

Figure 36: Model of the MKK6DDGRA/p38α T180V complex in the cryoEM density map at 7.9 Å. 94

Figure 37: Focus on the interaction between MKK6 αG-helix and p38α hydrophobic pocket (A) Cartoon and surface representation of four p38α structures (apo p38α (PDB 5etc), p38α bound to MKK6 KIM (pdb 2y8o), p38α bound to BOG (PDB 2npq) and salmon p38α (PDB 3oht)). The surface is coloured by hydrophobicity (cyan for most hydrophilic to yellow for most hydrophobic). An arrow points to the open hydrophobic pocket. (B) Sequences of human MAPKs and MAP2Ks aligned with Clustal Ω. The residues involved in the interaction are highly conserved. (C) Comparison between the cryoEM structure of MKK6DDGRA/p38α T180V and the crystal structures of p38α interacting with a peptide from its substrate transcription factor ATF2 (PDB 6zqs), and JNK1 interacting with MKP7 (PDB 4yr8). The interaction between the MAPK Tyr and a substrate Phe is conserved. 96

Figure 38: Nucleotides and their analogues. 104

Figure 39: Radiolabelled nucleotides. 105

Figure 40: Native PAGE gel of MKK6DD + p38α WT in presence of 10 mM ATP, ADP, or without nucleotides. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P and triphosphorylated 3P). 106

Figure 41: LC-MS/MS analysis of bands extracted from a Native PAGE gel of MKK6DD + p38α WT in the presence of 10 mM ATP, ADP or without nucleotides. The results are displayed in the same layout as the gel. For each analysed band, the results of the phosphorylation state of Thr180 (top) and Tyr182 (bottom) of p38α are represented as a percentage of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (orange) and phosphorylated (red) by MASCOT analysis. Bands that were significantly less abundant are represented in faint colours. 107

Figure 42: Native PAGE gel of MKK6DD + p38α WT in presence of ATP, ADP, or without nucleotides. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P and triphosphorylated 3P). 108

Figure 44: Native PAGE gel of MKK6DD + p38α WT with increasing concentrations of ATP and ADP. MKK6DD bands are smeared. p38α bands are well defined and run based on the total

phosphorylation number of the protein (not phosphorylated 0P, monophosphorylated 1P, biphosphorylated 2P, and triphosphorylated 3P).....109

Figure 43: Quantification of p38α bands from the native PAGE gel of Figure 44. 109

Figure 45: Phosphorylation sites of MKK6DD and p38α detected by MS and represented on the cryoEM model of MKK6DDGRA/p38α T180V presented in Chapter I. The phosphorylated sites are represented as balls (their size (small, medium and big) corresponds to the relative proportions of the phospho-site observed in MS). 111

Figure 46: native ESI-QTOF MS spectra of MKK6DD with p38α WT (A) and K53R (B) samples in the presence of 10 mM AMP, ADP, uADP, ATP or without nucleotide. For each sample, peak for each MKK6DD and p38α phosphorylation form is highlighted.....112

Figure 47: Phosphorylation state of p38α Thr180 and Tyr182 from the LC-MS/MS analysis of MKK6DD with p38α WT (A) and K53R (B) samples in presence of 10 mM AMP, ADP, uADP, ATP or without nucleotide. For each analysed sample, the results of phosphorylation state of Thr180 (top) and Tyr182 (bottom) of p38α are represented in % of peptide count detected as not phosphorylated (yellow), potentially phosphorylated (orange) and phosphorylated (red) by MASCOT analysis. 113

Figure 48: Native PAGE gel of MKK6DD + p38α WT in the presence of 10 mM ATP, ADP, AMP-CPP or without nucleotide. MKK6DD bands are smeared. p38α bands are well defined and run based on the total phosphorylation number of the protein (non-phosphorylated 0P, monophosphorylated 1P and biphosphorylated 2P). 114

Figure 49: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with 1 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and run on SDS-PAGE gel. The corresponding SDS-PAGE gel was stained after exposure and data acquisition as control. 115

Figure 50: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT or K53R were incubated with 10 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and run on SDS-PAGE gel. 116

Figure 51: In vitro phosphorylation assay with radiolabelled nucleotides and different cold nucleotide concentrations. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with (A) 10 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and (B) with only radiolabelled nucleotide, and run on SDS-PAGE gel. 116

Figure 52: Influence of cold nucleotide concentration on the incorporation of radiolabelled phosphorus. In vitro phosphorylation assay with radiolabelled nucleotides and increasing cold nucleotide concentrations. Samples of MKK6DD, p38α WT and MKK6DD + p38α WT were incubated with (A) cold ATP with γ-^{32}P ATP, (B) cold ATP with β-^{32}P ATP, (C) cold ADP with β-^{32}P ADP, and (D) cold ATP with β-^{32}P ADP, and run on SDS-PAGE gel. 118

Figure 53: In vitro phosphorylation assay with radiolabelled nucleotides. Samples of MKK6DD + p38α WT, or RIP2K were incubated with 1 mM cold nucleotide supplemented with radiolabelled nucleotide (ATP with either γ-^{32}P ATP or β-^{32}P ATP, ADP with β-^{32}P ADP), and run

on SDS-PAGE gel. The corresponding SDS-PAGE gel was stained after exposure and data acquisition as control. 119

Figure 54: Comparison of kinases with their substrate and nucleotide. In the top panel the kinases are represented as cartoon, the substrates and nucleotides as sticks. The bottom panel is a zoom on the kinase active site, with only key kinase residues displayed as sticks, the substrates and nucleotide s are represented as sticks. (A) MKK6DDGRA + AMP-PNP + p38α activation loop substrate (B) Thermococcus litoralis ADP-dependent glucokinase + ADP + α-D-glucopyranose substrate (PDB 1gc5) (C) human ATP-dependent glucokinase + AMP-PNP + β-D-glucopyranose substrate (PDB 3fgu). 121

Figure 55: GRA24 and its interaction with p38α (A) Diagram of GRA24 domains (SP signal peptide, NLS nuclear localization signal, R1 and R2 repeats containing KIM1 and KIM2 motifs) with corresponding IUPred2 (intrinsically disordered regions prediction) and ANCHOR2 (prediction binding regions that can undergo disorder-to-order transition upon binding) scores (B) GRA24 KIM class description and KIM1/2 sequences (C) Crystal structure of GRA24 KIM1 peptide bound to the docking site of p38α (PDB 5eta). The surface of p38α is displayed coloured by its electrostatic potential (from -10 (red) to 10 kTe^{-1} (blue)). GRA24 KIM peptide is represented as cartoon with sticks side chains. 127

Figure 56: SDS-PAGE gels of purified MAPKs (A) ERK1 (B) p38 (C) JNK1 134

Figure 57: Typical ITC results of binding of GRA24 KIM1 peptide to ERK1, p38α and JNK1. . 135

Figure 58: Comparison of ITC results of binding of GRA24 KIM1 peptide to ERK1 and p38α. The plot shows the mean and sd of 3 replicates for each sample. 135

Figure 59: Spherulites of ERK1 with bound GRA24 KIM peptide sample with corresponding UV signal. Sample at 10 mg/ml, setup at 20°C. Crystallization condition: 10% dioxane, 1.6 M ammonium sulfate, 0.1 M MS pH 6.5. 136

Figure 60: Crystals of ERK1 with bound GRA24 KIM1 peptide sample with corresponding UV signal. Sample at 13.1 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 6.5, 10% dioxane, 1.8 M ammonium sulfate. 137

Figure 61: Crystals of ERK1 with bound GRA24 KIM1 peptide sample. Sample at 15 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 6.5-7.5, 10% dioxane, 1.8 M ammonium sulfate. 138

Figure 62: Crystal of ERK1 with bound GRA24 KIM1 peptide that diffracted to a resolution of 2.4 Å. Sample at 17 mg/ml, setup at 20°C. Crystallization condition: 0.1 M MES pH 7, 10% dioxane, 1.8 M ammonium sulfate. 138

Figure 63: Structure of ERK1 bound to GRA24 KIM1 peptide (A) Simulated annealing omit map electron density (in green) in which the GRA24 KIM1 peptide was built (in orange). (B) The structure of ERK1 is shown as a solvent accessible surface coloured by electrostatic potential (-5 to 5 kTe^{-1}, red to blue) and the GRA24 KIM1 peptide is shown in cartoon representation (in grey). 141

Figure 64: Conformational change induced by GRA24 KIM1 binding. The N-terminal lobe rotates towards the active conformation of the kinase. The activation loop (indicated by dashed arrows) is embedded in the inactive conformation (PDB 5umo), ordered and accessible to solvent in the active conformation (PDB 2erk), and disordered when in complex with GRA24 KIM1 peptide 142

Figure 65: Comparison of the binding mode of KIM peptide to the docking site of ERK1, p38α (PDB 5eta) and JNK1 (PDB 4ux9). (A) GRA24 KIM peptide binds to the docking site of ERK1 through hydrogen bonds with the CD domain and through hydrophobic interactions. (B) GRA24 KIM peptide binds similarly to the conserved docking site of ERK1 and p38α. (C) ERK1 and JNK1 present different docking site conformations, explaining the different binding modes of KIM peptides. 144

Figure 66: Human cell HFF and ARP19 were infected with T. gondii RHku80 or RHku80Δgra24. Cytosolic and nuclear extracts were blotted with phospho-antibodies of ERK1/2 and p38, and controls (anti-AcH4 as nuclear marker and anti-QRS as infection marker). 145

Figure 67: Diagram of GRA16 domains (SP signal peptide, NLS nuclear localization signal, R1 and R2 repeats 1 and 2) with corresponding IUPred2 (intrinsically disordered regions prediction) and ANCHOR2 (prediction binding regions that can undergo disorder-to-order transition upon binding) scores. The two regions identified by Bougdour e al. (2013) to interact with USP7 and PP2A-B55 are indicated. 149

Figure 68: The USP7-MDM2-p53 axis (A) In normal cells, p53 has a short half-life. It is maintained at low levels by an equilibrium between ubiquitination by the E3 ligase Mdm2 and subsequent 26S proteolysis, and deubiquitination by USP7 that promotes its stabilization. Mdm2 also self-ubiquitinates leading to its degradation but can be rescued and stabilized by USP7 deubiquitination. (B) Under stress conditions, USP7 affinity for Mdm2 is lower, resulting in Mdm2 degradation and subsequent p53 accumulation. p53 acts as a transcription factor and induces cell cycle arrest and apoptosis. 150

Figure 69: USP7 domains and structures (A) USP7 domains (B) Model of the USP7 structure from combination of PDB 2f1z, 2ylm and 4wph (C) USP7 TRAF-like domain bound to MATH peptides from human interacting partners p53 (PDB 2foo) and MDM2 (PDB 2fop), and viral interacting partners EBNA1 (PDB 1yy6) 151

Figure 70: Assay principle. JPT's Peptide Microarrays were designed to elucidate protein-protein interactions. Each spot in the microarray represents a single individual peptide. After a direct labelling of the protein of interest and its incubation on the peptide microarray, the latter can directly be detected by measuring fluorescent signals. Resulting binding signatures represent the unique insights into the individual binding events. 156

Figure 71: Linear docking motifs described to bind to USP7. 160

Figure 72: Heatmap diagram (log) showing all incubations of labelled Human USP7 and all displayed GRA16 peptides on slide. The MMC2 values are shown as colours; y-axis represents peptide sequences in the library, x-axis shows samples applied. Colour coding ranges from white (0 or low intensity) over yellow (middle intensity) to red (high intensity), blue lines separate serial dilutions of each sample incubated twice. All signals were normalized towards the highest signal measured. 161

Figure 73: GRA16_TGGT1 sequence with potential linear motifs identified by elm.eu.org (in green and blue) and potential binding regions identified by peptide microarrays (in orange). Underlined in dash is the sequence (58-168) necessary for USP7 binding identified by Bougdour et al. (2013). 162

Figure 74: SDS PAGE gels of purified USP7 domains 164

Figure 75: Typical ITC result of the binding between GRA16_196-206 and USP7 N-ter domain. 167

Figure 76: Crystalline objects of USP7 GRA16. Crystallization condition: 0.1 M sodium acetate pH 5, 10 mM Zinc chloride, 20% PEG 6000. 168

Figure 77: Crystal form 1 of USP7 N-ter domain (53-206) with GRA16_196-206 peptide. Crystallization condition: 0.1 M sodium acetate pH 4.5, 10 mM Zinc chloride, 5% PEG 6000. 169

Figure 78: Crystal form 2 of USP7 N-ter domain (53-206) with GRA16_196-206 peptide. Crystallization condition: 0.1M sodium acetate pH 5, 10mM Zinc chloride, 5% PEG 6000...170

Figure 79: X-ray diffraction images from crystal form 1 (left) and crystal form 2 (right). 171

Figure 80: Alignment of the pairs of USP7 protein - GRA16 peptide from crystal form 1 and 2. 173

Figure 81: Crystal structure of USP7 N-terminal domain (53-206) bound to the GRA16_196-206 peptide (crystal form 2, chains A and B). (A) Interactions between GRA16 peptide (as sticks in yellow) and USP7 (as cartoon and surface in grey) (B) Omit map electron density (in green) in which the GRA16 peptide was built. (C) Detailed interactions between the GRA16_196-206 peptide (in yellow) and USP7 (in grey). (D) Comparison with EBNA1 peptide (in purple) binding to USP7 (PDB 1yy6). 174

Figure 82: Interaction between GRA16 peptide and symmetry and non-symmetry related USP7 N-terminal domain in crystal form 1. (A) GRA16 Arg204 from chain H stabilizes USP7 loop (aa 106-112) from non-symmetry related chain A through H-bond with Asp108 (B) GRA16 Ser196 coordinate symmetry related USP7 chain A and non-symmetry related USP7 chains E and G at the crystal interface. 175

Figure 83: Comparison of the crystal packing of the two crystal forms (crystal form 1 in red, orange, magenta and yellow; crystal form 2 in cyan and teal). 176

Figure 84: Comparison of the binding mode of MATH binding peptides (in stick representation, coloured according to table C) to the USP7 N-terminal domain (in cartoon and surface representation in grey) (A) GRA16 and viral interactors EBNA1 (PDB 1YY6), vIRF1 (PDB 4YSI), and vIRF4 (PDB 2XXN) (B) GRA16 and human interactors p53 (PDB 2FOO), MDM2 (PDB 2FOP), and MDMX (PDB 3MQR) (C) Alignment of MATH binding peptide sequence of GRA16, EBNA1, vIRF1, vIRF4, p53, MDM2 and MDMX. 177

Figure 85: The MATH2 motif of GRA16 is necessary for binding with USP7. T-REx cells were transfected with different GRA16 constructs. GRA16-HAFlag was immunoprecipitated from T-REx cells with anti-Flag antibodies and analysed by western blotting. 178

Figure 86: Diagram comparing our results with the results published in Bougdour et al. 2013. 179

Figure 87: Alignment of MATH binding motifs. 180

Figure 88: Possible mechanism of inhibition of USP7 by GRA16. USP7 inhibition could maintain low levels of p53 in infected cells despite the stress induced by infection and promote host cell survival.....182

Table of tables

Table 1: KIM classes 33

Table 2: T. gondii effectors, their targets and effects in host cells 42

Table 3: Recombinant protein constructs 52

Table 4: Parameters tested for sample preparation and crystallisation screens 71

Table 5: SAXS data analysis 72

Table 6: SAXS analysis of the mMKK6DDGRA/p38α complexes. Comparison of combination with p38α WT or mutants and with different nucleotides analogues (AMP-PCP and AlF_4^- + ADP). 76

