



The
University
Of
Sheffield.

Dissection of Enzymatic Phosphoryl Transfer: from Substrate Recognition to the Transition State

Thesis Submitted for the Degree of Doctor of Philosophy

Erika Pellegrini BSc MSc

Department of Molecular Biology and Biotechnology
The University Of Sheffield

March 2013

Abstract

Phosphoryl transfer is the most common class of biological reaction. Enzymes catalysing this type of reaction, such as kinases, mutases and small G proteins, are pharmaceutical targets in cancer, inflammation and infection therapy. The characterisation of the structural and energy framework of the enzymatic transition state is crucial for understanding enzyme catalysis, which is essential for successful drug design studies. In this thesis, a combination of ^{19}F -NMR and X-ray crystallography was used to define the structure of phosphoryl transfer enzymes in complex with metallofluorides, such as AlF_4^- and MgF_3^- , which are analogues of the transition state of the transfer of the phosphoryl group. In this thesis three different biological systems have been investigated.

The bacterial mutase β -Phosphoglucomutase (β PGM) catalyses the interconversion of β -D-glucose-1-phosphate (β G1P) and D-glucose-6-phosphate (G6P). The structural experiments performed on this protein, have led to the characterisation of:

- the first observed magnesium fluoride species bound in octahedral geometry in an enzyme active site;
- the discrimination in binding between the G1P and G6P using a α -fluorophosphonate analogue of β G1P, which is a potential lead compound for combating bacterial infections.

RhoA belongs to the Rho family of small G proteins, which cycles between an ‘on’ (GTP bound) and ‘off’ (GDP-bound) state. RhoGAP promotes the inactive conformation of RhoA by stimulating its intrinsic GTPase activity through the so called arginine finger. The structures solved in this thesis provide new information on this system by revealing:

- the contribution to catalysis by RhoGAP binding, in the absence of the arginine finger;
- the role of Tyr34_{RhoA}, one of the most conserved residues in small G protein, in RhoA switch cycle.

p38 α and MEK6 belong to one of the MAPK signalling cascades, where p38 α is the MAPK (Mitogen activated protein kinase) and MEK6 is its activator (MAPKK). Through structural and biochemical studies, how MEK6 binds and activates p38 α by phosphorylation, was investigated. The studies have led to the characterisation of:

- structural rearrangements caused by the docking interaction of a MEK6 D-motif peptide to the p38 α D-motif binding site;
- the discovery that MEK6 can use ADP as phosphate donor to activate p38 α .

Acknowledgements

I would like to express my gratitude towards all of those whom, over these past three years have helped me to achieve this important goal in my scientific career.

Firstly, I would like to thank my supervisors Dr. Matthew W. Bowler and Prof. Jon P. Waltho for giving me the opportunity to come to Grenoble and work on such an exciting scientific project, this made all my days working in the lab a pleasure rather than a chore. I'm particularly grateful for Matthew's support as it is with his patience and constant supervision which has made me a better scientist day by day.

Also, through my collaboration with Prof Jon Waltho's group in Sheffield I had the pleasure of meeting Prof. Mike Blackburn and Dr. Yi Jin, to whom I am grateful for the amazing time we spent working together and discussions over a lovely glass of wine. Without their help part of my Thesis would not exist.

Thank you to those who contributed to the results presented in this Thesis; Dr. J. Griffin and Dr. Jin Y. for the NMR work, Dr. M. Weiss for assisting me during my experiment at the BESSY synchrotron and Dr. D. Bhattasali for the synthesis of the phosphonates. In addition, thanks to all the people that sustain the PSB services with particular regards to the HTX laboratory.

I would also like to thank all the people who generously dedicated their time to me. *In primis*, all the Structural Biology group of Dr. Sean McSweeney in Grenoble. A special thanks to Dario who followed my first steps in the lab, Chloe for the helpful and inspiring scientific discussions, Samira and Ulrike for always being available to help me and to Mats, Tobias, Joanna, Steffie, Silvia, Salyha and everyone who has helped me during my PhD.

Thank you to my friends who have supported me emotionally such as Luca, Aili, Mats, Tobias, Ulrike, Taiana, Otilie, Samira, Michael, Nicolas, Filippo, Tommaso, Bjorn, Natasha and others. A special thanks goes to Gaëlle and Yi for their true friendship in always finding the time for me no matter what I needed.

It goes without saying that this PhD couldn't have been completed without the unconditional and kind support of my family who have always encouraged me to follow my dreams of being a scientist, even if it meant being far away from them. Indeed, the support of my mother, father and sister was one of the most precious supports I had over these years.

Finally, I would like to thank Paolo, who more than others knows how much time and energy this PhD and in particular the writing up of this Thesis took. I thank him for his tolerance, constant support and for reinforcing me day by day. This Thesis is dedicated to you.

Contents

Abstract	p v
Acknowledgements	p vii
Contents	p ix
List of Figures	p xvii
List of Tables	p xxii
List of Structures	p xxiv
List of Abbreviations	p xxvi
1. Chapter 1 Introduction	p 1
1.1 Relevance of studying phosphoryl transfer enzymes	p 1
1.2 The role of phosphorus and phosphoryl transfer in cell biology.....	p 2
1.2.1 Phosphate diesters and monoesters are fundamental building blocks in several macromolecules	p 3
1.2.2 Phosphorus is involved in energy production and transfer	p 3
1.2.3 Phosphorus is involved in protein regulation and signal transduction	p 4
1.3 How do enzymes work?	p 8
1.3.1 The enzymatic transition state	p 9
1.3.2 Enzymes lower the activation energy through binding energy.....	p 11
1.4 How do phosphoryl transfer enzymes achieve catalysis?	p 12
1.4.1 Mechanism possibilities: Associative-Concerted-Dissociative phosphoryl transfer transition state	p 13
1.4.2 Un-catalysed phosphoryl transfer reactions adopt a loose transition state	p 16
1.4.3 Enzyme challenges in accommodating the un-catalysed loose transition state	p 18
1.4.4 Enzymatic TS are studied through metallofluoride complexes	p 20
1.4.5 Activation of the nucleophile	p 24
1.4.5.1 Positioning	p 24
1.4.5.2 Increasing the nucleophilicity.....	p 25
1.4.5.2 Overcome the electrostatic repulsion	p 26
1.4.6 Stabilisation of negative charge development on the leaving group	p 26
1.4.7 Stabilisation of the charge and geometry of the phosphorus group in flight: the charge balance hypothesis	p 28

1.4.8 Is the enzymatic transition state dissociative or associative?	p 31
1.5 Subjects of this thesis.....	p 32
1.5.1 β -Phosphoglucomutase from <i>L. lactis</i>	p 32
1.5.2 RhoA GTP-hydrolysis	p 33
1.5.3 p38 α and MEK6	p 34
Chapter 2 Materials and Methods	p 37
2.1 General materials	p 37
2.2 Molecular Biology	p 37
2.2.1 Materials	p 37
2.2.2 Transformation of competent cells	p 38
2.3 Gel electrophoresis	p 38
2.3.1 Agarose gel electrophoresis	p 38
2.3.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	p 39
2.3.3 Native PAGE gel electrophoresis	p 40
2.4 Macromolecules quantification.....	p 40
2.4.1 UV adsorption	p 41
2.4.2 The Bradford assay	p 41
2.5. Protein expression and purification	p 42
2.5.1 Plasmid description	p 43
2.5.2 Transformation of competent cells	p 43
2.5.3 Large scale protein expression and protein harvesting.....	p 44
2.5.4 Cell lysis	p 44
2.5.5 Protein purification	p 44
2.5.5.1 Affinity chromatography and Thrombin cleavage	p 45
2.5.5.2 Size exclusion chromatography	p 46
2.5.5.3 Anion exchange chromatography.....	p 46
2.5.5.4 Buffer exchange	p 47
2.5.6 Dynamic light scattering	p 47
2.6 X-ray Crystallography	p 48
2.6.1 Crystallisation.....	p 48
2.6.2 Cryocooling	p 51
2.6.3 X-ray Crystallography analysis	p 52

2.6.3.1	Brief description of beamlines used	p 52
2.6.3.2	Data collection	p 53
2.6.3.3	Data processing	p 54
2.6.3.4	Data reduction	p 55
2.6.3.5	Molecular replacement	p 56
2.6.3.5	Model refinement	p 57
2.6.3.6.1	Refinement using Refmac5 and phenix.refine	p 58
2.6.3.6.2	Manual rebuilding	p 59
2.6.3.7	Structure validation.....	p 60
2.7	Figure preparation	p 61
Chapter 3 The Major and the Minor conformers in βPGM-G6P-MgF₄⁻-TSA		p 63
3.1	Introduction	p 63
3.1.1	β PGM is a pharmacological target	p 63
3.1.2	In <i>L. lactis</i> β PGM belongs to maltose and trehalose pathways.....	p 64
3.1.4	Structural characterisation of β PGM.....	p 65
3.1.4	The reaction catalysed by β PGM	p 67
3.1.4.1	β PGM activation.....	p 69
3.1.2.4	Isomerisation induce through an intermediate.....	p 70
3.1.2.4	Possible inhibition of the β PGM reaction.....	p 70
3.1.4	Challenges in studying the β PGM reaction.....	p 71
3.2	Aim of this project.....	p 73
3.3	Experimental procedures.....	p 74
3.3.1	Cloning, protein expression and purification.....	p 74
3.3.2	Crystallisation of β PGM-G6P-MgF ₃ ⁻ -TSA and β PGM _{K145A} -G6P-MgF _x -TSA complexes	p 74
3.3.4	Data collection at 100 K, structure solution and refinement	p 75
3.3.4	Data collection at room temperature, structure solution and refinement.....	p 75
3.4	Results and Discussion.....	p 77
3.4.1	¹⁹ F-NMR spectra revealed an alternative way of coordinating magnesium fluoride in the β PGM-G6P-MgF ₃ ⁻ -TSA complex.....	p 77
3.4.2	The β PGM _{K145A} -G6P-MgF _x -TSA structure is related to the minor form.....	p 80
3.4.4	Structure solved from room temperature data collection	

reveals the structure features of the minor form conformer	p 86
3.5 Conclusions	p 90