Table 7: Recombinant protein constructs 103

Table 8: Recombinant protein constructs 130

Table 9: Peptides 131

Table 10: binding affinity measured by ITC 134

Table 11: Data collection and refinement statistics. Values in parentheses are for the outer resolution shell. 139

Table 12: USP7 constructs 154

Table 13: GRA16 peptides. When the strain is not specified, the peptide sequence is conserved amongst the 3 strains TGGT1, TGME49 and TGVEG. 155

Table 14: ITC results of the interaction between GRA16 peptides and USP7 domains. Ø = no binding, / = not tested, rep = number of replicates. When affinity was detected, results present the mean and standard deviation of the K_D. 166

Table 15: Data collection and refinement statistics 172

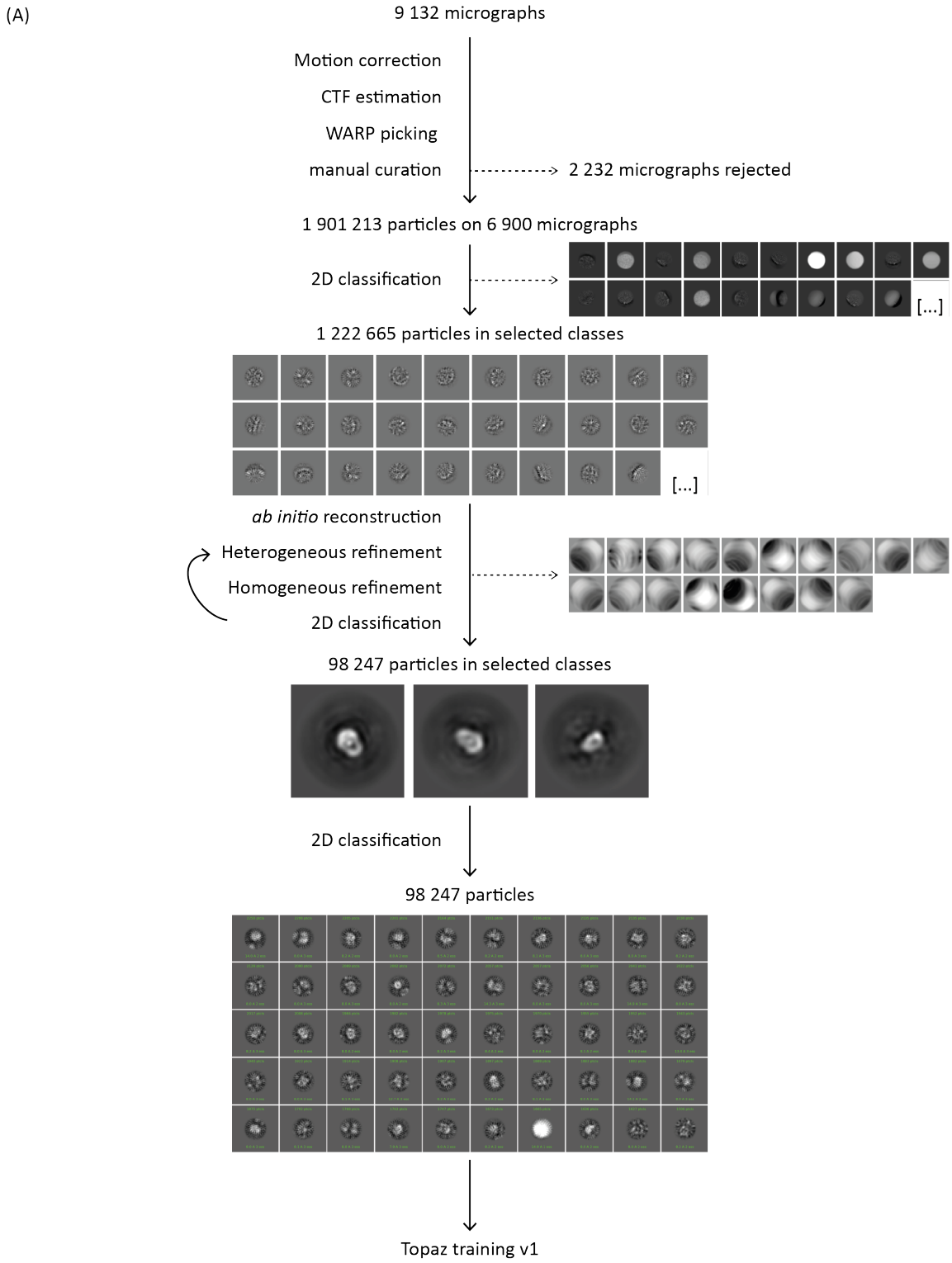
List of abbreviations

aa	amino acid
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
CCP4	Collaborative Computational Project Number 4
CD domain	common docking domain
CryoEM	Cryo electron microscopie
CSAID	cytokine-suppressive anti-inflammatory drug
CSBP2	CSAID binding protein
D-motif	Docking motif
DLS	Diamond Light Source
DUSP	Dual-specificity phosphatase
DVD	domain of versatile docking
EBNA1	Epstein–Barr nuclear antigen 1
ELM	Eukaryotic Linear Motif
ERK	extracellular signal-regulated kinase
ESRF	European Synchrotron Radiation Facility
FL	Full length
GPCR	G-protein-coupled receptor
GRA	dense granule proteins
HAUSP	herpesvirus associated ubiquitin-specific protease
HCC	hepatocellular carcinoma
HFF	Human foreskin fibroblast
HOG1	High osmolarity glycerol 1
IAB	Institute for Advanced Biosciences
IDR	intrinsically disordered region
ITC	Isothermal Titration Calorimetry
JNK	c-Jun amino-terminal kinase
KIM	Kinase Interacting Motif
KO	Knock-out
KSHV	Kaposi Sarcoma Herpesvirus
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography coupled to tandem Mass Spectrometry
MAP2K	mitogen-activated protein kinase kinase
MAP3K	mitogen-activated protein kinase kinase kinase
MAPK	mitogen-activated protein kinase
MIC	microneme protein
MKI	MAPK insert
NLS	nuclear localization sequence
PDB	Protein data bank
PEG	Polyethylene glycol
PRMT1	protein arginine methyltransferase 1
PTEN	phosphatase and tensin homolog
PTM	post-translational modification
PV	parasitophorous vacuole
Q-TOF MS	Quadrupole time-of-flight Mass Spectrometry

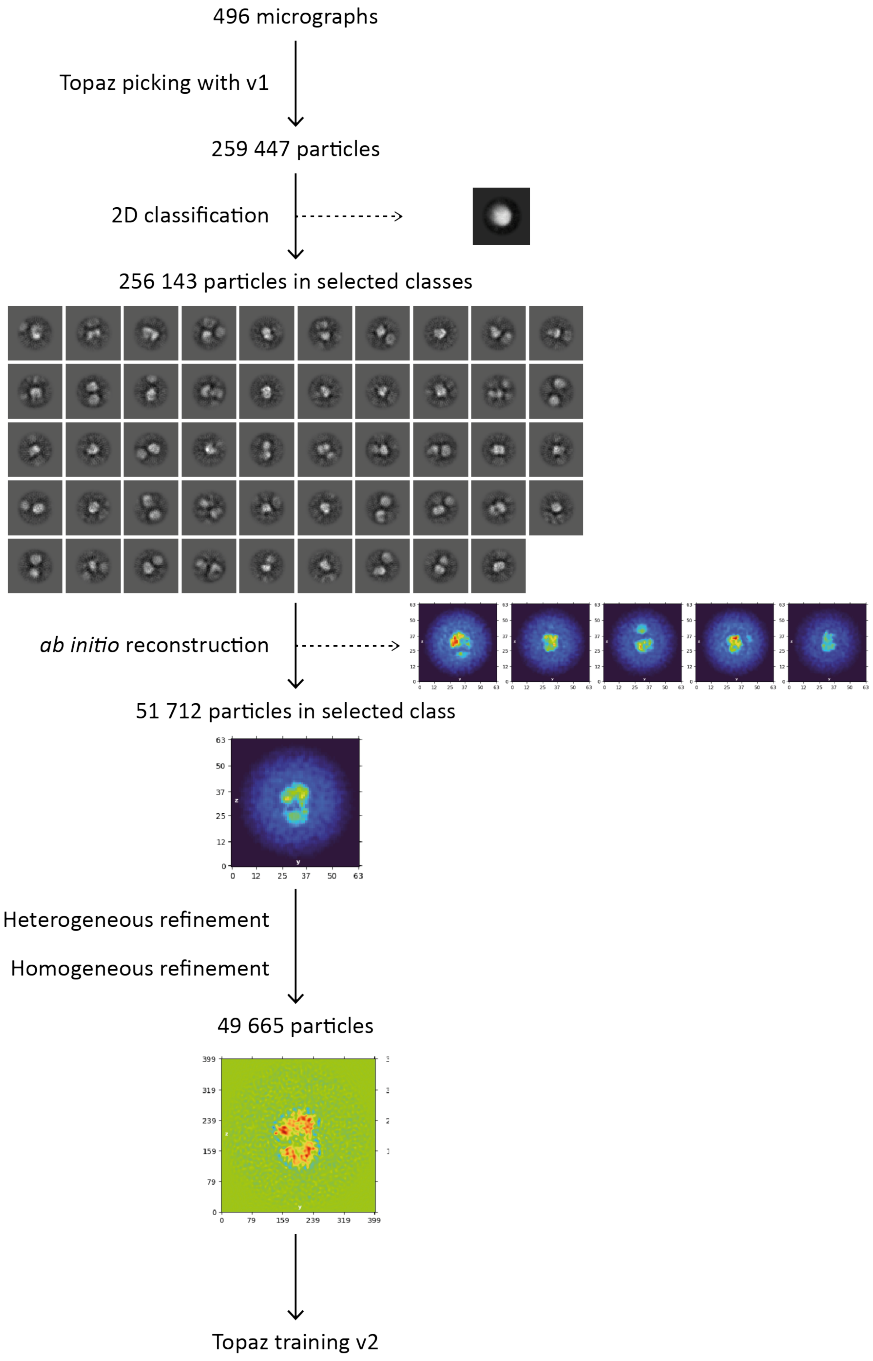
Relion	REgularised LIkelihood OptimisatioN
RIP2	Receptor Interacting protein 2
RK	MAPKAP kinase-2 reactivating kinase
RON	rhoptry neck protein
ROP	rhoptry bulb protein
RPE	retinal pigment epithelia
SAXS	Small Angle X-ray Scattering
SLiM	short linear motif
TAB1	TAK1-binding protein
TAK1	TGF-β-activated kinase 1
TCR	T-cell receptor
TKL	Tyrosine kinase-like
TLR	Toll-like receptor
TNF	Tumor necrosis factor
TRAF	TNF receptor-associated factors
Ubl	Ubiquitin-like domain
USP7	Ubiquitin-specific protease 7
VEGF	vascular endothelial growth factor
WT	Wild-type
β-OG	n-octyl-β-D-glucoside

Appendix

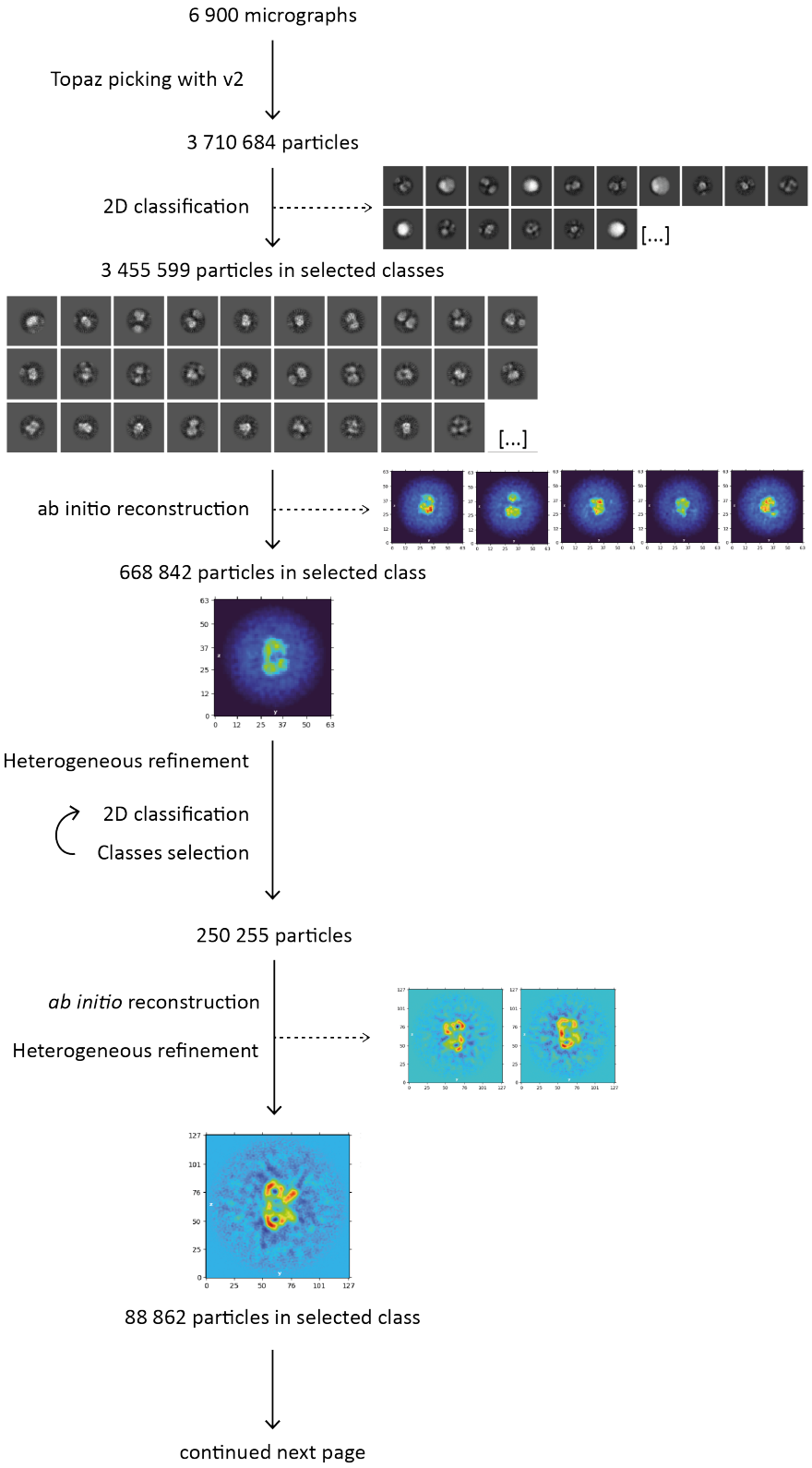
Appendix 1: CryoEM data processing of the MKK6DDGRA24/p38α T190V with ADP +AlF_4^- complex in CryoSPARC (A) Initial processing to obtain a set of particles that potentially contain the complex to train Topaz particle picker (B) Topaz particle picker optimization (C) Processing of full particle data set picked with optimized Topaz particle picker.



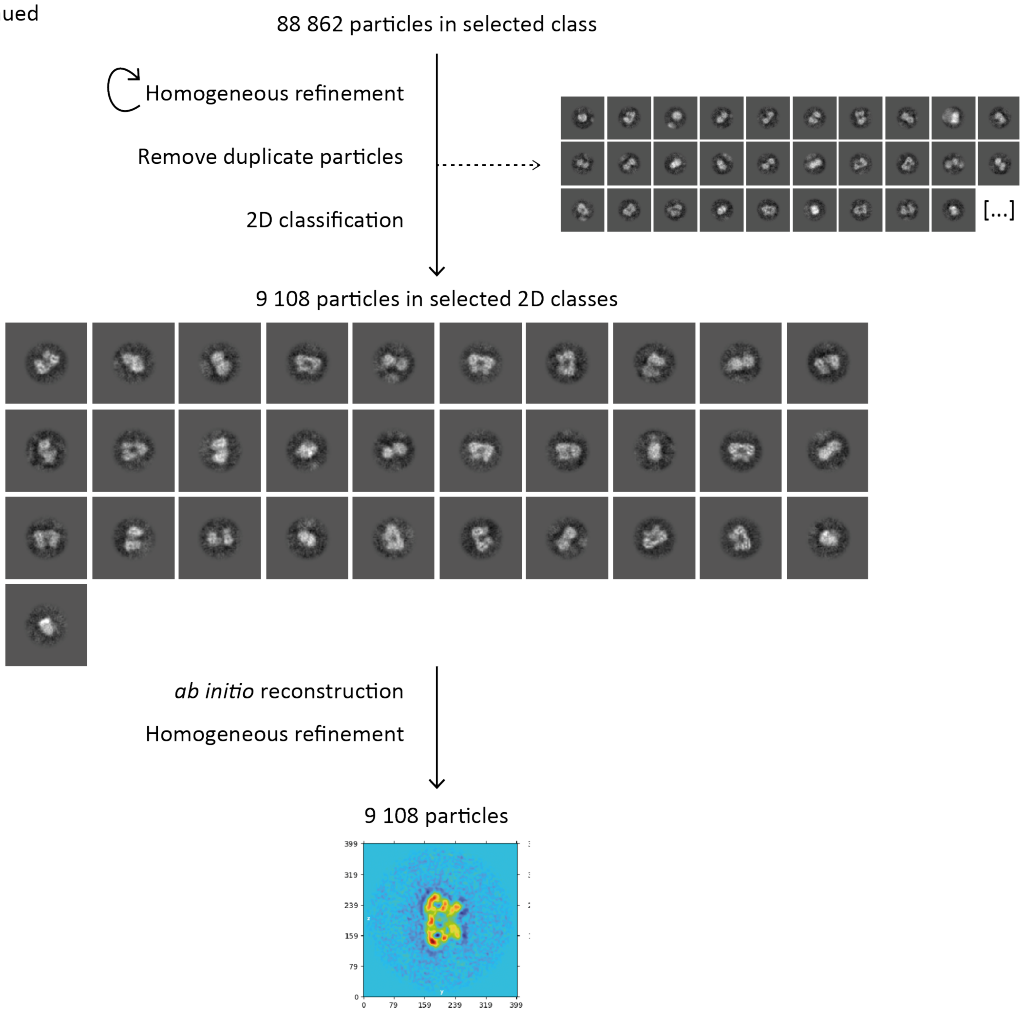
(B)


496 micrographs
Topaz picking with v1
259 447 particles
2D classification
256 143 particles in selected classes
ab initio reconstruction
51 712 particles in selected class
Heterogeneous refinement
Homogeneous refinement
49 665 particles
Topaz training v2

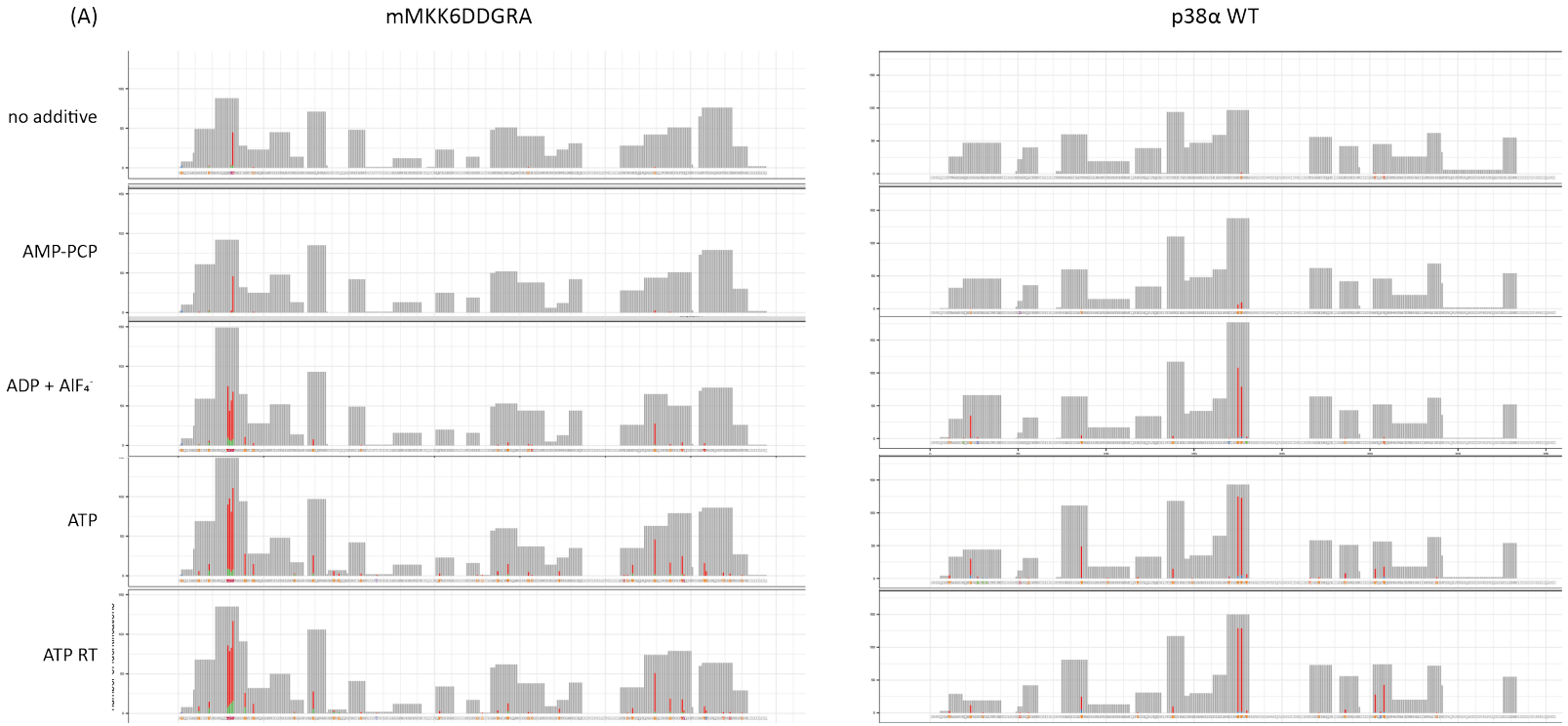
(C)

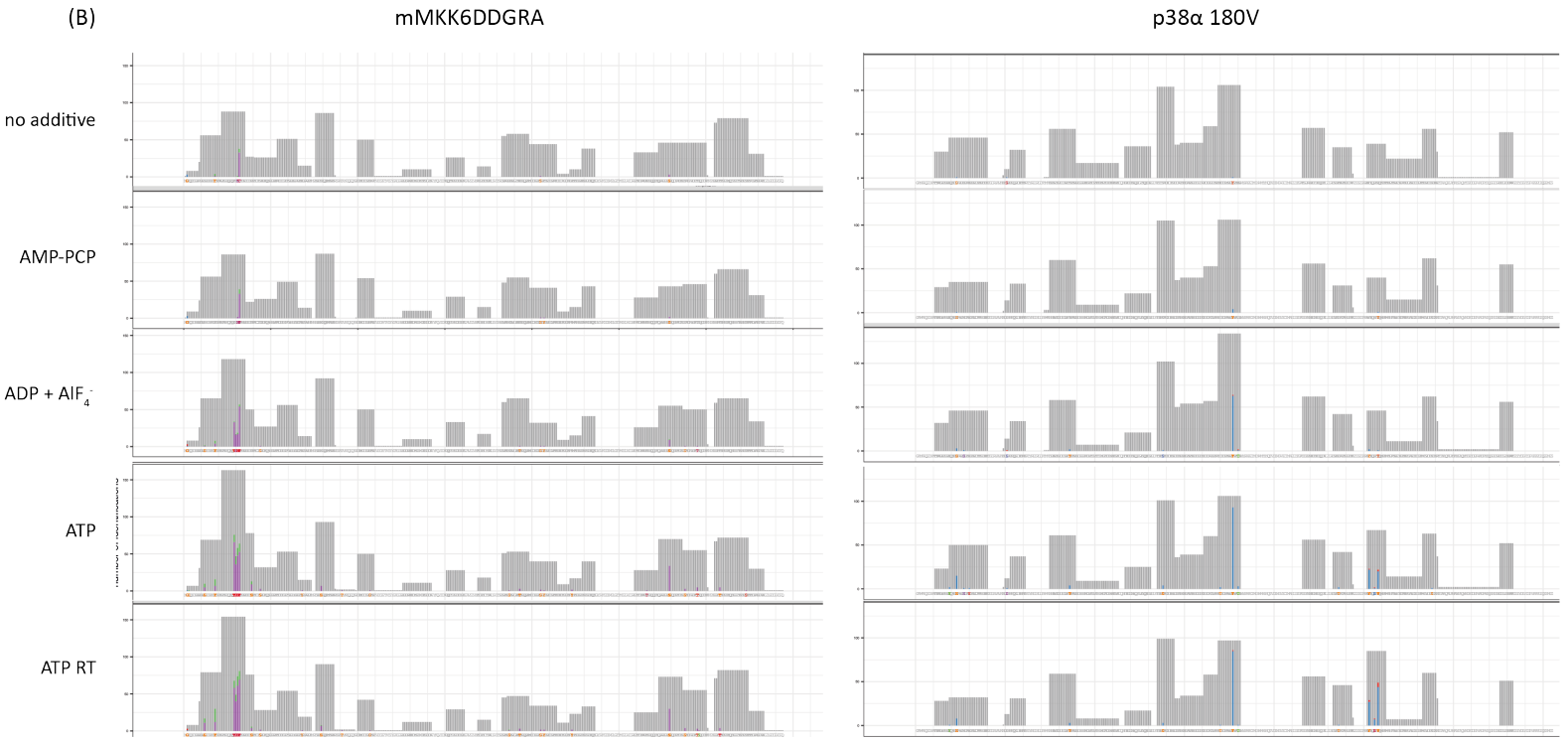

6 900 micrographs
Topaz picking with v2
3 710 684 particles
2D classification
[...]
3 455 599 particles in selected classes
[...]
ab initio reconstruction
668 842 particles in selected class
Heterogeneous refinement
2D classification
Classes selection
250 255 particles
ab initio reconstruction
Heterogeneous refinement
88 862 particles in selected class
continued next page

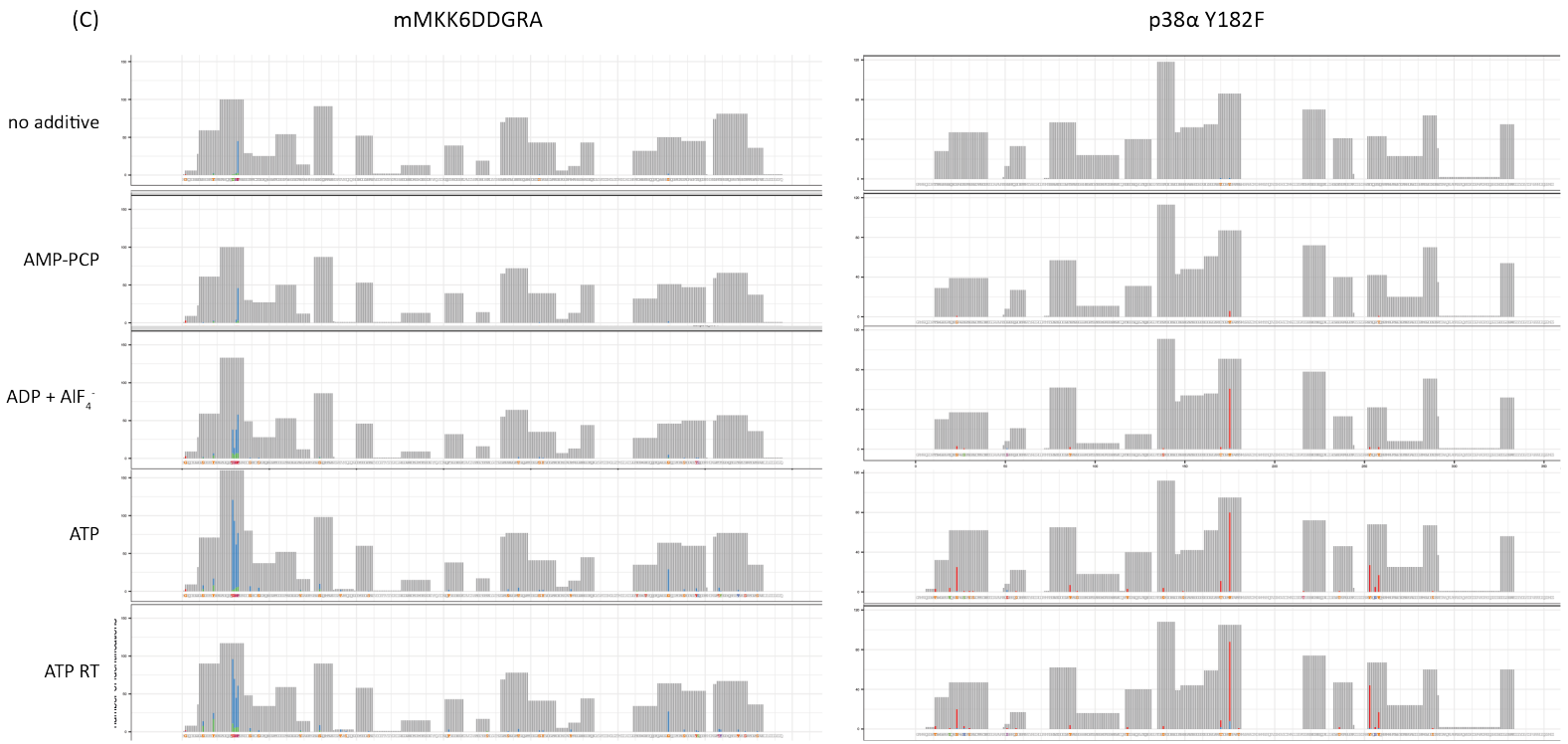
(C) continued


88 862 particles in selected class
Homogeneous refinement
Remove duplicate particles
2D classification
[...]
9 108 particles in selected 2D classes
ab initio reconstruction
Homogeneous refinement
9 108 particles
399
319
239
159
79
0
0
79
159
239
319
399

Appendix 2: LC-MS/MS analysis MKK6DD with p38α (A) WT, (B) T180V, and (C) Y182F samples incubated on ice with 10 mM AMP-PCP, ADP + AlF_4^-, ATP or without nucleotide, or incubated at RT with 10 mM ATP. For each analysed sample, the number of identifications of mMKK6DDGRA and p38α peptides is plotted over the amino acid sequence. Counts are coloured by their MASCOT analysis results: not phosphorylated (grey), and phosphorylated (coloured).