Chapter 4 Discrete analysis of both steps of the β PGM reaction using phosphonate analogues of Glucose 1- β -Phosphate p 91

4.1 Introduction	p 91
4.1.1 Challenges in trapping β PGM- β G1P-TSA complex	p 91
4.1.2 α -fluorophosphonates and fluorine in medicinal chemistry	p 92
4.1.2.1 How does fluorine atom affect bioavailability?	p 93
4.1.2.2 How does introduction of a fluorine atom affect protein binding affinity?	p 95
4.2 Aim of this project	p 99
4.3 Experimental procedures	p 100
4.3.1 Design and synthesis of the phosphonate analogues and ^{19}F -NMR experiments	p 100
4.3.2 Fluorescent titration binding assay	p 100
4.3.3 Crystallisation of, β PGM-YO5- MgF_3^- -TSA, β PGM-GRX- MgF_3^- -TSA and β PGM-GRX- AlF_3^- -TSA	p 101
4.3.4 Data collection, structure solution and refinement	p 101
4.4 Results and Discussion	p 102
4.4.1 Design and synthesis of β G1P phosphonate analogues	p 102
4.4.2 ^{19}F -NMR, Binding studies, and X-ray crystallography identified two viable candidates	p 103
4.4.2.1 ^{19}F -NMR experiments and binding studies	p 103
4.4.2.2 X-ray crystallography	p 105
4.4.2.2.1 Structure characterisation of β PGM- $\beta\text{CH}_2\text{G1P-MgF}_3^-$ -TSA	p 107
4.4.2.2.2 Specificity for the recognition of $\beta\text{CH}_2\text{G1P}$ and G6P	p 108
4.4.2.2.3 Structural characterisation of β PGM-(<i>S</i>)- $\beta\text{CHFG1P-TSA}$ structures	p 111
4.4.3 (<i>S</i>)- βCHFG1P is the best ligand for TSA formation	p 113
4.4.3.1 Why do phosphonates with a 1- α -hydroxyl phosphonate not bind?	p 113
4.4.3.2 Why does (<i>S</i>)- βCHFG1P bind more strongly than $\beta\text{CH}_2\text{G1P}$ to β PGM?	p 114

4.4.3.3 Why is there specificity for the fluorine (<i>S</i>)-stereochemistry?.....	p 116
4.5 Conclusions	p 118
Chapter 5 Novel insight into RhoA GTP-hydrolysis	p 119
5.1 Introduction	p 119
5.1.1 RhoA is a small G protein	p 119
5.1.1.1 The Ras superfamily of monomeric G proteins.....	p 119
5.1.1.2 The Small G protein switching cycle and structure correlation	p 120
5.1.1.3 Physiological roles of mammalian Rho family	p 123
5.1.1.4 The Rho family role in cancer and further diseases	p 126
5.1.2 How do GAPs stimulate GTP-hydrolysis?.....	p 128
5.2 Aim of this project.....	p 131
5.3 Experimental procedures.....	p 132
5.3.1 RhoA-GDP-AlF _x -TSA and RhoA-GDP	p 132
5.3.1.1 RhoA expression and purification	p 132
5.3.1.2 RhoA-GDP-AlF _x -TSA and RhoA-GDP crystallisation.....	p 134
5.3.1.3 Data collection, structure solution and refinement	p 135
5.3.2 RhoA-GDP-AlF _x -RhoGAP _{R85A} -TSA and RhoA-GDP-MgF _x -RhoGAP _{R85A} -TSA	p 135
5.3.2.1 Site-directed mutagenesis of RhoGAP to generate R85A mutant.....	p 135
5.3.2.2 RhoA and RhoGAP _{R85A} expression and purification.....	p 136
5.3.2.3 Generation and crystallisation of the RhoA-GDP-AlF _x -RhoGAP _{R85A} -TSA complex	p 137
5.3.2.4 Generation and crystallisation of RhoA-GDP-MgF _x -RhoGAP _{R85A} -TSA complex.....	p 139
5.3.2.5 Data collection, structure solution, and refinement	p 139
5.3.2.6 Sulphur-SAD data collection.....	p 139
5.4 Results and Discussion.....	p 141
5.4.1 RhoA-GDP-AlF ₃ -TSA crystallisation	p 141
5.4.2 The structures of RhoA- GDP-AlF ₄ ⁻ -RhoGAP _{R85A} -TSA and RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA	p 146
5.4.3 RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA and Rho-GDP provide novel insight into the GTP-hydrolysis reaction.....	p 150

5.4.3.1	The stabilisation step.....	p 151
5.4.3.2	The catalytic step.....	p 157
5.4.3.3	The realising step	p 157
5.4.4	Residual activity in RhoA and RhoA-RhoGAP _{R85A}	p 158
5.4.4.1	The intrinsic activity of RhoA	p 159
5.4.4.1	The RhoGAP _{R85A} residual activity	p 159
5.4.4.1.1	Potential counter ions around α -phosphate.....	p 160
5.4.4.1.2	The tightening effect around γ -phosphate during catalysis	p 163
5.5	Conclusions	p 165
Chapter 6 The effect of the MEK6 docking interaction on p38α		p 169
6.1	Introduction	p 169
6.1.1	MAPKs belong to the MAPKs cascade, a druggable pathway	p 169
6.1.2	Docking interactions are relevant in MAPK-specific drug design.....	p 172
6.1.3	The docking interaction by MAPKK promotes MAPK activation	p 174
6.1.3.1	Structural features of active state of protein kinase	p 175
6.1.3.2	Structural characterisation of the inactive and active form of p38	p 176
6.1.3.3	D-motif binding site in the inactive p38.....	p 181
6.1.3.4	How does the MAPKK docking interaction promote p38 activation?	p 182
6.2	Aim of this project.....	p 186
6.3	Experimental procedures	p 187
6.3.1	Plasmids.....	p 187
6.3.1.1	MEK6DD cloning in pET15b	p 188
6.3.1.2	Mutagenesis of p38 α in p38 α K53R	p 188
6.3.2	Protein expression and purification.....	p 189
6.3.3	Crystallisation experiments	p 193
6.3.4	Data collection, structure solution and refinement.....	p 193
6.4	Results and Discussion	p 194
6.4.1	The structure of p38 α K53R.....	p 195
6.4.1.1	p38 α K53R is an inactive mutant.....	p 196
6.4.1.2	The functional motifs of p38 α are well ordered in p38 α K53R structure	p 197
6.4.1.3	Is auto-inhibition in the p38 α K53R a crystal artefact?	p 201
6.4.2	The structure of p38 α K53R bound to the docking peptide of MEK6.....	p 202

6.4.2.1 MEK6 peptide binds in the D-motif binding site of p38 α K53R	p 203
6.4.2.2 How does the peptide bind specifically to the protein?	p 207
6.4.2.2.1 MEK6 binds differently form MKK3b and MEF2A in the hydrophobic docking groove	p 207
6.4.2.2.2 MEK6 shows unique binding with Loop16.....	p 212
6.4.3 Peptide-induced conformational changes in p38 α K53R	p 213
6.4.3.1 Conformational changes at the D-motif binding site.....	p 214
6.4.3.2 The peptide binding induce conformational changes at P-lip, DGF motif and Glycine-rich loop.....	p 214
6.5 Conclusions	p 215
6.5.1 The R-4 binding pocket of p38 α , a third element of the D-motif binding interaction	p 218
6.5.2 MEK6 might adopt a “scanning process” for recruiting p38 α	p 220
6.5.3 How does the phosphorylation lip become disordered?	p 222
Chapter 7 The non-canonical MEK6 phosphorylation pathway	p 225
7.1 Introduction	p 225
7.1.1 How does MAPKK double phosphorylate MAPK?	p 225
7.1.1.1 Processive vs distributive mechanism	p 225
7.1.1.2 In <i>vitro</i> vs in <i>vivo</i>	p 226
7.2 Aim of this project.....	p 229
7.3 Experimental procedures.....	p 230
7.3.1 Plasmids and protein expression-purification.....	p 230
7.3.2 Native PE gels.....	p 230
7.3.2.1 Native PE gels composition.....	p 230
7.3.2.2 Sample preparation for Native PE gel	p 231
7.3.3 Western blots of native PE gels	p 231
7.3.4 Kinetic assays.....	p 233
7.3.4.1 Phosphorylation Reactions	p 233
7.3.4.2 Phosphorylation reaction analysis	p 233
7.3.5 Nucleotides analogues	p 234
7.4 Results and Discussion.....	p 236

7.4.1 MEK6-p38 α complex run in a double band on native PAGE gel	p 236
7.4.2 ADP is a secondary phosphate donor in MEK6 phosphorylation reaction	p 238
7.4.2.1 Analysis of the native PAGE gels by mass spectrometry	p 239
7.4.2.2 Analysis of the native PAGE gels by western blot	p 242
7.4.2.3 Alternative explanations to the non-canonical phosphorylation pathway	p 243
7.4.3 The functional role of MEK6-ADP phosphorylation	p 244
7.4.4 Preliminary crystallographic results on Δ MEK6DD	p 249
7.5 Conclusions	p 252
Chapter 8 Perspective	p 256
8.1 How can the study of transition state complexes, which comprise fluoride analogues, benefit from the combination of ^{19}F -NMR and X-ray crystallography?	p 257
8.2 Is the charge balance hypothesis universal?	p 258
8.3 How is substrate recognition involved in catalysis?	p 260
Appendix A Crystallographic theory	p 263
A.1 Introduction	p 263
A.2 Hardware	p 263
A.2.1 Synchrotrons	p 263
A.2.2 Detectors	p 266
A.2.2.1 CCD detector	p 266
A.2.2.2 PILATUS detector	p 267
A.3 Crystal properties	p 269
A.4 Diffraction from protein crystals	p 271
A.5 Structure solution	p 274
A.5.1 From Structure Factor to Electron Density	p 274
A.5.2 The phase problem	p 276
A.5.2.1 Isomorphus replacement and MAD/SAD	p 277
A.5.2.2 Molecular replacement	p 280
A.5.1 Model building, Refinement and Validation	p 281
References	p 285

List of Figures

Figure 1.1 Examples of phosphate esters and phosphoanhydrides.	p 1
Figure 1.2 The six general types of signal transducers	p 6
Figure 1.3 Reaction coordinates diagram.....	p 10
Figure 1.4 Strategies adopted by enzymes for lowering the energy barrier.....	p 12
Figure 1.5 Hypothetical reaction mechanisms for phosphoryl transfer.	p 14
Figure 1.6 The More-O'Ferrall-Jencks diagram for phosphoryl transfer transition state	p 15
Figure 1.7 Transition state charge estimates for the un-catalysed hydrolysis of ATP.....	p 19
Figure 1.8 Transition state analogues in GTP-hydrolysis	p 21
Figure 1.9 Aluminium and magnesium fluoride in RhoA-RhoGAP transition state structures.....	p 22
Figure 1.10 Approximate distances for differing degrees of bonding to nucleophile/leaving group in transition states for phosphoryl transfer	p 24
Figure 1.11 Electrostatic surfaces for the ATP tetra-anion (left) and the AMP dianion (right).	p 26
Figure 1.12 Model of the transition state interactions in the reaction catalysed by <i>E. coli</i> alkaline phosphatase	p 27
Figure 1.13 Octahedral moiety of AlF_3 in the HsPGK _{K219A} -3PG- AlF_3 -ADP-TSA complex.....	p 30
Figure 2.1 Schematic view of solubility curve of a protein	p 49
Figure 2.2 Sitting drop and microbatch methods	p 50
Figure 3.1 The closely connected maltose and trehalose metabolic pathways of <i>L. lactis</i>	p 65
Figure 3.2 Ribbon representation of the cap domain movement	p 66
Figure 3.3 Schematic view of the β PGM-G6P-MgF ₄ ⁻ -TSA complex active site.....	p 67
Figure 3.4 Reactions catalysed by β PGM	p 68
Figure 3.5 Residues involved in General acid-base catalysis in β PGM.	p 68
Figure 3.6 Simplified reaction scheme of the reaction catalysed by β PGM in the presence of α G16BP as a cofactor	p 71
Figure 3.7 Crystal of β PGM _{K145A} -G6P-MgF _x ⁻ -TSA	p 75
Figure 3.8 ¹⁹ F-NMR spectra of β PGM-G6P-MgF ₄ ⁻ -TSA complex	p 78
Figure 3.9 Optimised geometry of MgF ₃ ⁻ in the gas phase (A) and in aqueous solution (B)	p 80

Figure 3.10 ^{19}F -NMR spectra of $\beta\text{PGM-G6P-MgF}_4^-$ -TSA and $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_2\text{-TSA}$ complexes	p 81
Figure 3.11 Difference Fourier map of the magnesium fluoride molecule	p 83
Figure 3.12 Magnesium is coordinated in octahedral geometry	p 84
Figure 3.13 View of the protein coordination of W-F _B	p 85
Figure 3.14 Overlay of the two magnesium fluoride structures on $\beta\text{PGM-G6P-MgF}_3^-$ -TSA Fourier map	p 86
Figure 3.15 View of $\beta\text{PGM-TSA}$ structure collected at room temperature.....	p 87
Figure 3.16 Geometric details of the Major and Minor form	p 89
Figure 4.1 Structure of G1P in the $\beta\text{PGM-G6P-BeF}_3^-$ active site.....	p 92
Figure 4.2 Development of Ezetimibe by optimisation of a lead compound	p 94
Figure 4.3 Structure of Fluorithromycin.....	p 95
Figure 4.4 Structure of UDP-galactose-4'-epimerase from <i>Trypanosoma brucei</i> in complex with the substrate analogue UDP-4'-deoxy-4'-fluoro- α -D-galactose (UFG).....	p 96
Figure 4.5 Binding of a fluorinated inhibitor to p38 kinase	p 96
Figure 4.6 Schematic representation of C-F...H-X bond	p 98
Figure 4.7 Chemical structures of βG1P analogues.....	p 102
Figure 4.8 ^{19}F -NMR spectra of (a) $\beta\text{PGM-}\beta\text{CH}_2\text{G1P-MgF}_3^-$ -TSA complex and (b) $\beta\text{PGM-(S)-}\beta\text{CHFG1P- MgF}_4^-$ -TSA complex	p 104
Figure 4.9 Fluorescence titration curve from a plot of binding fraction versus [ligand].....	p 104
Figure 4.10 Active site of the $\beta\text{PGM-}\beta\text{CH}_2\text{G1P-MgF}_3^-$ -TSA complex.....	p 107
Figure 4.11 Asp10 in the $\beta\text{PGM-}\beta\text{CH}_2\text{G1P-MgF}_3^-$ -TSA complex.....	p 108
Figure 4.12 G6P and $\beta\text{CH}_2\text{G1P}$ in βPGM active site	p 108
Figure 4.13 Schematic view of the $\beta\text{PGM-}\beta\text{CH}_2\text{G1P -MgF}_3^-$ -TSA and $\beta\text{PGM-G6P-MgF}_4^-$ -TSA	p 109
Figure 4.14 Structural details of the interaction between the region of residues 146-144 and the clamp loop	p 110
Figure 4.15 Active sites of the $\beta\text{PGM-(S)-}\beta\text{CHFG1P-MgF}_3^-$ -TSA and $\beta\text{PGM-(S)-GRX-AlF}_4^-$ -TSA complex.....	p 111
Figure 4.16 Schematic view of the $\beta\text{PGM-(S)-}\beta\text{CHFG1P -MgF}_3^-$	p 112
Figure 4.17 Structural details of the $\beta\text{PGM-}\beta\text{CH}_2\text{G1P-MgF}_3^-$ -TSA active site and βPGM region 136-144	p 113

Figure 4.18 Surface representation of His ₂₀ and βCHFG1P in βPGM-(S)-βCHFG1P-MgF ₃ ⁻ -TSA structure.....	p 114
Figure 4.19 Protons in the electron density of βPGM-(S)-βCHFG1P-MgF ₃ ⁻ -TSA structure	p 115
Figure 4.20 The C-F···H-N bond in βPGM-(S)-βCHFG1P-MgF ₃ ⁻ -TSA structure.....	p 116
Figure 4.21 VdW clash in (R)-βCHFG1P	p 117
Figure 5.1 The Ras superfamily G protein cycle	p 121
Figure 5.2 Structure of Switch I and II regions in RhoA _{G14V} -GTPγS and RhoA-GDP structures.....	p 122
Figure 5.3 The effects of Rac, Rho, and Cdc42 on actin organization in fibroblasts	p 124
Figure 5.4 Rho family signalling in cell motility	p 126
Figure 5.5 Structure conformation of conserved tyrosine in different GAP protein.....	p 130
Figure 5.6 Size exclusion chromatography profile of RhoA	p 132
Figure 5.7 12% SDS-PE of RhoA purification – initial tests	p 133
Figure 5.8 Crystals obtained from initial robot conditions	p 134
Figure 5.9 Size exclusion chromatography profile of RhoGAP	p 136
Figure 5.10 RhoA-GDP-AIF _x -RhoGAP _{R85A} -TSA crystals optimisation	p 138
Figure 5.11 ¹⁹ F NMR spectra of RhoA-GDP-AIF ₃ -TSA	p 142
Figure 5.12 The structure of RhoA-GDP at 1.3 Å resolution	p 144
Figure 5.13 Crystal packing contacts between RhoA molecules	p 145
Figure 5.14 Structure of Switch II region in RhoA-GDP and RhoA-GDP-1FTN.....	p 146
Figure 5.15 Difference electron density in the active site of the structures of RhoA-GDP-AIF ₄ ⁻ -RhoGAP _{R85A} -TSA and RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA.....	p 149
Figure 5.16 ¹⁹ F NMR spectrum of the RhoA-GDP-AIF ₄ ⁻ -RhoGAP _{R85A} -TSA and the RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA complexes.....	p 150
Figure 5.17 GTPγS bound to RhoA in the RhoA _{G14V} -GTPγS structure.....	p 152
Figure 5.18 Stabilisation and conformational changes of Tyr34 _{RhoA} in the Switch I region	p 153
Figure 5.19 Stabilisation and conformational changes in the Switch II region	p 154
Figure 5.20 Allosteric effects of RhoGAP binding.....	p 156
Figure 5.21 The Arginine finger flips out the Tyr34 _{RhoA}	p 157
Figure 5.22 Dramatic changes of the Switch regions in RhoA-GDP.....	p 158
Figure 5.23 GTP solvation	p 162

Figure 5.24 Snapshots of the RhoGAP mediated RhoA GTP-hydrolysis	p 166
Figure 5.25 Switch I conformational changes from the GTP-bound state to the GDP-bound state.....	p 168
Figure 6.1 MAPK cascades in mammals	p 170
Figure 6.2 The human MAPK family tree.....	p 171
Figure 6.3 Secondary structure in p38 α	p 175
Figure 6.4 Functional motifs in p38 α	p 177
Figure 6.5 p38 active site.....	p 178
Figure 6.6 DFG-in conformation in the active site of p38 γ	p 179
Figure 6.7 Activation of p38 α	p 179
Figure 6.8 D-motif binding domain in p38 α	p 182
Figure 6.9 Conformational changes induced by docking interaction in inactive p38 α	p 183
Figure 6.10 Conformational changes induced by docking interaction in inactive ERK2	p 184
Figure 6.11 Size exclusion chromatography profile of Δ MEK6DD, MEK6 and MEK6DD	p 190
Figure 6.12 Anionic exchange chromatography of p38 α K53R.....	p 191
Figure 6.13 Crystals of p38 α K53R and pepMEK6- p38 α _{K53R}	p 194
Figure 6.14 Effects of mutation K53R on the Glycine-rich loop	p 197
Figure 6.15 Possible effects of mutation K53R in the active site of active p38 conformation	p 198
Figure 6.16 Electron density map of functional motifs in p38 α K53R	p 199
Figure 6.17 Functional motifs in the structure of p38 α K53R.....	p 200
Figure 6.18 The N-terminal hinge	p 200
Figure 6.19 Docking interaction in symmetry related molecules of p38K53R	p 202
Figure 6.20 Peptide sequence nomenclature.....	p 204
Figure 6.21 B-factor analysis of MEK6 peptide.....	p 205
Figure 6.22 MEK6 peptide bound to the p38 α D-motif binding groove	p 206
Figure 6.23 Differences in MEK6 peptide conformation	p 207
Figure 6.24 Differences in MEK6, MEF2A and MKK3b peptide conformation.....	p 209
Figure 6.25 Hydrogen bonds in the inter-Pro peptide sequence.....	p 210
Figure 6.26 Hydrophobic interactions of the inter-Pro peptide sequence	p 211
Figure 6.27 Electrostatic interactions of MEK6 peptide N-terminus peptide	p 213
Figure 6.28 Conformational changes at the peptide binding site	p 214