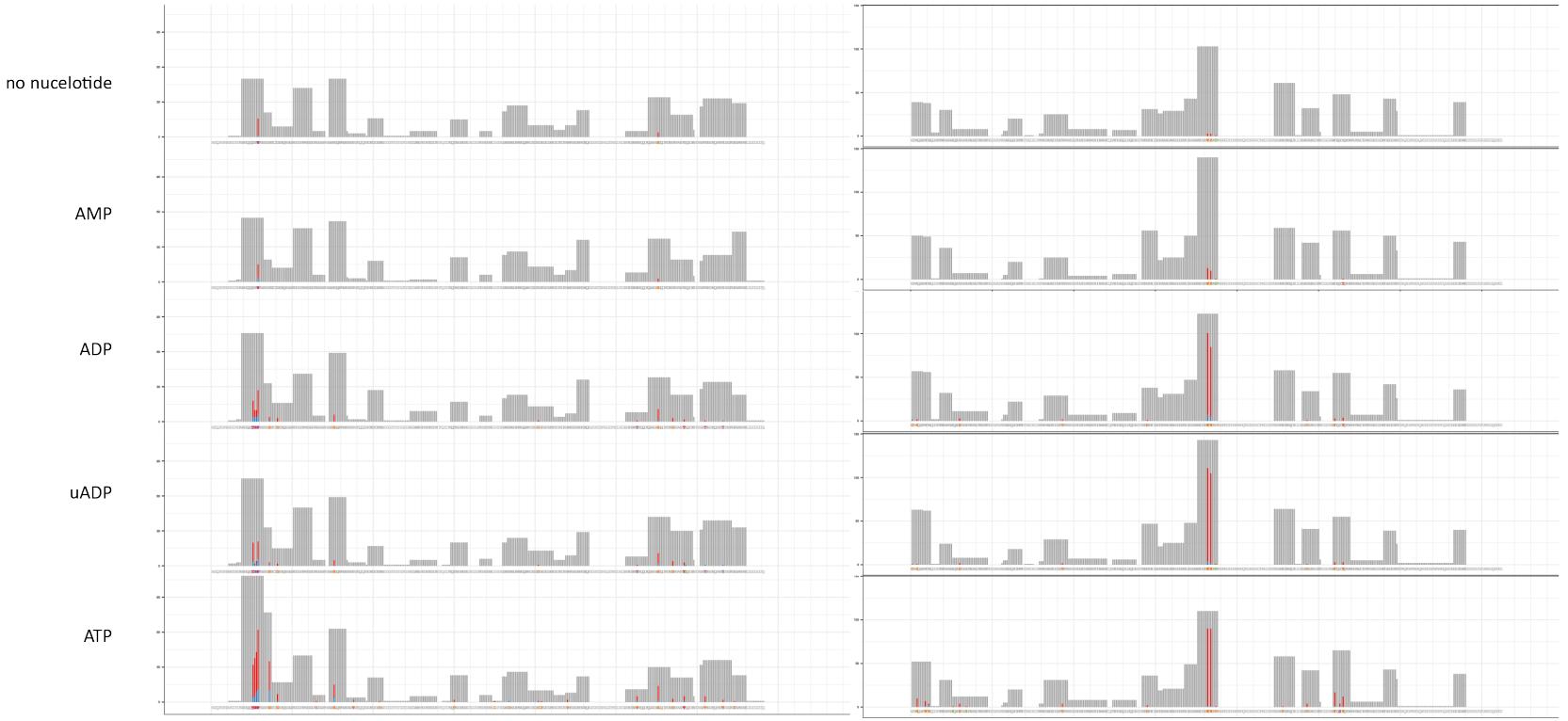


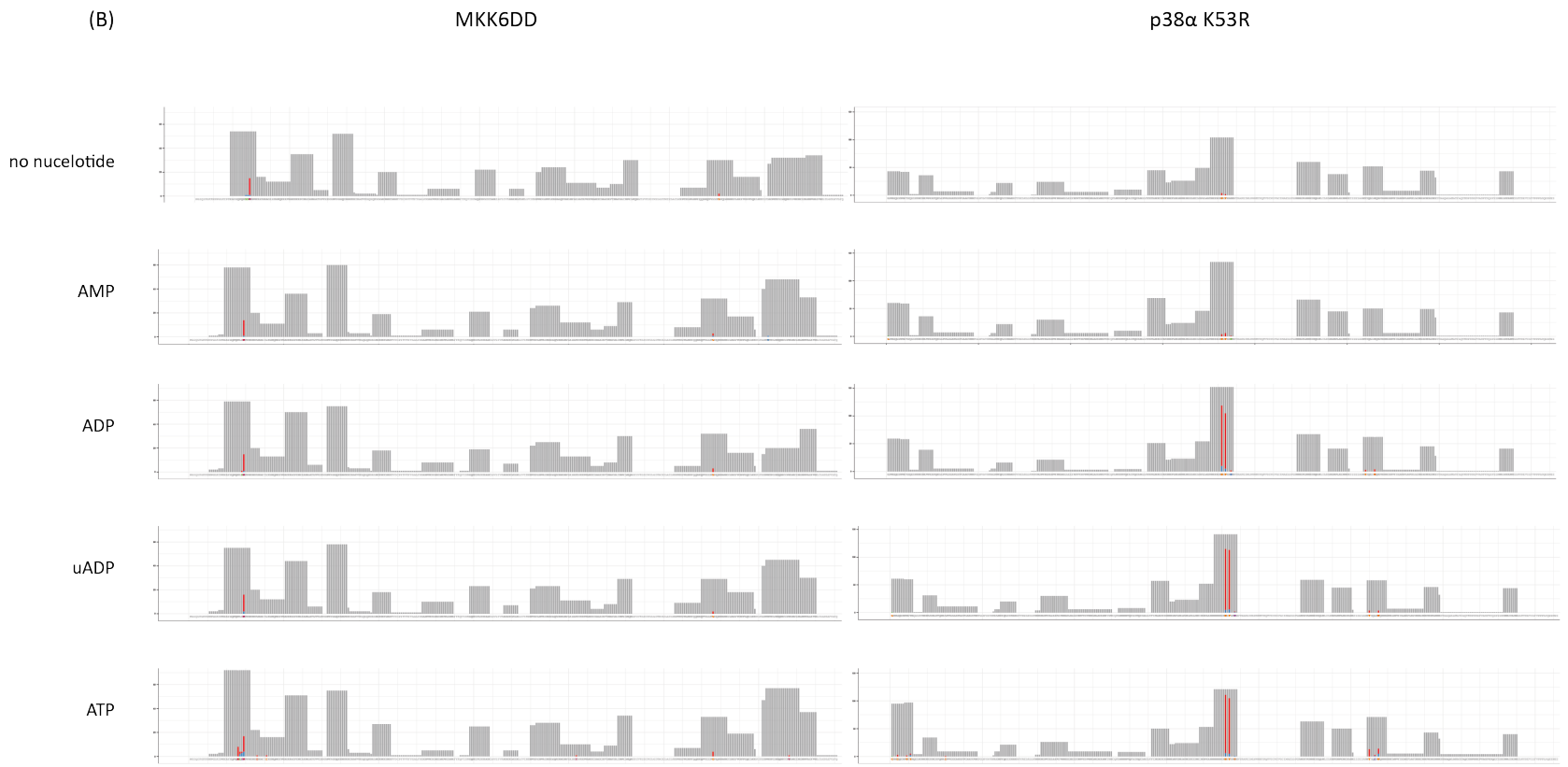

(B)
mMKK6DDGRA
p38α 180V
no additive
AMP-PCP
ADP + AlF4-
ATP
ATP RT

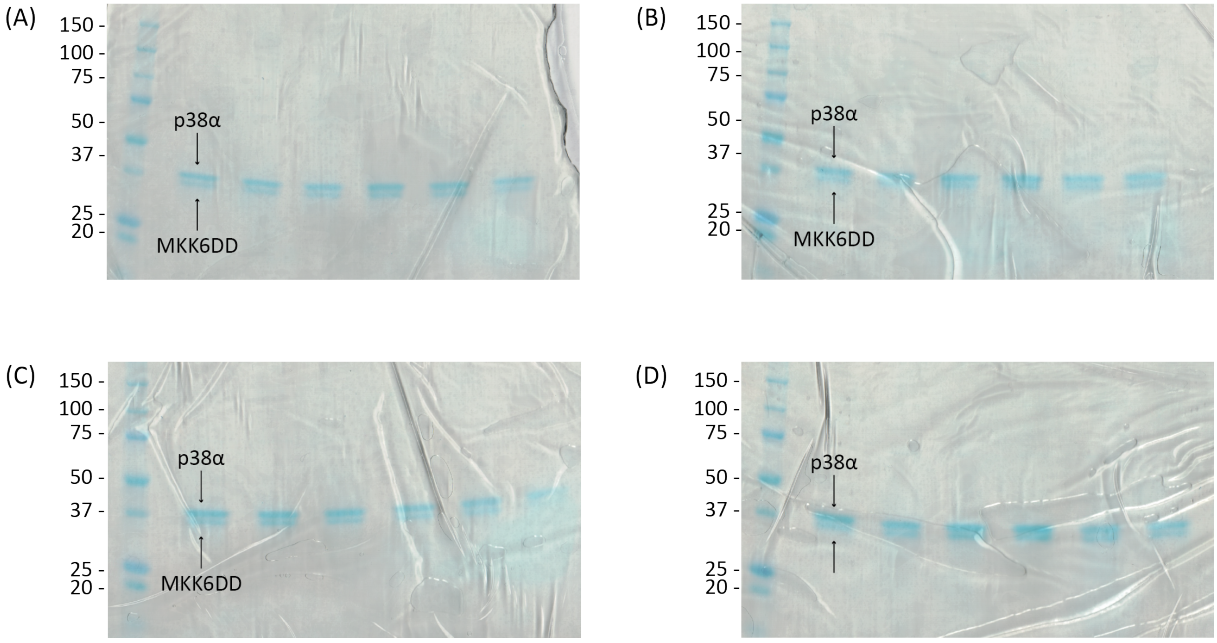

(C)
mMKK6DDGRA
p38α Y182F
no additive
AMP-PCP
ADP + AlF 4 -
ATP
ATP RT

Appendix 3: LC-MS/MS analysis of MKK6DD with p38α WT (A) and K53R (B) samples in presence of 10 mM AMP, ADP, uADP, ATP or without nucleotide. For each analysed sample, the number of identifications of MKK6DD and p38α peptides is plotted over their amino acid sequence. Counts are coloured by their MASCOT analysis results: not phosphorylated (grey), potentially phosphorylated (blue) and phosphorylated (red).

(A) MKK6DD p38α WT




(B)
MKK6DD
p38α K53R
no nucelotide
AMP
ADP
uADP
ATP



Appendix 4: control SDS PAGE gels of the radioactivity gel in Figure 52.

Appendix 5: GRA16 alleles from types I (TGGT1_208830), II (TGME49_208830), and III (TGVEG_208830) T. gondii strains were aligned using Clustal Ω. Unconserved residues are highlighted in red.

```
                          10        20        30        40        50
GRA16_TGGT1_208830        MYRNHSGIRLACRLFEVGALVLALENVSGIHRFVAGIEWNEGKEDFQYTTSPWVIPPDGL  60
GRA16_TGME49_208830       MYRNHSGIRLACRLFEVGALVLALENVSGIHRFVAGIEWNEGKEDFQYTTSPWVIPPDGL  60
GRA16_TGVEG_208830        MYRNHSGIRLACRLFEVGALVLALENVSGIHRFVAGIEWNEGKEDFQYTTSPWVIPPDGL  60

                          70        80        90        100       110
GRA16_TGGT1_208830        VSRRLAEEPPRKRLRKTNKSDRDSDSAQGSRTTSPGSLGGFGATVGRVATPRIRSGVVAS 120
GRA16_TGME49_208830       VSRRLAEEPPRKRLRKTNKSDRDSDSAQGSRTTSPGSLGGFGATGGRVAAPRIRSGVVAS 120
GRA16_TGVEG_208830        VSRRLAEEPPRKRLRKTNKSDRDSDSAQGSRTTSPGSLGGFGATGGRVAAPRIRSGVVAS 120

                          130       140       150       160       170
GRA16_TGGT1_208830        EAIRGTIWRRPGEVESTLKLRRTRPQYSQTDGDGLQGNRLSSTGERSGISHGAQSLAMRP 180
GRA16_TGME49_208830       EAIRGTIWRRPGEVEGTLKLRRTRPQYSQTDGDGLQGNRLSSTGERSGISHGAQSLAMRP 180
GRA16_TGVEG_208830        EAIRGTIWRRPGEVEGTLKLRRTRPQYSQTDGDGLQGNRLSSTGERSGISHGAQSLAMRP 180

                          190       200       210       220       230
GRA16_TGGT1_208830        RTMGQTMKSLESSWDSDPLEGTSRDWQYVPTSETAASPGLTGLGGIGRKFAPLYVRDRKF 240
GRA16_TGME49_208830       RTMGQTMKSLESSWDSDPLEGTSRDWLYVTTSETTASPGLTGFGGIGKKLAPLYVRDRKF 240
GRA16_TGVEG_208830        RTMGQTMKSLESSWDSDPLEGTSRDWLYVTTSETTASPGLTGFGGIGKKLAPLYVRDRKF 240

                          250       260       270       280       290
GRA16_TGGT1_208830        DLLQFVNLTRSKKQKLLMSSKSPSLRRLLMNDMAQEWALGILQIALQGRQRALQASHTTR 300
GRA16_TGME49_208830       DLLQFVNLTRSKKQKLLMSSKSPSLRKLLMNDMAQEWALGILQIVLQGRQRALQASHTTR 300
GRA16_TGVEG_208830        DLLQFVNLTRSKKQKLLMSSKSPSLRKLLMNDMAQEWALGILQIVLQGRQRALQASHTTR 300

                          310       320       330       340       350
GRA16_TGGT1_208830        TTEPASGTDGTSKSSEDEATRASEGNASVNQTSPAASYPRRPSSDEGQDSGRRKCSKRSP 360
GRA16_TGME49_208830       TTEPASGTDGTSKSSEDEATRASEGNASVNQTSPAASYPRRPSSDEGQDSGRRKCSKRSP 360
GRA16_TGVEG_208830        TTEPASGTDGTSKSSEDEATRASEGNASVNQTSPAASYPRRPSSDEGQDSGRRKCSKRSP 360

                          370       380       390       400       410
GRA16_TGGT1_208830        SRLVQNAPLFLKDDSHSLKDTLDLVKNKNRELTEKGRVHATPLRVVLLNSIMMKKLEKVL 420
GRA16_TGME49_208830       SRLVQNAPLFLKDDSHSLKDTLDLVKNKNRELTEKGRVHATPLRVVLLNSIMMKKLEKVL 420
GRA16_TGVEG_208830        SRLVQNAPLFLKDDSHSLKDTLDLVKNKNRELTEKGRVHATPLRVVLLNSIMMKKLEKVL 420

                          430       440       450       460       470
GRA16_TGGT1_208830        PVVESMDRALMARQTSSEAATVDDSSTSISHGMQGSTTSGAAAVQGPSTSVPGASGGLGP 480
GRA16_TGME49_208830       PVVESMDRALMARQTSSEAATVDDSSTSISHGMQGSTSSGAAAVQGPSTSVPGASGGLGP 480
GRA16_TGVEG_208830        PVVESMDRALMARQTSSEAATVDDSSTSISHGMQGSTSSGAAAVQGPSTSVPGASGGLGP 480

                          490       500
GRA16_TGGT1_208830        SGGKRKPDDEDDFDCSRAKRKNDQM                                     505
GRA16_TGME49_208830       SGGKRKPDDEDDFDCSRAKRKNDQM                                     505
GRA16_TGVEG_208830        SGGKRKPDDEDDFDCSRAKRKNDQM                                     505
```

Bibliography

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. and Adams, P. D. (2012) 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallographica Section D Biological Crystallography*, 68(4), pp. 352–367. doi: 10.1107/S0907444912001308.