Figure 6.29 Rearrangement of the lobes upon peptide binding	p 215
Figure 6.30 Conformational changes induced by peptide binding.....	p 216
Figure 6.31 Peptide binding in p38 α	p 219
Figure 6.32 R-pocket residues in p38 α , ERK2 and JNK1	p 220
Figure 6.33 Conformational changes of Tyr35	p 223
Figure 7.1 Steps performed to quantify gel bands	p 234
Figure 7.2 Nucleotides and nucleotides analogues	p 235
Figure 7.3 Complex detection on a native PAGE gel	p 237
Figure 7.4 Native PE gel on MEK6-p38 α K53R, MEK6DD-p38 α K53R, MEK6DD-p38 α in presence of ATP or ADP	p 240
Figure 7.5 Native PE gels 7.5% performed on MEK6DD-p38 α and MEK6DD-p38 α T180A	p 241
Figure 7.6 Western blot results	p 242
Figure 7.7 Effects of μ M or mM amount of Ap ₅ A on p38 α phosphorylation in presence of ADP	p 244
Figure 7.8 Kinetic experiments in the presence of ATP, ADP or ApCpp	p 246
Figure 7.9 Endpoint kinetics in the presence of ATP, ADP or ApCpp.....	p 247
Figure 7.10 Endpoint kinetics for ADP-ApCpp.....	p 249
Figure 7.11 Overlay of the four ANP molecules in the MEK4 structure.....	p 250
Figure A.1 The European Synchrotron Radiation facility (ESRF)	p 265
Figure A.2 The beamline experimental hutches.....	p 265
Figure A.3 CCD detector	p 267
Figure A.4 The PILATUS detector detects X-ray without readout noise	p 268
Figure A.5 A unit cell within the crystal lattice	p 269
Figure A.6 Schematic of the mosaicity in the protein crystal	p 271
Figure A.7 The Bragg's law in the diffraction pattern	p 272
Figure A.8 A two-dimensional representation of the Ewald sphere	p 273
Figure A.9 The Structure Factor	p 275
Figure A.10 The Patterson map.....	p 276
Figure A.11 Structure Factor of the heavy atom structure.....	p 277
Figure A.12 The Harker construction	p 278
Figure A.13 The Selenium k absorption edge.....	p 280

List of Tables

Table 2.1 SDS PAGE 12% (w/v) composition	p 40
Table 2.2 Essential features of the ESRF beam lines	p 53
Table 3.1 Simplified reaction scheme of the reaction catalysed by β PGM in the presence of α G16BP as a cofactor	p 71
Table 3.2 ^{19}F -NMR spectra of β PGM-G6P-MgF ₄ ⁻ -TSA and β PGM _{K145A} -G6P-MgF ₂ -TSA complexes	p 81
Table 3.4 Data collection and refinement statistics	p 82
Table 3.4 Difference Fourier map of the magnesium fluoride molecule.....	p 83
Table 3.5 Data collection and refinement statistics	p 88
Table 3.6 Geometric details of the Major and Minor form	p 89
Table 4.1 Effect of fluorine substitution on pK _a and pK _b values	p 94
Table 4.2 Data collection and refinement statistics	p 106
Table 4.3 Rmsd of C α displacement among (S)- β CHFG1P, β CH ₂ G1P and G6P structures	p 111
Table 5.1 Protein purification steps	p 133
Table 5.2 Crystals obtained from initial robot conditions	p 134
Table 5.3 Data collection and refinement statistics	p 143
Table 5.4 The structure of RhoA-GDP at 1.3 Å of resolution.....	p 144
Table 5.5 Data collection and refinement statistics	p 148
Table 5.6 Tightening effects	p 164
Table 6.1 Structure-based sequence alignment of the D-motifs	p 173
Table 6.2 Protein purification steps	p 192
Table 6.3 Data collection and refinement statistics	p 196
Table 6.4 Data collection and refinement statistics	p 203
Table 6.5 B-factor analysis of MEK6 peptide	p 205
Table 6.6 Alignment of MEK6, MKK3b and MEF2A peptides	p 208

Table 7.1 Native gel composition table.....	p 230
Table 7.2 Native PAGE gel sample composition.....	p 231
Table 7.3 Western blot protocol.....	p 232
Table 7.4 Endpoint kinetics in presence of ATP, ADP or ApCpp.....	p 247
Table 7.5 Preliminary crystallization results for MKK6 Δ	p 251
Table A.1 The seven crystal systems	p 270

List of Structures

In this list are reported structures solved in this thesis, and published structures, which are included in the discussion sections. Structures are reported in alphabetic order.

1. **Active p38 γ** : structure of bis-phosphorylated p38 γ from human. AMP-PNP is bound to the active site (PDB code 1CM8, Bellon *et al.*, 1999).
2. **β PGM^P**: structure of phosphorylated β PGM (P-Asp8) from *L. lactis* (PDB code 1O03, Lahiri *et al.*, 2003).
3. **β PGM-YO5-MgF₃⁻-TSA** or **β PGM- β CH₂G1P-MgF₃⁻-TSA**: structure of WT β PGM from *L. lactis* in complex with phosphonate analogue of β G1P (β CH₂G1P) and MgF₃⁻ (**solved in this thesis**).
4. **β PGM-GRX-MgF₃⁻-TSA** or **β PGM-(S)- β CHFG1P-MgF₃⁻-TSA**: structure of WT β PGM from *L. lactis* in complex with fluorophosphonate analogue of β G1P ((S)- β CHFG1P) and MgF₃⁻ (**solved in this thesis**).
5. **β PGM-GRX-AlF₄⁻-TSA** or **β PGM-(S)- β CHFG1P-AlF₄⁻-TSA**: structure of WT β PGM from *L. lactis* in complex with fluorophosphonate analogue of β G1P ((S)- β CHFG1P) and AlF₄⁻ (**solved in this thesis**).
6. **β PGM-G6P-MgF₃⁻-TSA**: structure of WT β PGM from *L. lactis* in complex with G6P and MgF₃⁻ (PDB code 2WF5, Baxter *et al.*, 2003).
7. **β PGM_{K145A}-G6P-MgF₂(H₂O)-TSA**: structure of mutated β PGM from *L. lactis* in complex with G6P and MgF₂(H₂O). The protein carries the mutation K145A. The magnesium is coordinated by two fluorine atoms and one water molecule (**solved in this thesis**).
8. **β PGM-G6P-MgF₃⁻-TSA_{RT}**: structure of WT β PGM from *L. lactis* in complex with G6P and MgF₃⁻. Data collection was performed at RT (**solved in this thesis**).
9. **HsPGK_{K219A}-3PG-AlF₃-ADP-TSA**: structure of human and mutated PGK in complex with 3PG (3-phospho-glycerate), ADP and AlF₃. The protein carries the mutation K219A. The aluminium is coordinated by three fluorine atoms and one water molecule (PDB code 2WZD, Cliff *et al.*, 2010).
10. **human p38 α** : structure of un-phosphorylated WT p38 α from human (PDB code 1WFC, Wilson *et al.*, 1996a).

11. **murine p38 α** : structure of un-phosphorylated WT p38 α from mouse (PDB code 1P38, Wang *et al.*, 1997).
12. **p38 α K53R**: structure of un-phosphorylated p38 α from human. The protein carries the mutation K53R (**solved in this thesis**).
13. **pepMEF2A-p38 α** : structure of un-phosphorylated WT p38 α from mouse, in complex with a D-motif peptide from MEF2A. MEF2A is a transcription factor (PDB code 1LEW, Chang *et al.*, 2002).
14. **pepMKK3b-p38 α** : structure of un-phosphorylated WT p38 α from mouse, in complex with a D-motif peptide from MKK3b. MKK3b is a MAPKK (PDB code 1LEZ, Chang *et al.*, 2002).
15. **pepMEK6-p38 α _{C162S}**: structure of un-phosphorylated and mutated p38 α from human, in complex with a D-motif peptide from MEK6. The protein carries the mutation C162A. MEK6 is a MAPKK (PDB codes 2Y8O, Garai *et al.*, 2012).
16. **pepMEK6-p38 α _{K53R}**: structure of un-phosphorylated and mutated p38 α from human, in complex with a D-motif peptide from MEK6. The protein carries the mutation C162A. MEK6 is a MAPKK (**solved in this thesis**).
17. **RhoA-GDP**: structure of RhoA from human, in complex with GDP (PDB code 1FTN, Wei *et al.*, 1997; **solved in this thesis** at higher resolution).
18. **RhoA-GDP-AlF₄⁻-RhoGAP-TSA**: structure of RhoA from human, in complex with WT RhoGAP, GDP and AlF₄⁻ (PDB code 1TX4, Rittinger *et al.*, 1997).
19. **RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA**: structure of RhoA from human, in complex with mutated RhoGAP, GDP and AlF₄⁻. The RhoGAP protein carries the mutation R85A (**solved in this thesis**).
20. **RhoA-GDP-MgF₃⁻-RhoGAP-TSA**: structure of RhoA from human, in complex with WT RhoGAP, GDP and MgF₃⁻ (PDB code 1OW3, Graham *et al.*, 2002).
21. **RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA**: structure of RhoA from human, in complex with mutated RhoGAP, GDP and MgF₃⁻. The RhoGAP protein carries the mutation R85A (**solved in this thesis**).
22. **RhoA_{G14V}-GTP γ S**: structure of constitutively active RhoA from human, in complex with GTP γ S. The protein carries the mutation G14V (PDB code 1A2B, Ihara *et al.*, 1999).

List of abbreviations

Å	Ångström (1 Å = 0.1 nm)
AlF ₃	aluminium trifluoride
AlF ₄ ⁻	aluminium tetrafluoride
AMPPCP	Adenosine-5'-[(β-γ) methylene]diphosphate
AMPPNP	Adenosine-5'-[(β-γ) imido]diphosphate
ApCp (or AMPCP)	Adenosine-5'-[(α-β) methylene]diphosphate
ApCpp (or AMPCPP)	Adenosine-5'-[(α-β) methylene]triphosphate
AP	Alkaline Phosphatase
APS	ammonium persulfate
ATP, ADP, AMP	adenosine tri-, di-, mono-phosphate
β	Brønsted value
BB	binding buffer
BeF ₃ ⁻	beryllium fluoride
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
BSA	bovine serum albumin
CD	common domain
CH ₂ βG1P or YO5	methylenephosphonate-D-glucopyranose
(<i>R</i>)-βCHFG1P, (<i>S</i>)-βCHFG1P	monofluoro-methylenephosphonate-D-glucopyranose
C-terminal	carboxy terminal
DLS	dynamic light scattering
D-motif	docking motif
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ε	molecular extinction coefficient
E	enzyme
EB	elution buffer
ED	glutamate-aspartate motif
EDNA	enhanced automated collection of data
EDTA	Ethylenediaminetetraacetic acid
EP	enzyme-product
ES	enzyme-substrate
ESRF	European Synchrotron Radiation Facility
EMBL	European Molecular Biology Laboratory
ERK	extracellular signal-regulated kinase
F _o	observed structure factors
F _s	Structure factor
FFT	fast Fourier transformation
¹⁹ F-NMR	Fluorine Nuclear Magnetic Resonance
G	free energy

ΔG^\ddagger	activation energy
$\Delta G'^\circ$	difference in energy between ES-G and EP-G
G1P	D-glucose-1-phosphate
G16BP	D-glucose-1,6-biphosphate
G6P	D-glucose-6-phosphate
GABC	general acid base catalysis
GAP	GTPase activating protein
GEF	guanidine nucleotide exchange factor
cGMP	cyclic guanosine monophosphate
GPCR	G protein coupled receptor
GST-tag	glutathione-S-transferase tag
GTP, GDP	guanosine tri-, di-phosphate
GTPase	guanosine triphosphatase
HAD superfamily	haloacid dehalogenase superfamily
H-bonds	hydrogen bonds
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HC1b	EMBL humidity control device
His ₆ -tag	hexa-histidine tag
HTX laboratory	EMBL High Throughput Crystallization Laboratory
I	Intensity
IPTG	Isopropyl- β -D-thiogalactopyranoside
JNK/SAPK	c-Jun N-terminal kinase / Stress activated protein kinase
K_{cat}	turnover number
KIE	kinetic isotope effect
k_m	Michaelis constant
LB	Lysogeny broth
LFER	linear free energy relationship
LyB	lysis buffer
MAPK	mitogen-activated protein kinase
MAPKK/ MEK# /MKK#	kinase of MAPK
MAPKKK	kinase of MAPKK
MgF ₂	magnesium difluoride
MgF ₃ ⁻	magnesium trifluoride
MR	molecular replacement
NCS	non-crystallographic symmetry
Ni-NTA column	Nickel-nitriloacetic acid column
NMR	Nuclear Magnetic Resonance
N-terminus	amino terminus
OD ₆₀₀	optical density at 600 nm
ON	over night
P	reaction product
P-	phosphorylation

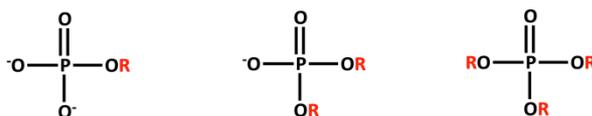
PCR	polymerase chain reaction
PDB	protein data bank
PDI	polydispersity index
PEG	polyethylene glycol
PGK	Phosphoglycerate kinase
PGM	phosphoglucomutase
PGM ^P	phospho-beta-phosphoglucomutase
Pi	inorganic phosphate
PIXE	proton-induced X-ray emission spectroscopy
pKa	protonation constant
RAS	Rat sarcoma
RH	Relative Humidity
RHO	Ras homologous
rmsd	root mean square deviation
RNA	Ribonucleic Acid
rpm	revolutions per minute
RT	Room Temperature
RTK	Receptor tyrosine kinase
S	substrate
SEC	size exclusion chromatography
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
TCB	Thrombin cleavage buffer
TBP	trigonal bipyramidal geometry
TBE	Tris-Boric Acid-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TPO	binding site for phosphorylation of Thr180 in p38 α
TF	translational function
TLS	Translation Libration Screw-motion
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TS	Transition State
TSA	Transition State Analogue
U	units of enzyme activity
UV	ultra violet
VdW	Van der Waals
VSEPR	Valence Shell Electron Pair Repulsion
v/v	volume per volume
W	water molecule
w/v	weight per volume
X	secondary messenger
YPO	binding site for phosphorylated Tyr182 in p38 α

Chapter 1 Introduction

1.1 Relevance of studying phosphoryl transfer enzymes

Phosphoryl transfer is the most common reaction catalysed by enzymes (Admiraal & Herschlag, 1995). Reactions at the phosphorus atom of phosphate esters and anhydrides form the chemical basis for many of the most fundamental processes in living systems. Typical phosphate monoesters are glucose or inositol phosphate, diesters are DNA and RNA. Common phosphoanhydrides are ATP or GTP, well known enzyme substrates and energy suppliers (Knowles, 1980; Lassila *et al.*, 2011; Westheimer, 1987) (Figure 1.1).

Phosphate esters



Phosphoanhydrides

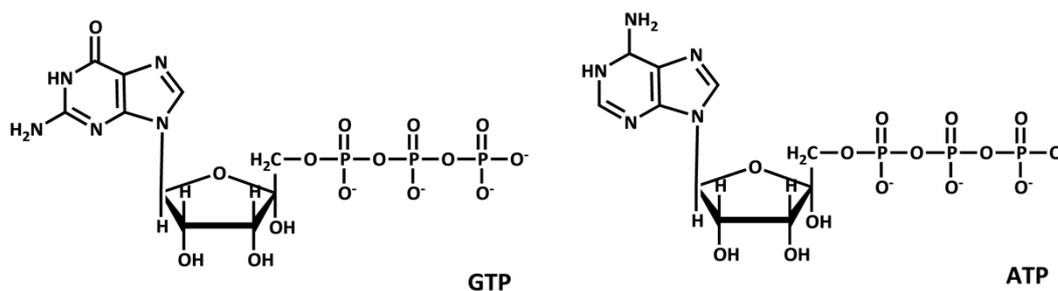


Figure 1.1 Examples of phosphate esters and phosphoanhydrides. Phosphate esters can be mono-, di- or tri- based on the number of R groups covalently linked to the phosphate. Phosphoanhydrides contain one or more anhydride linkages between the phosphate groups. The protonation state shown is the dominant form at pH 7-8 (Lassila *et al.*, 2011).

One of the most relevant aspects of phosphoric acid is that phosphorous is employed by nature in many aspects of cell biology. For example, it is used as raw material for building stable molecules such as nucleic acids and lipids, or as transient element in post-translational modifications, such as regulatory signalling or energy production (Bowler *et al.*, 2010a). Many specific phosphoryl transfer enzymes are employed in these processes. Enzymes are a specific class of proteins or ribozymes that are considered the most selective and powerful biological catalysts known (Alberts, 2008; Nelson *et al.*, 2008). In phosphoryl transfer, enzymes are fundamental, because they catalyse a reaction which has some of the slowest non-enzymatic rates in solution and thus require enormous rate accelerations (Lad *et al.*, 2003; Lassila *et al.*, 2011; Schroeder *et al.*, 2006; Wolfenden & Snider, 2001). Indeed, the understanding of how these biological catalysts promote and perform the phosphoryl transfer reaction has become a fundamental issue for investigating the relevance of phosphorus in life science.

The goal of this introduction is to provide an overview of the *state of the art* of enzymatic phosphoryl transfer. First, the relevance of phosphorus and phosphoryl transfer enzymes in life science will be discussed. Secondly, theoretical concepts of enzymatic catalysis will be explored with emphasis on the investigations reported on the catalysed phosphoryl transfer reaction. The role that metallofluoride complexes have assumed in this context will be discussed. Finally, a general background will be reviewed for understanding the relevance of the results contained in this thesis, which provided new knowledge of biological systems involving phosphoryl transfer.

1.2 The role of phosphorus and phosphoryl transfer in cell biology

Phosphoryl transfer reactions are involved throughout cell biology. Phosphorus is a stable part of biological elements, it is involved in energy production and transfer, it is widely used in the regulation of protein activity and signal transduction. All these cell aspects involve enzymes, which have been classified by the specific phosphoryl reaction performed.

1.2.1 Phosphate diesters and monoesters are fundamental building blocks in several macromolecules

Since the 1970s, the structural roles for phosphate diesters and monoesters have been well understood. Phosphate diesters, which are stable and easy to repair, are employed in both DNA and RNA. Monoesters, such as phospholipids, are membrane building blocks. Phosphorus itself is used as a mineral component of the skeletal structure. Several classes of enzymes are involved in the metabolism of these phosphor-elements. **Phosphodiesterases** catalyse the hydrolysis of nucleic acids, at the phosphoester linkage (Cleland & Hengge, 2006). **Phospholipases** are enzymes that hydrolyse phospholipids into fatty acids and other lipophilic substances (Nelson *et al.*, 2008). Aspartate Transcarbamoylase, which belong to the **Carbamoyl and Carboxyl Transferases** family, catalyses carbamoyl transfer from carbamoyl-P to aspartate to give carbamoylaspartate, providing the raw materials for pyrimidine ring synthesis (Gerhart & Pardee, 1962). Part of the **Kinases** and **Phosphatases** families are also involved in metabolism of phosphor-elements. They catalyse, respectively, the addition or hydrolysis of a phosphate group onto or off small molecules, such as sugars and nucleotides.