Akella, R., Min, X., Wu, Q., Gardner, K. H. and Goldsmith, E. J. (2010) 'The third conformation of p38α MAP kinase observed in phosphorylated p38α and in solution', *Structure*, 18(12), pp. 1571–1578. doi: 10.1016/j.str.2010.09.015.

Albanese, S. K., Parton, D. L., Işlk, M., Rodríguez-Laureano, L., Hanson, S. M., Behr, J. M., Gradia, S., Jeans, C., Levinson, N. M., Seeliger, M. A. and Chodera, J. D. (2018) 'An Open Library of Human Kinase Domain Constructs for Automated Bacterial Expression', *Biochemistry*, 57(31), pp. 4675–4689. doi: 10.1021/acs.biochem.7b01081.

Alexa, A., Varga, J. and Reményi, A. (2010) 'Scaffolds are "active" regulators of signaling modules', *FEBS Journal*, 277(21), pp. 4376–4382. doi: 10.1111/j.1742-4658.2010.07867.x.

Ambrosino, C., Mace, G., Galban, S., Fritsch, C., Vintersten, K., Black, E., Gorospe, M. and Nebreda, A. R. (2003) 'Negative Feedback Regulation of MKK6 mRNA Stability by p38α Mitogen-Activated Protein Kinase', *Molecular and Cellular Biology*, 23(1), pp. 370–381. doi: 10.1128/mcb.23.1.370-381.2003.

Aoki, K., Takahashi, K., Kaizu, K. and Matsuda, M. (2013) 'A Quantitative Model of ERK MAP Kinase Phosphorylation in Crowded Media', *Scientific Reports*, 3(1), p. 1541. doi: 10.1038/srep01541.

Aoki, K., Yamada, M., Kunida, K., Yasuda, S. and Matsuda, M. (2011) 'Processive phosphorylation of ERK MAP kinase in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 108(31), pp. 12675–12680. doi: 10.1073/pnas.1104030108.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G. and Muzio, L. Lo (2017) 'The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review)', *International Journal of Molecular Medicine*, 40(2), pp. 271–280. doi: 10.3892/ijmm.2017.3036.

Askari, N., Beenstock, J., Livnah, O. and Engelberg, D. (2009) 'p38α Is Active in Vitro and in Vivo When Monophosphorylated at Threonine 180', *Biochemistry*, 48(11), pp. 2497–2504. doi: 10.1021/bi900024v.

Beardmore, V. A., Hinton, H. J., Eftychi, C., Apostolaki, M., Armaka, M., Darragh, J., McIlrath, J., Carr, J. M., Armit, L. J., Clacher, C., Malone, L., Kollias, G. and Arthur, J. S. C. (2005) 'Generation and Characterization of p38β (MAPK11) Gene-Targeted Mice', *Molecular and Cellular Biology*, 25(23), pp. 10454–10464. doi: 10.1128/mcb.25.23.10454-10464.2005.

Becher, I., Savitski, M. M., Savitski, M. F., Hopf, C., Bantscheff, M. and Drewes, G. (2013) 'Affinity profiling of the cellular kinome for the nucleotide cofactors ATP, ADP, and GTP', *ACS Chemical Biology*, 8(3), pp. 599–607. doi: 10.1021/cb3005879.

Beenstock, J., Mooshayef, N. and Engelberg, D. (2016) 'How Do Protein Kinases Take a Selfie (Autophosphorylate)?', *Trends in Biochemical Sciences*, 41(11), pp. 938–953. doi: 10.1016/j.tibs.2016.08.006.

Bell, M. and Engelberg, D. (2003) 'Phosphorylation of Tyr-176 of the yeast MAPK Hog1/p38 is not vital for Hog1 biological activity', *Journal of Biological Chemistry*, 278(17), pp. 14603–14606. doi: 10.1074/jbc.C300006200.

Bepler, T., Noble, A. J. and Berger, B. (2019) 'Topaz-Denoise: General deep denoising models for cryoEM', *bioRxiv*. doi: 10.1101/838920.

Boivin, S., Kozak, S. and Meijers, R. (2013) 'Optimization of protein purification and characterization using Thermofluor screens', *Protein Expression and Purification*, 91(2), pp. 192–206. doi: 10.1016/j.pep.2013.08.002.

Bojagora, A. and Saridakis, V. (2020) 'USP7 manipulation by viral proteins', *Virus Research*, 286, p. 198076. doi: 10.1016/j.virusres.2020.198076.

Bougdour, A., Durandau, E., Brenier-Pinchart, M. P., Ortet, P., Barakat, M., Kieffer, S., Curt-Varesano, A., Curt-Bertini, R. L., Bastien, O., Coute, Y., Pelloux, H. and Hakimi, M. A. (2013) 'Host cell subversion by Toxoplasma GRA16, an exported dense granule protein that targets the host cell nucleus and alters gene expression', *Cell Host and Microbe*, 13(4), pp. 489–500. doi: 10.1016/j.chom.2013.03.002.

Bowler, M. W. *et al.* (2015) 'MASSIF-1: A beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules', *Journal of Synchrotron Radiation*, 22, pp. 1540–1547. doi: 10.1107/S1600577515016604.

Braun, L., Brenier-Pinchart, M.-P., Yogavel, M., Curt-Varesano, A. A., Curt-Bertini, R.-L., Hussain, T., Kieffer-Jaquinod, S., Coute, Y., Pelloux, H., Tardieux, I., Sharma, A., Belrhali, H., Bougdour, A. and Hakimi, M. A. (2013) 'A Toxoplasma dense granule protein, GRA24, modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation.', *The Journal of experimental medicine*, 210(10), pp. 2071–86. doi: 10.1084/jem.20130103.

Braun, L., Brenier-Pinchart, M. P., Hammoudi, P. M., Cannella, D., Kieffer-Jaquinod, S., Vollaire, J., Josserand, V., Touquet, B., Couté, Y., Tardieux, I., Bougdour, A. and Hakimi, M. A. (2019) 'The Toxoplasma effector TEEGR promotes parasite persistence by modulating NF-κB signalling via EZH2', *Nature Microbiology*, 4(7), pp. 1208–1220. doi: 10.1038/s41564-019-0431-8.

Brichkina, A., Nguyen, N. T. M., Baskar, R., Wee, S., Gunaratne, J., Robinson, R. C. and Bulavin, D. V. (2016) 'Proline isomerisation as a novel regulatory mechanism for p38MAPK activation and functions', *Cell Death and Differentiation*, 23(10), pp. 1592–1601. doi: 10.1038/cdd.2016.45.

Brunet, J., Pfaff, A. W., Abidi, A., Unoki, M., Nakamura, Y., Guinard, M., Klein, J.-P., Candolfi, E. and Mousli, M. (2008) 'Toxoplasma gondii exploits UHRF1 and induces host cell cycle arrest at G2 to enable its proliferation', *Cellular Microbiology*, 10(4), pp. 908–920. doi: 10.1111/j.1462-5822.2007.01093.x.

Bührmann, M., Hardick, J., Weisner, J., Quambusch, L. and Rauh, D. (2017) 'Covalent Lipid

Pocket Ligands Targeting p38α MAPK Mutants', *Angewandte Chemie - International Edition*, 56(43), pp. 13232–13236. doi: 10.1002/anie.201706345.

Bührmann, M., Wiedemann, B. M., Müller, M. P., Hardick, J., Ecke, M. and Rauh, D. (2017) 'Structure-based design, synthesis and crystallization of 2-arylquinazolines as lipid pocket ligands of p38α MAPK', *PLoS ONE*, 12(9), pp. 1–27. doi: 10.1371/journal.pone.0184627.

Burton, J. C., Antoniades, W., Okalova, J., Roos, M. M. and Grimsey, N. J. (2021) 'Atypical p38 signaling, activation, and implications for disease', *International Journal of Molecular Sciences*, 22(8). doi: 10.3390/ijms22084183.

Caffrey, D. R., O'Neill, L. A. J. and Shields, D. C. (1999) 'The Evolution of the MAP Kinase Pathways: Coduplication of Interacting Proteins Leads to New Signaling Cascades', *Journal of Molecular Evolution*, 49(5), pp. 567–582. doi: 10.1007/PL00006578.

Canovas, B. and Nebreda, A. R. (2021) 'Diversity and versatility of p38 kinase signalling in health and disease', *Nature Reviews Molecular Cell Biology*, 6. doi: 10.1038/s41580-020-00322-w.

Chang, C. I., Xu, B. E., Akella, R., Cobb, M. H. and Goldsmith, E. J. (2002) 'Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b', *Molecular Cell*, 9(6), pp. 1241–1249. doi: 10.1016/S1097-2765(02)00525-7.

Chavoshi, S., Egorova, O., Lacdao, I. K., Farhadi, S., Sheng, Y. and Saridakis, V. (2016) 'Identification of kaposi sarcoma herpesvirus (KSHV) vIRF1 protein as a novel interaction partner of human deubiquitinase USP7', *Journal of Biological Chemistry*, 291(12), pp. 6281–6291. doi: 10.1074/jbc.M115.710632.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: All-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.

Chen, Z., Gibson, T. B., Robinson, F., Silvestro, L., Pearson, G., Xu, B. E., Wright, A., Vanderbilt, C., Cobb, M. H., Raf, B., Raf, B. and Raf, A. (2001) 'MAP Kinases', *Chemical Review*, 101(101), pp. 2449–2476. doi: 10.1021/cr000241p.

Cho, J. H., Hong, W. G., Jung, Y. J., Lee, J., Lee, E., Hwang, S. G., Um, H. D. and Park, J. K. (2016) 'Γ-Ionizing radiation-induced activation of the EGFR–p38/ERK–STAT3/CREB-1–EMT pathway promotes the migration/invasion of non-small cell lung cancer cells and is inhibited by podophyllotoxin acetate', *Tumor Biology*, 37(6), pp. 7315–7325. doi: 10.1007/s13277-015-4548-y.

Cliff, M. J., Bowler, M. W., Varga, A., Marston, J. P., Szabó, J., Hounslow, A. M., Baxter, N. J., Blackburn, G. M., Vas, M. and Waltho, J. P. (2010) 'Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis', *Journal of the American Chemical Society*, 132(18), pp. 6507–6516. doi: 10.1021/ja100974t.

Cobb, M. H. and Goldsmith, E. J. (1995) 'How MAPK are regulated', *The Journal of biological chemistry*, pp. 14843–14846. doi: 10.1074/jbc.270.25.14843.

Cuadrado, A. and Nebreda, A. R. R. (2010) 'Mechanisms and functions of p38 MAPK signalling.', *The Biochemical journal*, 429(3), pp. 403–417. doi: 10.1042/BJ20100323.

Cuenda, A., Alonso, G., Morrice, N., Jones, M., Meier, R., Cohen, P. and Nebreda, A. R. (1996) 'Purification and cDNA cloning of SAPKK3, the major activator of RK/p38 in stress- and cytokine-stimulated monocytes and epithelial cells', *EMBO Journal*, 15(16), pp. 4156–4164. doi: 10.1002/j.1460-2075.1996.tb00790.x.

Cuenda, A. and Sanz-Ezquerro, J. J. (2017) 'p38γ and p38δ: From Spectators to Key Physiological Players', *Trends in Biochemical Sciences*, 42(6), pp. 431–442. doi: 10.1016/j.tibs.2017.02.008.

Davey, N. E., Travé, G. and Gibson, T. J. (2011) 'How viruses hijack cell regulation', *Trends in Biochemical Sciences*, 36(3), pp. 159–169. doi: 10.1016/j.tibs.2010.10.002.

Dimasi, N., Flot, D., Dupeux, F. and Márquez, J. A. (2007) 'Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1', *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63(3), pp. 204–208. doi: 10.1107/S1744309107004903.

Diskin, R., Engelberg, D. and Livnah, O. (2008) 'A Novel Lipid Binding Site Formed by the MAP Kinase Insert in p38α', *Journal of Molecular Biology*, 375(1), pp. 70–79. doi: 10.1016/j.jmb.2007.09.002.

Donoghue, C., Cubillos-Rojas, M., Gutierrez-Prat, N., Sanchez-Zarzalejo, C., Verdaguer, X., Riera, A. and Nebreda, A. R. (2020) 'Optimal linker length for small molecule PROTACs that selectively target p38α and p38β for degradation', *European Journal of Medicinal Chemistry*, 201, p. 112451. doi: 10.1016/j.ejmech.2020.112451.

Dosztányi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005) 'IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content', *Bioinformatics*, 21(16), pp. 3433–3434. doi: 10.1093/bioinformatics/bti541.

Dosztányi, Z., Mészáros, B. and Simon, I. (2009) 'ANCHOR: Web server for predicting protein binding regions in disordered proteins', *Bioinformatics*, 25(20), pp. 2745–2746. doi: 10.1093/bioinformatics/btp518.

Dupeux, F., Röwer, M., Seroul, G., Blot, D. and Márquez, J. A. (2011) 'A thermal stability assay can help to estimate the crystallization likelihood of biological samples', *Acta Crystallographica Section D: Biological Crystallography*, 67(11), pp. 915–919. doi: 10.1107/S0907444911036225.

Emsley, P. and Cowtan, K. (2004) 'Coot : model-building tools for molecular graphics', *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp. 2126–2132. doi: 10.1107/S0907444904019158.

Endicott, J. A., Noble, M. E. M. and Johnson, L. N. (2012) 'The Structural Basis for Control of Eukaryotic Protein Kinases', *Annual Review of Biochemistry*, 81(1), pp. 587–613. doi: 10.1146/annurev-biochem-052410-090317.

Ferrao, R., Zhou, H., Shan, Y., Liu, Q., Li, Q., Shaw, D. E., Li, X. and Wu, H. (2014) 'IRAK4 Dimerization and trans-Autophosphorylation Are Induced by Myddosome Assembly', *Molecular Cell*, 55(6), pp. 891–903. doi: 10.1016/j.molcel.2014.08.006.

Ferrell, J. E. and Ha, S. H. (2014) 'Ultrasensitivity part II: Multisite phosphorylation,

stoichiometric inhibitors, and positive feedback', *Trends in Biochemical Sciences*, 39(11), pp. 556–569. doi: 10.1016/j.tibs.2014.09.003.

Fischmann, T. O., Smith, C. K., Mayhood, T. W., Myers, J. E., Reichert, P., Mannarino, A., Carr, D., Zhu, H., Wong, J., Yang, R. S., Le, H. V. and Madison, V. S. (2009) 'Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors', *Biochemistry*, 48(12), pp. 2661–2674. doi: 10.1021/bi801898e.

Förster, S., Apostol, L. and Bras, W. (2010) 'Scatter: Software for the analysis of nano-and mesoscale small-angle scattering', *Journal of Applied Crystallography*, 43(3), pp. 639–646. doi: 10.1107/S0021889810008289.

Freshney, N. W., Rawlinson, L., Guesdon, F., Jones, E., Cowley, S., Hsuan, J. and Saklatvala, J. (1994) 'Interleukin-1 activates a novel protein kinase cascade that results in the phosphorylation of hsp27', *Cell*, 78(6), pp. 1039–1049. doi: 10.1016/0092-8674(94)90278-X.

Galanis, A., Yang, S. H. and Sharrocks, A. D. (2001) 'Selective targeting of MAPKs to the ETS domain transcription factor SAP-1', *Journal of Biological Chemistry*, 276(2), pp. 965–973. doi: 10.1074/jbc.M007697200.

Galarreta, A., Valledor, P., Ubieto-Capella, P., Lafarga, V., Zarzuela, E., Muñoz, J., Malumbres, M., Lecona, E. and Fernandez-Capetillo, O. (2021) 'USP7 limits CDK1 activity throughout the cell cycle', *The EMBO Journal*, pp. 1–13. doi: 10.15252/embj.201899692.

Garai, A., Zeke, A., Gogl, G., Toro, I., Fordos, F., Blankenburg, H., Barkai, T., Varga, J., Alexa, A., Emig, D., Albrecht, M. and Remenyi, A. (2012) 'Specificity of Linear Motifs That Bind to a Common Mitogen-Activated Protein Kinase Docking Groove', *Science Signaling*, 5(245), pp. ra74–ra74. doi: 10.1126/scisignal.2003004.

Ge, B. X., Gram, H., Di Padova, F., Huang, B., New, L., Ulevitch, R. J., Luo, Y. and Han, J. H. (2002) 'MAPKK-independent activation of p38 alpha mediated by TAB1- dependent autophosphorylation of p38 alpha', *Science*, 295(5558), pp. 1291–1294. doi: 10.1126/science.1067289.

Grimes, J. M. and Grimes, K. V. (2020) 'p38 MAPK inhibition: A promising therapeutic approach for COVID-19', *Journal of Molecular and Cellular Cardiology*, 144, pp. 63–65. doi: 10.1016/j.yjmcc.2020.05.007.

Guixé, V. and Merino, F. (2009) 'The ADP-dependent sugar kinase family: Kinetic and evolutionary aspects', *IUBMB Life*, 61(7), pp. 753–761. doi: 10.1002/iub.217.

Ha, J., Kang, E., Seo, J. and Cho, S. (2019) 'Phosphorylation dynamics of jnk signaling: Effects of dual-specificity phosphatases (dusps) on the jnk pathway', *International Journal of Molecular Sciences*, 20(24), pp. 1–19. doi: 10.3390/ijms20246157.

Hakimi, M.-A., Olias, P. and Sibley, L. D. (2017) 'Toxoplasma Effectors Targeting Host Signaling and Transcription', *Clinical Microbiology Reviews*, 30(3), pp. 615–645. doi: 10.1128/CMR.00005-17.

Han, J., Lee, J.-D., Jiang, Y., Li, Z., Feng, L. and Ulevitch, R. J. (1996) 'Characterization of the Structure and Function of a Novel MAP Kinase Kinase (MKK6)', *Journal of Biological Chemistry*, 271(6), pp. 2886–2891. doi: 10.1074/jbc.271.6.2886.

Han, J., Lee, J., Bibbs, L. and Ulevitch, R. (1994) 'A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells', *Science*, 265(5173), pp. 808–811. doi: 10.1126/science.7914033.

Haystead, T. A. J., Dent, P., Wu, J., Haystead, C. M. M. and Sturgill, T. W. (1992) 'phosphorylation of p42mapk by MAP kinase kinase', *FEBS Letters*, 306(1), pp. 17–22. doi: 10.1016/0014-5793(92)80828-5.

He, H., Braun, L., Kraut, A., Tardieux, I., Hakimi, M. A., Touquet, B., Coute, Y. and Bougdour, A. (2018) 'Characterization of a Toxoplasma effector uncovers an alternative GSK3/β-catenin-regulatory pathway of inflammation', *eLife*, pp. 1–28. doi: 10.7554/eLife.39887.

Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S. and Adams, P. D. (2012) 'Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution', *Acta Crystallographica Section D: Biological Crystallography*, 68(4), pp. 381–390. doi: 10.1107/S0907444911047834.

Herzik, M. A., Wu, M. and Lander, G. C. (2018) 'High-resolution structure determination of sub-100 kilodalton complexes using conventional cryo-EM'. doi: 10.1101/489898.

Hu, M., Gu, L., Li, M., Jeffrey, P. D., Gu, W. and Shi, Y. (2006) 'Structural basis of competi tive recognition of p53 and MDM2 by HAUSP/USP7: Implications for the regulation of the p53-MDM2 pathway', *PLoS Biology*, 4(2), pp. 228–239. doi: 10.1371/journal.pbio.0040027.

Humphreys, J. M., Piala, A. T., Akella, R., He, H. and Goldsmith, E. J. (2013) 'Precisely ordered phosphorylation reactions in the p38 mitogen-activated protein (MAP) kinase cascade', *Journal of Biological Chemistry*, 288(32), pp. 23322–23330. doi: 10.1074/jbc.M113.462101.

Hunter, C. A. and Sibley, L. D. (2012) 'Modulation of innate immunity by Toxoplasma gondii virulence effectors', *Nature Reviews Microbiology*, 10(11), pp. 766–778. doi: 10.1038/nrmicro2858.

Ihle, J. N. (2000) 'The challenges of translating knockout phenotypes into gene function', *Cell*, 102(2), pp. 131–134. doi: 10.1016/S0092-8674(00)00017-9.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H. and Wakagi, T. (2001) 'Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon', *Structure*, 9(3), pp. 205–214. doi: 10.1016/S0969-2126(01)00577-9.

Jacobs, D., Glossip, D., Xing, H., Muslin, A. J. and Kornfeld, K. (1999) 'Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase', *Genes and Development*, 13(2), pp. 163–175. doi: 10.1101/gad.13.2.163.

Jagannathan, V., Kaur, P. and Datta, S. (2010) 'Polyphosphate kinase from m. tuberculosis: An interconnect between the genetic and biochemical role', *PLoS ONE*, 5(12). doi: 10.1371/journal.pone.0014336.

Jensen, K. D. C., Wang, Y., Wojno, E. D. T., Shastri, A. J., Hu, K., Cornel, L., Boedec, E., Ong, Y. C., Chien, Y. H., Hunter, C. A., Boothroyd, J. C. and Saeij, J. P. J. (2011) 'Toxoplasma polymorphic effectors determine macrophage polarization and intestinal inflammation', *Cell Host and Microbe*, 9(6), pp. 472–483. doi: 10.1016/j.chom.2011.04.015.

Jin, Y., Molt, R. W. and Blackburn, G. M. (2017) 'Metal Fluorides: Tools for Structural and Computational Analysis of Phosphoryl Transfer Enzymes', *Topics in Current Chemistry*, 375(2), pp. 1–31. doi: 10.1007/s41061-017-0130-y.

Jin, Y., Richards, N. G., Waltho, J. P. and Blackburn, G. M. (2017) 'Metal Fluorides as Analogues for Studies on Phosphoryl Transfer Enzymes', *Angewandte Chemie - International Edition*, 56(15), pp. 4110–4128. doi: 10.1002/anie.201606474.

Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 125–132. doi: 10.1107/S0907444909047337.

Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Lührmann, R. and Stark, H. (2007) 'GraFix: sample preparation for single-particle electron cryomicroscopy', *Nature Methods*, 5(1), pp. 53–55. doi: 10.1038/nmeth1139.

Kato, M., Chuang, J. L., Tso, S. C., Wynn, R. M. and Chuang, D. T. (2005) 'Crystal structure of pyruvate dehydrogenase kinase 3 bound to lipoyl domain 2 of human pyruvate dehydrogenase complex', *EMBO Journal*, 24(10), pp. 1763–1774. doi: 10.1038/sj.emboj.7600663.

Kengen, S. W. M., De Bok, F. A. M., Van Loo, N. D., Dijkema, C., Stams, A. J. M. and De Vos, W. M. (1994) 'Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by Pyrococcus furiosus', *Journal of Biological Chemistry*, 269(26), pp. 17537–17541. doi: 10.1016/s0021-9258(17)32474-2.

Kim, S. G., Seo, S. H., Shin, J. H., Yang, J. P., Lee, S. H. and Shin, E. H. (2019) 'Increase in the nuclear localization of PTEN by the Toxoplasma GRA16 protein and subsequent induction of p53-dependent apoptosis and anticancer effect', *Journal of Cellular and Molecular Medicine*, 23(5), pp. 3234–3245. doi: 10.1111/jcmm.14207.

Kirsch, K., Zeke, A., Tőke, O., Sok, P., Sethi, A., Sebő, A., Kumar, G. S., Egri, P., Póti, Á. L., Gooley, P., Peti, W., Bento, I., Alexa, A. and Reményi, A. (2020) 'Co-regulation of the transcription controlling ATF2 phosphoswitch by JNK and p38', *Nature Communications*, 11(1), pp. 1–15. doi: 10.1038/s41467-020-19582-3.

Kornev, A. P. and Taylor, S. S. (2010) 'Defining the conserved internal architecture of a protein kinase', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(3), pp. 440–444. doi: 10.1016/j.bbapap.2009.10.017.

Kuhlbrandt, W. (2014) 'The Resolution Revolution', *Science*, 343(6178), pp. 1443–1444. doi: 10.1126/science.1251652.

Kumar, G. S., Page, R. and Peti, W. (2021) 'The interaction of p38 with its upstream kinase MKK6', *Protein Science*, 6(January), p. pro.4039. doi: 10.1002/pro.4039.

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Signalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B. and Gibson, T. J. (2020) 'ELM-the eukaryotic linear motif resource in 2020', *Nucleic Acids Research*, 48(D1), pp. D296–D306. doi: 10.1093/nar/gkz1030.

Kumphune, S., Chattipakorn, S. and Chattipakorn, N. (2012) 'Role of p38 inhibition in cardiac ischemia/reperfusion injury', *European Journal of Clinical Pharmacology*, 68(5), pp. 513–524.

doi: 10.1007/s00228-011-1193-2.

Kyriakis JM, A. J. (2001) 'Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation.', *Physiol Rev*, 81(2), pp. 807–869.

LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z. and Rout, M. P. (2016) 'Protein complex purification by affinity capture', *Cold Spring Harbor Protocols*, 2016(7), pp. 601–605. doi: 10.1101/pdb.top077545.

Lavoie, H., Gagnon, J. and Therrien, M. (2020) 'ERK signalling: a master regulator of cell behaviour, life and fate', *Nature Reviews Molecular Cell Biology*, 21(10), pp. 607–632. doi: 10.1038/s41580-020-0255-7.

Lawrence, M. C., Jivan, A., Shao, C., Duan, L., Goad, D., Zaganjor, E., Osborne, J., McGlynn, K., Stippec, S., Earnest, S., Chen, W. and Cobb, M. H. (2008) 'The roles of MAPKs in disease', *Cell Research*, 18(4), pp. 436–442. doi: 10.1038/cr.2008.37.

Lecona, E., Rodriguez-Acebes, S., Specks, J., Lopez-Contreras, A. J., Ruppen, I., Murga, M., Muñoz, J., Mendez, J. and Fernandez-Capetillo, O. (2016) 'USP7 is a SUMO deubiquitinase essential for DNA replication', *Nature Structural and Molecular Biology*, 23(4), pp. 270–277. doi: 10.1038/nsmb.3185.

Lee, H.-R., Choi, W.-C., Lee, S., Hwang, J., Hwang, E., Guchhait, K., Haas, J., Toth, Z., Jeon, Y. H., Oh, T.-K., Kim, M. H. and Jung, J. U. (2011) 'Bilateral inhibition of HAUSP deubiquitinase by a viral interferon regulatory factor protein', *Nature Structural & Molecular Biology*, 18(12), pp. 1336–1344. doi: 10.1038/nsmb.2142.

Lee, J. C., Laydon, J. T., McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Keys, J. R., Land Vatter, S. W., Strickler, J. E., McLaughlin, M. M., Siemens, I. R., Fisher, S. M., Livi, G. P., White, J. R., Adams, J. L. and Young, P. R. (1994) 'A protein kinase involved in the regulation of inflammatory cytokine biosynthesis', *Nature*, 372(6508), pp. 739–746. doi: 10.1038/372739a0.

Lee, P., Chandel, N. S. and Simon, M. C. (2020) 'Cellular adaptation to hypoxia through hypoxia inducible factors and beyond', *Nature Reviews Molecular Cell Biology*, 21(5), pp. 268–283. doi: 10.1038/s41580-020-0227-y.

Li, M., Chen, D., Shiloh, A. and Luo, J. (2002) 'Analysis of mRNA expression using oligonucleotide microarrays', 416(April). doi: 10.1038/nature735.1.

Li, S., Gong, P., Zhang, N., Li, X., Tai, L., Wang, X., Yang, Z., Yang, J., Zhu, X., Zhang, X. and Li, J. (2019) '14-3-3 protein of neospora caninum modulates host cell innate immunity through the activation of MAPK and NF-κB pathways', *Frontiers in Microbiology*, 10(JAN), pp. 1–13. doi: 10.3389/fmicb.2019.00037.

Liu, M. Y., Hua, W. K., Chen, C. J. and Lin, W. J. (2020) 'The MKK-dependent phosphorylation of p38α is augmented by arginine methylation on arg49/arg149 during erythroid differentiation', *International Journal of Molecular Sciences*, 21(10), pp. 1–21. doi: 10.3390/ijms21103546.

Liu, X., Zhang, C.-S., Lu, C., Lin, S.-C., Wu, J.-W. and Wang, Z.-X. (2016) 'A conserved motif in JNK/p38-specific MAPK phosphatases as a determinant for JNK1 recognition and inactivation', *Nature Communications*, 7, p. 10879. doi: 10.1038/ncomms10879.

López-Santalla, M., Salvador-Bernáldez, M., González-Alvaro, I., Castañeda, S., Ortiz, A. M., García-García, M. I., Kremer, L., Roncal, F., Mulero, J., Martínez-A, C. and Salvador, J. M. (2011) 'Tyr^{323}-dependent p38 activation is associated with rheumatoid arthritis and correlates with disease activity.', *Arthritis and rheumatism*, 63(7), pp. 1833–1842. doi: 10.1002/art.30375.

Ma, J. S., Sasai, M., Ohshima, J., Lee, Y., Bando, H., Takeda, K. and Yamamoto, M. (2014) 'Selective and strain-specific NFAT4 activation by the Toxoplasma gondii polymorphic dense granule protein GRA6', *Journal of Experimental Medicine*, 211(10), pp. 2013–2032. doi: 10.1084/jem.20131272.

Mace, P. D., Wallez, Y., Egger, M. F., Dobaczewska, M. K., Robinson, H., Pasquale, E. B. and Riedl, S. J. (2013) 'Structure of ERK2 bound to PEA-15 reveals a mechanism for rapid release of activated MAPK', *Nature Communications*, 4(1), p. 1681. doi: 10.1038/ncomms2687.

Machius, M., Chuang, J. L., Wynn, R. M., Tomchick, D. R. and Chuang, D. T. (2001) 'Structure of rat BCKD kinase: Nucleotide-induced domain communication in a mitochondrial protein kinase', *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), pp. 11218–11223. doi: 10.1073/pnas.201220098.

Makino, Y., Sato, T., Kawamura, H., Hachisuka, S. I., Takeno, R., Imanaka, T. and Atomi, H. (2016) 'An archaeal ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism', *Nature Communications*, 7, pp. 1–12. doi: 10.1038/ncomms13446.

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I. and Franke, D. (2021) ' ATSAS 3.0 : expanded functionality and new tools for small-angle scattering data analysis ', *Journal of Applied Crystallography*, 54(1), pp. 343–355. doi: 10.1107/s1600576720013412.

Manning, G. (2002) 'The Protein Kinase Complement of the Human Genome', *Science*, 298(5600), pp. 1912–1934. doi: 10.1126/science.1075762.

Mansour, S. J., Candia, J. M., Matsuura, J. E., Manning, M. C. and Ahn, N. G. (1996) 'Interdependent domains controlling the enzymatic activity of mitogen activated protein kinase kinase 1', *Biochemistry*, 35(48), pp. 15529–15536. doi: 10.1021/bi961854s.