1.2.2 Phosphorus is involved in energy production and transfer

Cell metabolism is the result of the coordination of two opposing streams of chemical reactions: the catabolic and anabolic pathways. Catabolic reactions break down foodstuffs into smaller molecules, thereby generating both a useful form of energy for the cell and some of the small molecules that the cell needs as building blocks. Anabolic reactions use the energy produced by catabolism to drive the synthesis of the many other molecules that form the cell. The energy released in catabolism, or driven force, is stored as chemical bond energy in a small set of molecules named carrier molecules. These molecules diffuse rapidly throughout the cell and provide energy through two different mechanisms: transfer of specific chemical group or transfer of electrons/protons. The phosphoanhydrides ATP and GTP belong to the first group. In these molecules energy is realised by the hydrolysis of one or more phosphate groups. ATP is the most common currency of energy in all cells and several enzymes, such as

Kinases, couple reactions with ATP hydrolysis. **ATPases**, such as myosin in muscle tissue, use the energy released by ATP hydrolysis to drive motion (Smith & Rayment, 1996).

ATP production or glycolysis, is indeed the fundamental catabolic chain reaction, which occurs in both aerobic and anaerobic organisms. The pathway produces the energy necessary for synthesising ATP. After the breakdown of food molecules into sugars, glucose enters glycolysis, a sequence of reactions which produces 2 ATP molecules for each glucose molecule without the involvement of molecular oxygen (O₂ gas). Phosphoryl transfer enzymes are involved in production of glucose and glycolysis itself. For example, **Mutases** catalyse the interconversion of sugar-phosphate molecules through a phosphor-protein intermediate. Phosphoglucomutases catalyse the interconversion of G1P to G6P and *vice versa*. β Phosphoglucomutases, one of the subjects of this thesis, catalyses the interconversion of β -D-glucose-1-phosphate (β G1P) and D-glucose-6-phosphate (G6P) in bacteria. Furthermore, the same enzyme converts G6P to glucose and inorganic phosphate (Golicnik *et al.*, 2009). Finally, in oxidative phosphorylation energy from the Krebs cycle in mitochondria, the most remarkable phosphoryl transfer enzyme is **ATP Synthase**, which catalyses the formation of ATP from ADP and phosphate, using a gradient of protons, or in some cases Na⁺, to provide the energy for the synthesis (Senior *et al.*, 2002).

1.2.3 Phosphorus is involved in protein regulation and signal transduction

The covalent addition of a phosphate group is a method commonly used by eukaryotic cells to regulate a protein's function. It induces conformational changes that switch a protein's state to the active one. The removal of the phosphoryl group returns the protein to the original state. Furthermore, transient phosphorylation is used for changing the geometrical and charge properties of a specific protein region. This induces the formation of a "module", a specific recognition zone for another protein. These processes are typically regulated by two main classes of enzymes: **Kinases** and **Phosphatases**. Protein kinases catalyse the transfer of a phosphate group from a nucleotide, such as ATP, to a protein substrate. In eukaryotic cells, protein kinases primarily phosphorylate serine, threonine, or tyrosine residues. In prokaryotes, histidine

and aspartate can also be phosphorylated by the so-called two-component mechanism, in which a phosphoryl group is transferred from ATP first to histidine and then to an aspartate (Stock *et al.*, 2000). In contrast, protein phosphatases catalyse the hydrolysis of phosphate monoesters, by transferring a phosphoryl group to a water molecule, producing inorganic phosphate (Cleland & Hengge, 2006).

Kinase and Phosphatase reactions are involved in cell signal transduction. A third class of phosphoryl transfer enzymes participate in this process, which are called **Regulatory Proteins** (Cleland & Hengge, 2006). These are, for example, small GTP proteins, in which the active *vs* inactive state is regulated by GTP hydrolysis.

Signal transduction is the process that foresees the conversion of the “signal” information into a chemical change. The signal is a molecule or ligand detected by specific receptors. Then the interaction between the activated receptor and the cellular machinery produces a second signal or a change in the activity of a cellular process, which in turn is converted into a specific cell response. There are six basic receptor types, which are briefly described in figure 1.2.

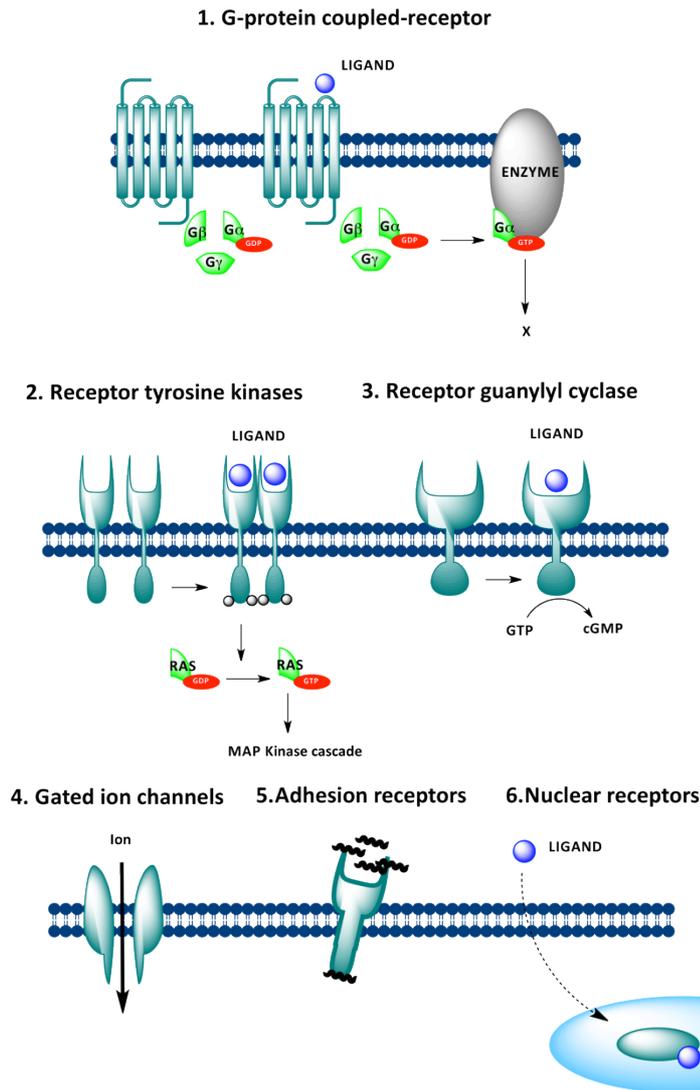


Figure 1.2 The six general types of signal transducers. There are six basic receptor types:

- 1. G protein coupled receptors (GPCR)**, which activate the second signal or messenger through the activation of heterometric G proteins, which in turn regulate an enzyme that generates the second messenger X.
- 2. Receptor tyrosine kinases (RTKs)**, which catalyse the phosphorylation of downstream effector proteins. Ligand binding activates the receptor by promoting dimerisation and auto-phosphorylation of the receptor. Typically, a small G protein, such as RAS, is activated which in turn activates a kinase cascade.
- 3. Receptor guanylyl cyclase**, which act through cyclic guanosine monophosphate (cGMP).
- 4. Gated ions channels** of the plasma membrane, which open and close in response to the binding of chemical ligands or changes of trans membrane potential.
- 5. Adhesion receptors**, which interact with macromolecular complex components of the extracellular matrix.
- 6. Nuclear receptors**, that bind specific ligands, which are able to pass the membrane by diffusion and alter gene expression.

The representation is simplified for clarity (Nelson *et al.*, 2008).

Signal transduction has several features: high specificity, high sensitivity, desensitisation and integration (Alberts, 2008; Nelson *et al.*, 2008). Phosphoryl transfer is one of the most used reaction in the mechanisms behind these features.

High **specificity** is achieved by precise molecular complementarity between the signal and receptor molecules. It is mediated by the same kinds of weak (non-covalent) forces that mediate enzyme-substrate and antigen-antibody interactions. In multicellular organisms further specificity is achieved by the fact that receptors or the intracellular targets of a given signal pathway, are present only in certain cell types. Phosphorylation can change the protein complementarity, by affecting the chemical and geometrical features of the regions involved in the binding. Examples are G proteins, which are proteins that cycle between a GTP or a GDP bound state. The two states are structurally different. Only the GTP bound state, or active state, is able to bind other enzymes, altering its activity. There are two distinct families of G proteins: heterotrimeric G proteins and small G proteins. Heterotrimeric G proteins, also called Large G proteins, are activated by G protein-coupled receptors and are made up of α , β , and γ subunits. Small G proteins are monomeric proteins homologous to the α -subunit found in the heterotrimers and are activated by RTK or GPCR signalling effectors (Figure 1.2). RhoA, one of the subjects of this thesis, is a small G protein, that belongs to the Ras superfamily of monomeric G proteins.

High **sensitivity** is achieved through three different factors: the high affinity of receptors for the signal molecule (in the picomolar range), cooperativity in the ligand-receptor interaction, and amplification of the signal by enzyme cascades. Signal amplification through a signal cascade results when an enzyme associated with a signal receptor is activated and in turn catalyses the activation of many molecules of a second enzyme, each of which activates many molecules of a third enzyme and so on. The activation process is often through a phosphorylation reaction. A typical cascade is the Mitogen activated protein kinase (MAPK) cascade, which is activated by both GPCR and RTK signalling. It comprises three kinases (MAPKKK, MAPKK and MAPK), which phosphorylate, in turn, each other. MEK6 and p38 α , which are one of the subjects of this thesis, belong to one of these cascades in mammals. p38 α is a MAPK, while MEK6 is its activator, a MAPKK.