Marino, N. D., Panas, M. W., Franco, M., Theisen, T. C., Naor, A., Rastogi, S., Buchholz, K. R., Lorenzi, H. A. and Boothroyd, J. C. (2018) 'Identification of a novel protein complex essential for effector translocation across the parasitophorous vacuole membrane of Toxoplasma gondii', *PLoS Pathogens*, 14(1), pp. 1–26. doi: 10.1371/journal.ppat.1006828.

Matsumoto, T., Kinoshita, T., Matsuzaka, H., Nakai, R., Kirii, Y., Yokota, K. and Tada, T. (2012) 'Crystal structure of non-phosphorylated MAP2K6 in a putative auto-inhibition state', *Journal of Biochemistry*, 151(5), pp. 541–549. doi: 10.1093/jb/mvs023.

Matsushita, M., Nakamura, T., Moriizumi, H., Miki, H. and Takekawa, M. (2020) 'Stress-responsive MTK1 SAPKKK serves as a redox sensor that mediates delayed and sustained activation of SAPKs by oxidative stress', *Science Advances*, 6(26). doi: 10.1126/sciadv.aay9778.

Meister, M., Tomasovic, A., Banning, A. and Tikkanen, R. (2013) 'Mitogen-Activated Protein (MAP) Kinase Scaffolding Proteins: A Recount', *International Journal of Molecular Sciences*, 14(3), pp. 4854–4884. doi: 10.3390/ijms14034854.

Mercer, H. L., Snyder, L. M., Doherty, C. M., Fox, B. A., Bzik, D. J. and Denkers, E. Y. (2020) 'Toxoplasma gondii dense granule protein GRA24 drives MyD88-independent p38 MAPK activation, IL-12 production and induction of protective immunity', *PLoS Pathogens*, 16(5), pp. 1–24. doi: 10.1371/journal.ppat.1008572.

Mészáros, B., Sámano-Sánchez, H., Alvarado-Valverde, J., Čalyševa, J., Martínez-Pérez, E., Alves, R., Shields, D. C., Kumar, M., Rippmann, F., Chemes, L. B. and Gibson, T. J. (2021) 'Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications', *Science Signaling*, 14(665), p. eabd0334. doi: 10.1126/scisignal.abd0334.

Min, X., Akella, R., He, H., Humphreys, J. M., Tsutakawa, S. E., Lee, S. J., Tainer, J. A., Cobb, M. H. and Goldsmith, E. J. (2009) 'The Structure of the MAP2K MEK6 Reveals an Autoinhibitory Dimer', *Structure*, 17(1), pp. 96–104. doi: 10.1016/j.str.2008.11.007.

Mittelstadt, P. R., Yamaguchi, H., Appella, E. and Ashwell, J. D. (2009) 'T Cell Receptor-mediated Activation of p38α by Mono-phosphorylation of the Activation Loop Results in Altered Substrate Specificity', *Journal of Biological Chemistry*, 284(23), pp. 15469–15474. doi: 10.1074/jbc.M901004200.

Modi, V. and Dunbrack, R. L. (2019a) 'A Structurally-Validated Multiple Sequence Alignment of 497 Human Protein Kinase Domains', *Scientific Reports*, 9(1), pp. 1–16. doi: 10.1038/s41598-019-56499-4.

Modi, V. and Dunbrack, R. L. (2019b) 'Defining a new nomenclature for the structures of active and inactive kinases', *Proceedings of the National Academy of Sciences*, 116(14), pp. 6818–6827. doi: 10.1073/pnas.1814279116.

Moriguch, T., Toyoshima, F., Gotoh, Y., Iwamatsu, A., Irie, K., Mori, E., Kuroyanagi, N., Hagiwara, M., Matsumoto, K. and Nishida, E. (1996) 'Purification and identification of a major activator for p38 from osmotically shocked cells. Activation of mitogen-activated protein kinase kinase 6 by osmotic shock, tumor necrosis factor-α, and H2O2', *Journal of Biological Chemistry*, 271(43), pp. 26981–26988. doi: 10.1074/jbc.271.43.26981.

Mota, C. M., Oliveira, A. C. M., Davoli-Ferreira, M., Silva, M. V., Santiago, F. M., Nadipuram, S. M., Vashisht, A. A., Wohlschlegel, J. A., Bradley, P. J., Silva, J. S. J. S., Mineo, J. R. J. R. and Mineo, T. W. P. (2016) 'Neospora caninum activates p38 MAPK as an evasion mechanism against innate immunity', *Frontiers in Microbiology*, 7(SEP), pp. 1–14. doi: 10.3389/fmicb.2016.01456.

Mudgett, J. S., Ding, J., Guh-Siesel, L., Chartrain, N. A., Yang, L., Gopal, S. and Shen, M. M. (2000) 'Essential role for p38α mitogen-activated protein kinase in placental angiogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 97(19), pp. 10454–10459. doi: 10.1073/pnas.180316397.

Mukherjee, S., Keitany, G., Li, Y., Wang, Y., Ball, H. L., Goldsmith, E. J. and Orth, K. (2006) 'Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation', *Science*, 312(5777), pp. 1211–1214. doi: 10.1126/science.1126867.

Naor, A., Panas, M. W., Marino, N., Coffey, M. J., Tonkin, C. J. and Boothroyd, J. C. (2018) 'MYR1-Dependent Effectors Are the Major Drivers of a Host Cell's Early Response to Toxoplasma , Including Counteracting MYR1-Independent Effects', *mBio*. Edited by L. M.

Weiss, 9(2), pp. 1–17. doi: 10.1128/mBio.02401-17.

Nguyen, W., Hodder, A. N., de Lezongard, R. B., Czabotar, P. E., Jarman, K. E., O'Neill, M. T., Thompson, J. K., Jousset Sabroux, H., Cowman, A. F., Boddey, J. A. and Sleebs, B. E. (2018) 'Enhanced antimalarial activity of plasmepsin V inhibitors by modification of the P2 position of PEXEL peptidomimetics', *European Journal of Medicinal Chemistry*, 154, pp. 182–198. doi: 10.1016/j.ejmech.2018.05.022.

De Nicola, G. F., Martin, E. D., Chaikuad, A., Bassi, R., Clark, J., Martino, L., Verma, S., Sicard, P., Tata, R., Atkinson, R. A., Knapp, S., Conte, M. R. and Marber, M. S. (2013) 'Mechanism and consequence of the autoactivation of p38α mitogen-activated protein kinase promoted by TAB1.', *Nature structural & molecular biology*, 20(10), pp. 1182–90. doi: 10.1038/nsmb.2668.

Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. and Berger, I. (2014) 'Multiprotein complex production in insect cells by using polyproteins', *Methods in Molecular Biology*, 1091, pp. 131–141. doi: 10.1007/978-1-62703-691-7_8.

Nithianandarajah-Jones, G. N., Wilm, B., Goldring, C. E. P., Müller, J. and Cross, M. J. (2012) 'ERK5: Structure, regulation and function', *Cellular Signalling*, 24(11), pp. 2187–2196. doi: 10.1016/j.cellsig.2012.07.007.

Noble, M. E. M., Endicott, J. A. and Johnson, L. N. (2004) 'Protein Kinase Inhibitors: Insights into Drug Design from Structure', *Science*, 303(5665), pp. 1800–1805. doi: 10.1126/science.1095920.

Ouldridge, T. E. and Rein Ten Wolde, P. (2014) 'The robustness of proofreading to crowding-induced pseudo-processivity in the MAPK pathway', *Biophysical Journal*, 107(10), pp. 2425–2435. doi: 10.1016/j.bpj.2014.10.020.

Pan, C. Q., Sudol, M., Sheetz, M. and Low, B. C. (2012) 'Modularity and functional plasticity of scaffold proteins as p(l)acemakers in cell signaling', *Cellular Signalling*, 24(11), pp. 2143–2165. doi: 10.1016/j.cellsig.2012.06.002.

Panas, M. W. and Boothroyd, J. C. (2020) 'Toxoplasma Uses GRA16 To Upregulate Host c-Myc', *mSphere*. Edited by K. S. Ralston, 5(3), pp. 1–7. doi: 10.1128/mSphere.00402-20.

Panas, M. W. and Boothroyd, J. C. (2021) ' Seizing control: how dense granule effector proteins enable Toxoplasma to take charge ', *Molecular Microbiology*, pp. 0–2. doi: 10.1111/mmi.14679.

Park, E. C. and Rongo, C. (2016) 'The p38 MAP kinase pathway modulates the hypoxia response and glutamate receptor trafficking in aging neurons', *eLife*, 5, pp. 1–25. doi: 10.7554/elife.12010.

Park, E., Rawson, S., Li, K., Kim, B. W., Ficarro, S. B., Pino, G. G. Del, Sharif, H., Marto, J. A., Jeon, H. and Eck, M. J. (2019) 'Architecture of autoinhibited and active BRAF–MEK1–14-3-3 complexes', *Nature*, 575(7783), pp. 545–550. doi: 10.1038/s41586-019-1660-y.

Parton, D. L., Hanson, S. M., Rodríguez-laureano, L., Albanese, S. K., Gradia, S., Jeans, C., Seeliger, M., Levinson, N. and Chodera, J. D. (2016) 'An open library of human kinase domain constructs for automated bacterial expression', *bioRxiv*.

Patwardhan, P. and Miller, W. T. (2007) 'Processive phosphorylation: Mechanism and

biological importance', *Cellular Signalling*, 19(11), pp. 2218–2226. doi: 10.1016/j.cellsig.2007.06.006.

Pease, B. N., Huttlin, E. L., Jedrychowski, M. P., Talevich, E., Harmon, J., Dillman, T., Kannan, N., Doerig, C., Chakrabarti, R., Gygi, S. P. and Chakrabarti, D. (2013) 'Global Analysis of Protein Expression and Phosphorylation of Three Stages of Plasmodium falciparum Intraerythrocytic Development', *Journal of Proteome Research*, 12(9), pp. 4028–4045. doi: 10.1021/pr400394g.

Pellegrini, E. (2013) *Dissection of Enzymatic Phosphoryl Transfer : from Substrate Recognition to the Transition State*. University Of Sheffield.

Pellegrini, E., Palencia, A., Braun, L., Kapp, U., Bougdour, A., Belrhali, H., Bowler, M. W. and Hakimi, M.-A. (2017) 'Structural Basis for the Subversion of MAP Kinase Signaling by an Intrinsically Disordered Parasite Secreted Agonist', *Structure*, 25(1), pp. 16–26. doi: 10.1016/j.str.2016.10.011.

Pellegrini, E., Signor, L., Singh, S., Erba, E. B. and Cusack, S. (2017) 'Structures of the inactive and active states of RIP2 kinase inform on the mechanism of activation', *PLoS ONE*, 12(5), pp. 1–27. doi: 10.1371/journal.pone.0177161.

Peregrin, S., Jurado-Pueyo, M., Campos, P. M., Sanz-Moreno, V., Ruiz-Gomez, A., Crespo, P., Mayor, F. and Murga, C. (2006) 'Phosphorylation of p38 by GRK2 at the Docking Groove Unveils a Novel Mechanism for Inactivating p38MAPK', *Current Biology*, 16(20), pp. 2042–2047. doi: 10.1016/j.cub.2006.08.083.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L. and McSweeney, S. (2013) 'Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution', *Journal of Synchrotron Radiation*, 20(4), pp. 660–664. doi: 10.1107/S0909049513010431.

Peti, W. and Page, R. (2013) 'Molecular basis of MAP kinase regulation', *Protein Science*, 22(12), pp. 1698–1710. doi: 10.1002/pro.2374.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science*, 30(1), pp. 70–82. doi: 10.1002/pro.3943.

Piala, A. T., Humphreys, J. M. and Goldsmith, E. J. (2014) 'MAP Kinase Modules: The Excursion Model and the Steps that Count', *Biophysical Journal*, 107(9), pp. 2006–2015. doi: 10.1016/j.bpj.2014.09.024.

Pillai, V. B., Sundaresan, N. R., Samant, S. A., Wolfgeher, D., Trivedi, C. M. and Gupta, M. P. (2011) 'Acetylation of a Conserved Lysine Residue in the ATP Binding Pocket of p38 Augments Its Kinase Activity during Hypertrophy of Cardiomyocytes', *Molecular and Cellular Biology*, 31(11), pp. 2349–2363. doi: 10.1128/mcb.01205-10.

Pozhidaeva, A. and Bezsonova, I. (2019) 'USP7: Structure, substrate specificity, and inhibition', *DNA Repair*, 76, pp. 30–39. doi: 10.1016/j.dnarep.2019.02.005.

Punjani, A., Rubinstein, J. L., Fleet, D. J. and Brubaker, M. A. (2017) 'CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination', *Nature Methods*, 14(3), pp. 290–296. doi: 10.1038/nmeth.4169.

Quan, J. H., Ismail, H. A. H. A., Cha, G. H., Jo, Y. J., Gao, F. F., Choi, I. W., Chu, J. Q., Yuk, J. M. and Lee, Y. H. (2020) 'VEGF Production Is Regulated by the AKT/ERK1/2 Signaling Pathway and Controls the Proliferation of Toxoplasma gondii in ARPE-19 Cells', *Frontiers in Cellular and Infection Microbiology*, 10(April), pp. 1–18. doi: 10.3389/fcimb.2020.00184.

Rastogi, S., Cygan, A. M. and Boothroyd, J. C. (2019) 'Translocation of effector proteins into host cells by Toxoplasma gondii', *Current Opinion in Microbiology*, 52, pp. 130–138. doi: 10.1016/j.mib.2019.07.002.

Richter, J. P., Goroncy, A. K., Ronimus, R. S. and Sutherland-Smith, A. J. (2016) 'The structural and functional characterization of mammalian ADP-dependent glucokinase', *Journal of Biological Chemistry*, 291(8), pp. 3694–3704. doi: 10.1074/jbc.M115.679902.

Rivas-Pardo, J. A., Herrera-Morande, A., Castro-Fernandez, V., Fernandez, F. J., Vega, M. C. and Guixé, V. (2013) 'Crystal Structure, SAXS and Kinetic Mechanism of Hyperthermophilic ADP-Dependent Glucokinase from Thermococcus litoralis Reveal a Conserved Mechanism for Catalysis', *PLoS ONE*, 8(6). doi: 10.1371/journal.pone.0066687.

Robichon, A. (2020) 'Protein Phosphorylation Dynamics: Unexplored Because of Current Methodological Limitations: Dynamics of Processive Phosphorylation', *BioEssays*, 42(4), pp. 1–10. doi: 10.1002/bies.201900149.

Robinson, E., De Valk, H., Villena, I., Le Strat, Y. and Tourdjman, M. (2021) 'National perinatal survey demonstrates a decreasing seroprevalence of Toxoplasma gondii infection among pregnant women in France, 1995 to 2016: Impact for screening policy', *Eurosurveillance*, 26(5). doi: 10.2807/1560-7917.ES.2021.26.5.1900710.

Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J. and Davey, N. E. (2014) 'Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation', *Chemical Reviews*, 114(13), pp. 6733–6778. doi: 10.1021/cr400585q.

Ronimus, R. S. and Morgan, H. W. (2004) 'Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase', *Biochemical and Biophysical Research Communications*, 315(3), pp. 652–658. doi: 10.1016/j.bbrc.2004.01.103.

van Rooyen, J. M., Murat, J.-B. B., Hammoudi, P.-M. M., Kieffer-Jaquinod, S., Coute, Y., Sharma, A., Pelloux, H., Belrhali, H. and Hakimi, M.-A. A. (2014) 'Assembly of the novel five-component apicomplexan multi-aminoacyl-tRNA synthetase complex is driven by the hybrid scaffold protein Tg-p43', *PLoS ONE*. Edited by I. J. Blader, 9(2), p. e89487. doi: 10.1371/journal.pone.0089487.

Rosowski, E. E., Lu, D., Julien, L., Rodda, L., Gaiser, R. a, Jensen, K. D. C. and Saeij, J. P. J. (2011) 'Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein.', *The Journal of experimental medicine*, 208(1), pp. 195–212. doi: 10.1084/jem.20100717.

Rothweiler, U., Åberg, E., Johnson, K. A., Hansen, T. E., Jørgensen, J. B. and Engh, R. A. (2011) 'P38α MAP kinase dimers with swapped activation segments and a novel catalytic loop conformation', *Journal of Molecular Biology*, 411(2), pp. 474–485. doi: 10.1016/j.jmb.2011.06.013.

Rouse, J., Cohen, P., Trigon, S., Morange, M., Alonso-Llamazares, A., Zamanillo, D., Hunt, T. and Nebreda, A. R. (1994) 'A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins', *Cell*, 78(6), pp. 1027–1037. doi: 10.1016/0092-8674(94)90277-1.

Sakauchi, C., Wakatsuki, H., Ichijo, H. and Hattori, K. (2017) 'Pleiotropic properties of ASK1', *Biochimica et Biophysica Acta - General Subjects*. Elsevier B.V., pp. 3030–3038. doi: 10.1016/j.bbagen.2016.09.028.

Salvador, J. M., Mittelstadt, P. R., Guszczynski, T., Copeland, T. D., Yamaguchi, H., Appella, E., Fornace, A. J. and Ashwell, J. D. (2005) 'Alternative p38 activation pathway mediated by T cell receptor–proximal tyrosine kinases', *Nature Immunology*, 6(4), pp. 390–395. doi: 10.1038/ni1177.

Sámano-Sánchez, H. and Gibson, T. J. (2020) 'Mimicry of Short Linear Motifs by Bacterial Pathogens: A Drugging Opportunity', *Trends in Biochemical Sciences*, pp. 1–19. doi: 10.1016/j.tibs.2020.03.003.

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S. and Vargas, J. (2020) 'DeepEMhancer: A deep learning solution for cryo-EM volume post-processing', *bioRxiv*. doi: 10.1101/2020.06.12.148296.

Saridakis, V., Sheng, Y., Sarkari, F., Holowaty, M. N., Shire, K., Nguyen, T., Zhang, R. G., Liao, J., Lee, W., Edwards, A. M., Arrowsmith, C. H. and Frappier, L. (2005) 'Structure of the p53 binding domain of HAUSP/USP7 bound to epstein-barr nuclear antigen 1: Implications for EBV-mediated immortalization', *Molecular Cell*, 18(1), pp. 25–36. doi: 10.1016/j.molcel.2005.02.029.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9(7), pp. 671–675. doi: 10.1038/nmeth.2089.

Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. and Mastronarde, D. N. (2019) 'Software tools for automated transmission electron microscopy', *Nature Methods*, 16(6), pp. 471–477. doi: 10.1038/s41592-019-0396-9.

Schröder, M., Tan, L., Wang, J., Liang, Y., Gray, N. S., Knapp, S. and Chaikuad, A. (2020) 'Catalytic Domain Plasticity of MKK7 Reveals Structural Mechanisms of Allosteric Activation and Diverse Targeting Opportunities', *Cell Chemical Biology*, 27(10), pp. 1285-1295.e4. doi: 10.1016/j.chembiol.2020.07.014.

Schrodinger LLC (2015) 'The PyMOL Molecular Graphics System, Version 1.8'.

Schwartz, P. A. and Murray, B. W. (2011) 'Protein kinase biochemistry and drug discovery', *Bioorganic Chemistry*, 39(5–6), pp. 192–210. doi: 10.1016/j.bioorg.2011.07.004.

Seshacharyulu, P., Pandey, P., Datta, K. and Batra, S. K. (2013) 'Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer', *Cancer Letters*, 335(1), pp. 9–18. doi: 10.1016/j.canlet.2013.02.036.

Sharrocks, A. D., Yang, S. H. and Galanis, A. (2000) 'Docking domains and substrate-specificity determination for MAP kinases', *Trends in Biochemical Sciences*, 25(9), pp. 448–453. doi: 10.1016/S0968-0004(00)01627-3.

Sheng, Y., Saridakis, V., Sarkari, F., Duan, S., Wu, T., Arrowsmith, C. H. and Frappier, L. (2006) 'Molecular recognition of p53 and MDM2 by USP7/HAUSP', *Nature Structural and Molecular Biology*, 13(3), pp. 285–291. doi: 10.1038/nsmb1067.

Sheridan, D. L., Kong, Y., Parker, S. A., Dalby, K. N. and Turk, B. E. (2008) 'Substrate discrimination among mitogen-activated protein kinases through distinct docking sequence motifs', *Journal of Biological Chemistry*, 283(28), pp. 19511–19520. doi: 10.1074/jbc.M801074200.

Sievers, F. and Higgins, D. G. (2018) 'Clustal Omega for making accurate alignments of many protein sequences', *Protein Science*, 27(1), pp. 135–145. doi: 10.1002/pro.3290.

Slobodnyuk, K., Radic, N., Ivanova, S., Llado, A., Trempolec, N., Zorzano, A. and Nebreda, A. R. (2019) 'Autophagy-induced senescence is regulated by p38α signaling', *Cell Death and Disease*, 10(6). doi: 10.1038/s41419-019-1607-0.

Smith, N. C., Goulart, C., Hayward, J. A., Kupz, A., Miller, C. M. and van Dooren, G. G. (2021) 'Control of human toxoplasmosis', *International Journal for Parasitology*, 51(2–3), pp. 95–121. doi: 10.1016/j.ijpara.2020.11.001.

Song, N., Ma, J., Meng, X. W., Liu, H., Wang, H., Song, S. Y., Chen, Q. C., Liu, H. Y., Zhang, J., Peng, K. and Ji, F. H. (2020) 'Heat Shock Protein 70 Protects the Heart from Ischemia/Reperfusion Injury through Inhibition of p38 MAPK Signaling', *Oxidative Medicine and Cellular Longevity*, 2020. doi: 10.1155/2020/3908641.

Studier, F. W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification*, 41(1), pp. 207–234. doi: 10.1016/j.pep.2005.01.016.

Sun, J., Yi, M., Yang, L., Wei, W., Ding, Y. and Jia, Y. (2014) 'Enhancement of tunability of MAPK cascade due to coexistence of processive and distributive phosphorylation mechanisms', *Biophysical Journal*, 106(5), pp. 1215–1226. doi: 10.1016/j.bpj.2014.01.036.

Suwanmajo, T. and Krishnan, J. (2015) 'Mixed mechanisms of multi-site phosphorylation', *Journal of the Royal Society Interface*, 12(107). doi: 10.1098/rsif.2014.1405.

Svensson, O., Gilski, M., Nurizzo, D. and Bowler, M. W. (2018) 'Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1', *Acta Crystallographica Section D: Structural Biology*, 74(5), pp. 433–440. doi: 10.1107/S2059798318003728.

Svergun, D. I. (1992) 'Determination of the regularization parameter in indirect-transform methods using perceptual criteria', *Journal of Applied Crystallography*, 25(pt 4), pp. 495–503. doi: 10.1107/S0021889892001663.

Takekawa, M., Tatebayashi, K. and Saito, H. (2005) 'Conserved docking site is essential for activation of mammalian MAP kinase kinases by specific MAP kinase kinase kinases', *Molecular Cell*, 18(3), pp. 295–306. doi: 10.1016/j.molcel.2005.04.001.

Tanoue, T., Maeda, R., Adachi, M. and Nishida, E. (2001) 'Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions', *EMBO Journal*, 20(3), pp. 466–479. doi: 10.1093/emboj/20.3.466.

Tanoue, T. and Nishida, E. (2002) 'Docking interactions in the mitogen-activated protein kinase

cascades', *Pharmacology & Therapeutics*, 93(2–3), pp. 193–202. doi: 10.1016/S0163-7258(02)00188-2.

Taylor, S. S. and Radzio-Andzelm, E. (1994) 'Three protein kinase structures define a common motif', *Structure*, 2(5), pp. 345–355. doi: 10.1016/S0969-2126(00)00036-8.

Tegunov, D. and Cramer, P. (2019) 'Real-time cryo-electron microscopy data preprocessing with Warp', *Nature Methods*, 16(11), pp. 1146–1152. doi: 10.1038/s41592-019-0580-y.

Tesker, M., Selamat, S. E., Beenstock, J., Hayouka, R., Livnah, O. and Engelberg, D. (2016) 'Tighter αC-helix-αL16-helix interactions seem to make p38α less prone to activation by autophosphorylation than Hog1.', *Bioscience reports*, 36(2), pp. 1–18. doi: 10.1042/BSR20160020.

Tzarum, N., Eisenberg-Domovich, Y., Gills, J. J., Dennis, P. A. and Livnah, O. (2012) 'Lipid molecules induce p38α activation via a novel molecular switch', *Journal of Molecular Biology*, 424(5), pp. 339–353. doi: 10.1016/j.jmb.2012.10.007.

Valère, A., Garnotel, R., Villena, I., Guenounou, M., Pinon, J. M. and Aubert, D. (2003) 'Activation of the cellular mitogen-activated protein kinase pathways ERK, P38 and JNK during Toxoplasma gondii invasion', *Parasite*, 10(1), pp. 59–64. doi: 10.1051/parasite/2003101p59.

Vázquez, B., Soto, T., del Dedo, J. E., Franco, A., Vicente, J., Hidalgo, E., Gacto, M., Cansado, J. and Madrid, M. (2015) 'Distinct biological activity of threonine monophosphorylated MAPK isoforms during the stress response in fission yeast', *Cellular Signalling*, 27(12), pp. 2534–2542. doi: 10.1016/j.cellsig.2015.09.017.

Visconti, R., Gadina, M., Chiariello, M., Chen, E. H., Stancato, L. F., Gutkind, J. S. and O'Shea, J. J. (2000) 'Importance of the MKK6/p38 pathway for interleukin-12-induced STAT4 serine phosphorylation and transcriptional activity', *Blood*, 96(5), pp. 1844–1852. doi: 10.1182/blood.v96.5.1844.

Wagner, T. *et al.* (2019) 'SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM', *Communications Biology*, 2(1), pp. 1–13. doi: 10.1038/s42003-019-0437-z.

Wang, C., Deng, L., Hong, M., Akkaraju, G. R., Inoue, J. and Chen, Z. J. (2001) 'TAK1 is a ubiquitin-dependent kinase of MKK and IKK', *Nature*, 412(6844), pp. 346–351. doi: 10.1038/35085597.

Wang, J., Wu, J. W. and Wang, Z. X. (2011) 'Structural insights into the autoactivation mechanism of p21-activated protein kinase', *Structure*, 19(12), pp. 1752–1761. doi: 10.1016/j.str.2011.10.013.

Wang, Y., Zhang, Y., Lu, C., Zhang, W., Deng, H., Wu, J., Wang, J. and Wang, Z. (2019) 'Kinetic and mechanistic studies of p38α phosphorylation by MKK6', *The FEBS Journal*, 286(5), pp. 1030–1052. doi: 10.1111/febs.14762.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: Homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46(W1), pp. W296–W303. doi: 10.1093/nar/gky427.

Weaver, B. P., Weaver, Y. M., Omi, S., Yuan, W., Ewbank, J. J. and Han, M. (2020) 'Non-

Canonical Caspase Activity Antagonizes p38 MAPK Stress-Priming Function to Support Development', *Developmental Cell*, 53(3), pp. 358-369.e6. doi: 10.1016/j.devcel.2020.03.015.

Widmann, C., Gibson, S., Jarpe, M. B. and Johnson, G. L. (1999) 'Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human.', *Physiological reviews*, 79(1), pp. 143–80. doi: 10.1152/physrev.1999.79.1.143.

Wilsbacher, J. L. and Cobb, M. H. (2001) 'Bacterial expression of activated mitogen-activated protein kinases', in *Methods*, pp. 387–400. doi: 10.1016/S0076-6879(01)32217-6.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) 'Overview of the CCP 4 suite and current developments', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 235–242. doi: 10.1107/S0907444910045749.

Witzel, F. and Blüthgen, N. (2018) 'When More Is Less: Dual Phosphorylation Protects Signaling Off State against Overexpression', *Biophysical Journal*, 115(7), pp. 1383–1392. doi: 10.1016/j.bpj.2018.08.019.

Wybenga-Groot, L. E., Ho, C. S., Sweeney, F. D., Ceccarelli, D. F., McGlade, C. J., Durocher, D. and Sicheri, F. (2014) 'Structural basis of Rad53 kinase activation by dimerization and activation segment exchange', *Cellular Signalling*, 26(9), pp. 1825–1836. doi: 10.1016/j.cellsig.2014.05.004.

Zander, U., Hoffmann, G., Cornaciu, I., Marquette, J. P., Papp, G., Landret, C., Seroul, G., Sinoir, J., Röwer, M., Felisaz, F., Rodriguez-Puente, S., Mariaule, V., Murphy, P., Mathieu, M., Cipriani, F. and Márquez, J. A. (2016) 'Automated harvesting and processing of protein crystals through laser photoablation', *Acta Crystallographica Section D: Structural Biology*, 72(4), pp. 454–466. doi: 10.1107/S2059798316000954.

Zeke, A., Bastys, T., Alexa, A., Garai, A., Mészáros, B., Kirsch, K., Dosztányi, Z., Kalinina, O. V and Reményi, A. (2015) 'Systematic discovery of linear binding motifs targeting an ancient protein interaction surface on MAP kinases.', *Molecular systems biology*, 11(11), p. 837. doi: 10.15252/msb.20156269.

Zhang, Y. Y., Wu, J. W. and Wang, Z. X. (2011) 'Mitogen-activated protein kinase (MAPK) phosphatase 3-mediated cross-talk between MAPKs ERK2 and p38α', *Journal of Biological Chemistry*, 286(18), pp. 16150–16162. doi: 10.1074/jbc.M110.203786.

Zhao, Y. O., Khaminets, A., Hunn, J. P. and Howard, J. C. (2009) 'Disruption of the Toxoplasma gondii parasitophorous vacuole by IFNγ-inducible immunity-related GTPases (IRG proteins) triggers necrotic cell death', *PLoS Pathogens*, 5(2). doi: 10.1371/journal.ppat.1000288.

Zheng, B., Lou, D., Ding, J., Zhuo, X., Ding, H., Kong, Q. and Lu, S. (2019) 'GRA24-Based DNA Vaccine Prolongs Survival in Mice Challenged With a Virulent Toxoplasma gondii Strain', *Frontiers in Immunology*, 10(March), pp. 1–11. doi: 10.3389/fimmu.2019.00418.

Zhou, B., Wang, Z.-X., Zhao, Y., Brautigan, D. L. and Zhang, Z.-Y. (2002) 'The specificity of extracellular signal-regulated kinase 2 dephosphorylation by protein phosphatases.', *The Journal of biological chemistry*, 277(35), pp. 31818–25. doi: 10.1074/jbc.M203969200.

Zhou, T., Sun, L., Humphreys, J. and Goldsmith, E. J. (2006) 'Docking Interactions Induce

Exposure of Activation Loop in the MAP Kinase ERK2', *Structure*, 14(6), pp. 1011–1019. doi: 10.1016/j.str.2006.04.006.

Zivanov, J., Nakane, T., Forsberg, B., Kimanius, D., Hagen, W. J. H., Lindahl, E. and Scheres, S. H. W. (2018) 'RELION-3: New tools for automated high-resolution cryo-EM structure determination', *bioRxiv*, pp. 1–22. doi: 10.1101/421123.