Desensitisation is achieved by receptor system modification. When a signal is present

continuously, the receptor activation triggers a feedback circuit that shuts off one of the receptor effectors, the receptor itself or removes the receptor from the cell surface. Controlled GTP hydrolysis in small G proteins or of the G α subunits of Large G protein promotes the inactive conformation. In this thesis, how RhoGAP stimulates the intrinsic GTP hydrolysis in RhoA-GTP will be discussed. Furthermore, a receptor could be switched-off by de-phosphorylation of a specific site, when phosphorylation was responsible for an active state. Stronger desensitisation is achieved by phosphorylation of an intracellular domain, which will facilitate receptor sequestration from the plasma membrane.

Finally, **Integration**, is the ability of the system to receive multiple signals and produce a unified response appropriate to the needs of the cell or organism. Different signalling pathways converse with each other at several levels, generating a wealth of interactions that maintains homeostasis in the cell and in the organism.

1.3 How do enzymes work?

Most of the chemical reactions occurring in living organisms are energetically favourable. However, the reactants are usually in a relatively stable state and cannot be changed to a lower energy state without an input of energy, this is called the activation energy. In a living cell, enzymes can drastically reduce this energy barrier, which prevents a specific reaction from occurring. By reducing the energy barrier reaction, times can be increased or become feasible. A remarkable example is alkaline phosphatase which catalyses the hydrolysis of a methyl phosphate dianion with a rate acceleration of more than 10^{27} -fold (Westheimer, 1987). The investigation of this biocatalytic system is one of the most important and fundamental goals of life science. An understanding of an enzyme's mechanism provides a critical tool for the discovery of new drugs, for the large-scale industrial synthesis of useful chemicals, and for the appreciation of the chemistry in cells and organisms (Alberts, 2008).

1.3.1 The enzymatic transition state

Linus Pauling (Pauling, 1948) suggested that the incredible catalytic rate enhancement of enzymes could be achieved by the ability of enzymes to bind tightly to substrates distorted toward the highest energy state in the reaction coordinates, which is called the transition state (TS).

In a potential energy diagram (Figure 1.3) where the free energy (G) is plotted against the progress of a reaction, the transition state (or activated complex) resides on the highest point of the energy surface. In an un-catalysed reaction, the starting point is the reactants A and B. A certain amount of free energy, named the activation energy (ΔG^\ddagger), is required to obtain the activated complex or TS. Then the reaction goes on to product at an extremely rapid rate (Lowry & Richardson, 1987). In an enzyme-catalysed reaction, the reactants are called substrates (S). The starting point of the reaction is the ground state (ES), where Substrate (S) is bound to Enzyme (E), in a specific region called the active site. Enzymes increase the reaction rate by decreasing the activation energy that would be necessary if the catalyst was absent. Then the reaction goes on to the Enzyme-Product (EP) complex. Afterwards, the enzyme releases the product (Nelson *et al.*, 2008).

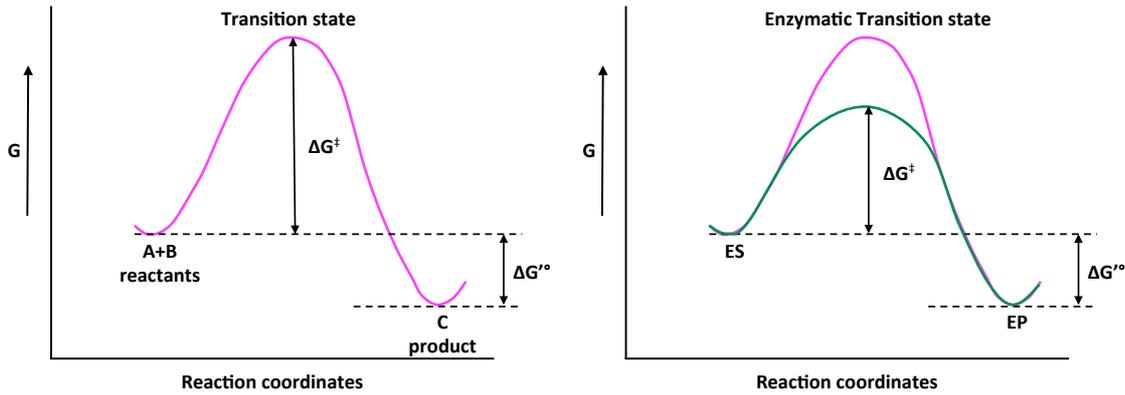


Figure 1.3 Reaction coordinates diagram. The reaction coordinate for an un-catalysed reaction (in *magenta*) and an enzyme-catalysed reaction (in *green*) are shown. The curves plot the free energy of the reaction complex as it progresses along the reaction coordinate from reactants to product. Reactants/ES and products/EP reside in energy wells on this surface, and the lowest energy course between these wells is travelled during the chemical transformation. ΔG^\ddagger is the activation energy, defined as the difference in free energy between the reactants/ES and the transition state. G is the free energy. ΔG° is the difference of energy level between reactants/ES and product/EP. When product resides in an energy well lower than reactants (negative ΔG°) the reaction is favourable (Lowry & Richardson, 1987; Nelson *et al.*, 2008).

The rate of the reaction performed will depend on the activation energy (ΔG^\ddagger) and the substrate concentration, as shown by the following sequence of formulas.

The activation energy is expressed by:

$$\Delta G^\ddagger = -RT \ln K$$

where R is the gas constant, T is the temperature and K is the equilibrium constant. The value of K will depend on the concentrations of each component in the reaction:



Thus, the rate for the reaction will be:

$$\text{Rate} = k_r [A][B], \text{ where } k_r = e^{-\Delta G^\ddagger/RT} (\kappa kT/h)$$

However, due to the exponential, the main factor that determines the reaction rate will

be ΔG^\ddagger , implying that the proficiency of an enzyme is related to its ability to recognise and stabilise the transition state.

1.3.2 Enzymes lower the activation energy through binding energy

As suggested by Linus Pauling, enzymes preferentially stabilise the transition state relative to the ground state, through the formation of non-covalent interactions. Some weak interactions are formed in the ES complex, but the full complement of such interactions between substrate and enzyme is formed only when the substrate reaches the transition state. The formation of each interaction produces energy. The total energy produced is called the binding energy (ΔG_B), and it is the major source of free energy used by enzymes to lower the activation energies of reactions (Jencks, 1987; Nelson *et al.*, 2008; Pauling, 1948; Wolfenden, 1969; Wolfenden & Snider, 2001). The binding energy contributes to enzyme specificity, because the amount of weak interactions formed will depend on the ability of the enzyme to discriminate between a substrate and a competing molecule.

Enzymes affect in different ways the physical and thermodynamic factor of ΔG^\ddagger :

- the binding of substrate to enzyme decreases the entropy of the reaction, by lowering the relative motion and promoting the correct orientation.
- The substrate-enzyme interactions cause substrate desolvation, by replacing most of the hydrogen bonds usually present between substrate and water; these weak bonds help to compensate thermodynamically for any distortion that the substrate could undergo, such as electron re-distribution.
- Binding of the substrate to enzyme causes distortion in both, which promotes the substrate TS structure and a new enzyme conformation. In the new enzyme conformation, catalytic properties are enhanced because the catalytic functional groups are properly aligned.

Some of these mechanisms are shown in figure 1.4 (Alberts, 2008; Jencks, 1972; Johnson, 2008; Laidler & King, 1983; Nelson *et al.*, 2008).

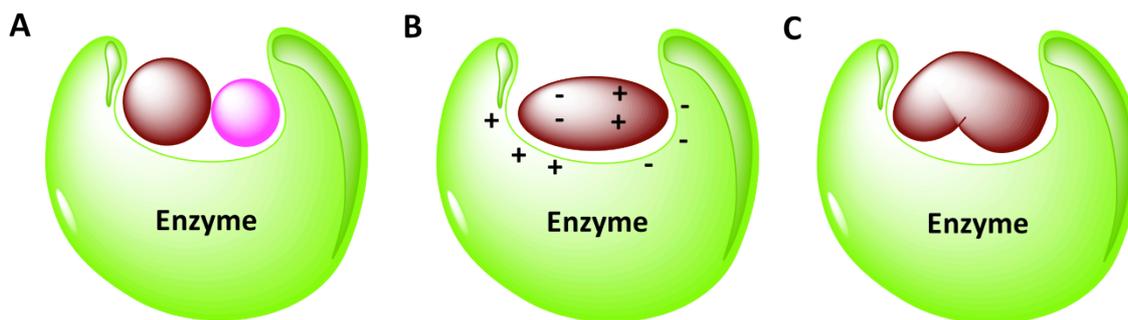


Figure 1.4 Strategies adopted by enzymes for lowering the energy barrier. Several strategies can be assumed by an enzyme for promoting catalysis: A) enzyme binds the substrates molecules and orients them precisely; B) binding of substrate to enzyme rearranges electrons in the substrate, creating partial negative and positive charge; C) the binding of the enzyme changes substrate shapes by bending bonds (modified from Alberts, 2008).

Thus, characterising the geometry, bonding, and charge distribution of the enzymatic transition state is crucial for understanding catalysis. A high resolution X-ray structure of the Michaelis complex between an enzyme and its substrate cannot provide a full description of the mechanism by which an enzyme achieves enormous catalytic rate acceleration. This determination requires, minimally, knowledge of the structure of the transition state for the enzyme-catalysed reaction. This information, combined with the Michaelis enzyme-substrate complex, which describes the substrate recognition in the ES complex, will provide a better-defined picture of catalysis (Pauling L., 1946; Amyes TL and Richard P, 2007).

In this thesis, a combination of solution NMR and X-ray crystallographic techniques has been used to investigate the requirements of specific phosphoryl transfer enzymes for recognising substrates and then overcoming the energy barrier. Before describing the questions addressed, the *state of the art* of how phosphoryl transfer enzymes achieve catalysis will be discussed.

1.4 How do phosphoryl transfer enzymes achieve catalysis?

The understanding of how enzymes catalyse phosphoryl transfer reactions, and in particular, what are the transition state features, has been studied by the comparison with what occurs in an un-catalysed reaction. Thus, the comprehension of enzymatic

phosphoryl transfer, first requires the knowledge of transition state features in un-catalysed phosphoryl transfer (Lassila *et al.*, 2011).

Three main questions are raised:

- Are un-catalysed phosphoryl-transfer reactions concerted, or do they proceed through stable intermediates?
- Do enzymes alter transition states from those in solution?
- How do phosphoryl-transfer enzymes achieve catalysis?

The following paragraphs will go through these questions. As mentioned above, the enzymes studied in this thesis are kinases, mutases or small GTP proteins. Since these three systems involve the phosphomonoester hydrolysis reaction, reactions involving nucleic acids or diesters, will not be discussed.

1.4.1 Mechanism possibilities: Associative-Concerted-Dissociative phosphoryl transfer transition states

One of the most confusing issues is whether phosphoryl-transfer reactions are concerted or whether they proceed through stepwise processes with discrete intermediates. This question has largely been resolved for un-catalysed reactions.

The transition states for phosphoryl transfer reactions are typically assigned to a position along a continuum between associative and dissociative extremes (Admiraal & Herschlag, 1995). The associative and dissociative nature of a phosphoryl transfer reaction is defined by the extent of bond formation between the incoming nucleophile and phosphorus, and the extent of bond cleavage between phosphorus and the leaving group in the TS. The associative transition state has a larger amount of bond formation to the incoming nucleophile, a small amount of bond cleavage to the outgoing leaving group and charge accumulation on the non-bridging phosphoryl oxygen atoms. In contrast, the dissociative transition state has a small amount of bond formation to the incoming nucleophile, a large amount of bond cleavage to the outgoing leaving group, and charge donation from the non-bridging phosphoryl oxygen atoms to phosphorus.

Thus, dissociative and associative transition states are very different. In the case of a

dissociative TS, a stepwise elimination-addition mechanism will lead to the formation of a metaphosphate intermediate; in the associative TS a stepwise addition-elimination mechanism will produce a pentavalent phosphorane intermediate. A third scenario has been proposed for concerted mechanism, an S_N2 -type reaction proceeding through a single transition state with simultaneous breaking of one bond and formation of a new bond (Lassila *et al.*, 2011) (Figure 1.5).

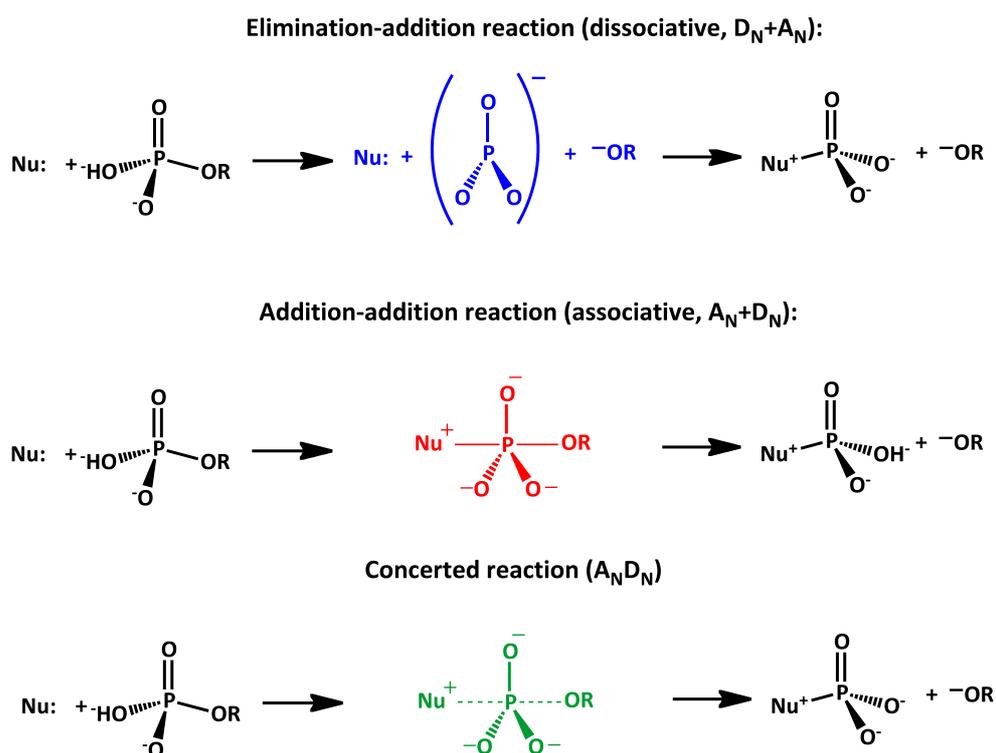


Figure 1.5 Hypothetical reaction mechanisms for phosphoryl transfer. The three hypothetical reaction mechanisms described for phosphoryl transfer are shown. The IUPAC nomenclature is used: $D_N + A_N$, $A_N + D_N$, A_ND_N (modified from Lassila *et al.*, 2011).

All the possibilities are represented by two-dimensional reaction coordinate diagrams (Figure 1.6), also known as More-O'Ferrall-Jencks diagrams (More O'Ferrall, 1970). In this diagram, the two extremes are represented in the upper-left corner (a phosphorane intermediate) and in the lower-right corner (a metaphosphate intermediate). Around the perfect synchronous transition state a number of possibilities arise from how much the synchronisation is concerted. Thus the concerted transition state could be defined as

tight or **loose**, if it is more similar to the associative or dissociative intermediate, respectively. Transition states in which the phosphorus atom sees an increase in the total bond order to the nucleophile and the leaving group relative to that in the reactants are referred to as tight or associative-like transition states. Those with a decrease in total bond order between phosphorus and the nucleophile and leaving group are called loose or dissociative-like transition states.

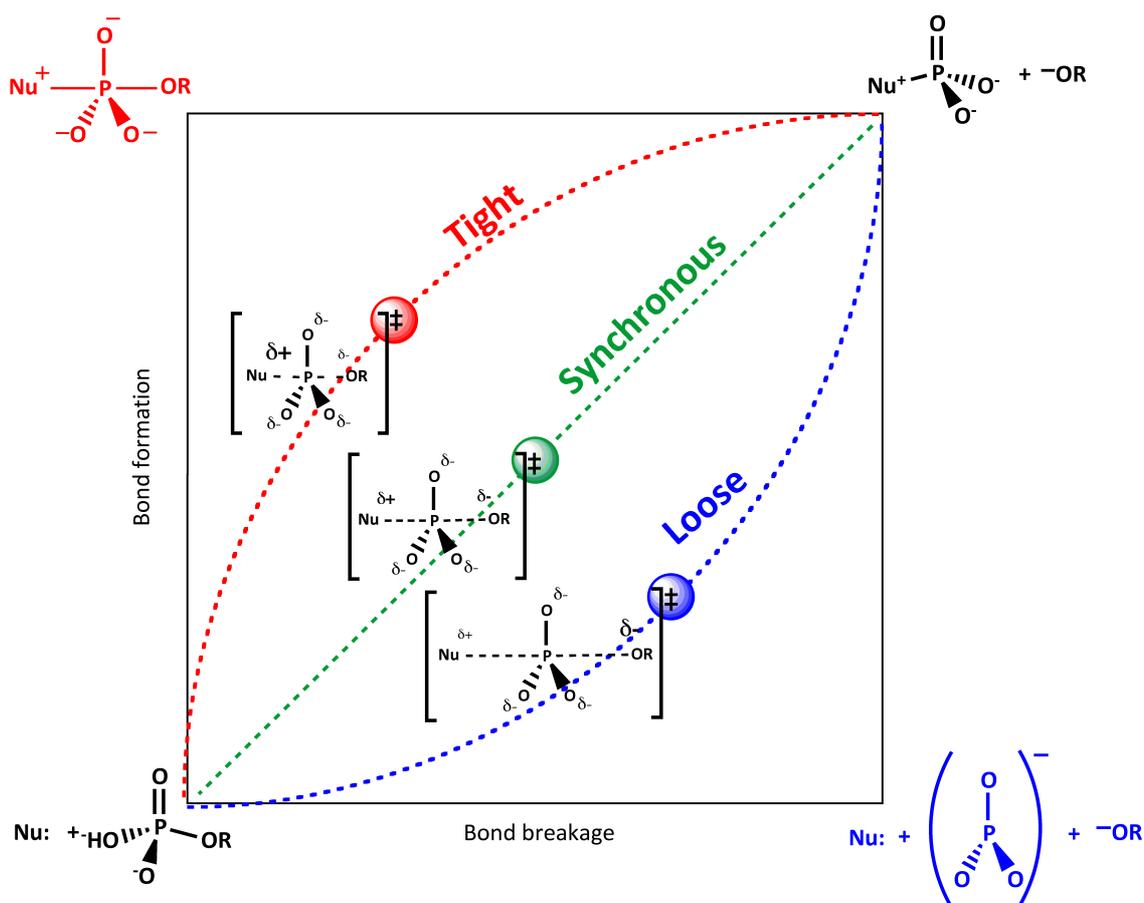


Figure 1.6 The More-O'Ferrall-Jencks diagram for phosphoryl transfer transition states. In the two-dimensional reaction coordinate diagram shown the reactants are depicted in the lower left corner and the products in the upper right. Bond breaking proceeds along the x-axis and bond formation proceeds along the y-axis. This two-dimensional reaction coordinate defines a three-dimensional free energy surface in which the free energy axis is perpendicular to the page. Starting from the reactants located in a free energy well at the bottom left, a reaction will proceed across this surface via the pathway with the lowest barrier. The range of possible concerted transition states are shown (Tight, Synchronous and Loose). The charge distribution for each of these three TS is shown (modified from Lassila *et al.*, 2011).

1.4.2 Un-catalysed phosphoryl transfer reactions adopt a loose transition state

Both phosphomonoesters and phosphoanhydrides undergo phosphoryl transfer by P-O and P-N cleavage. Phosphate monoester hydrolysis or phosphate monoester dianion un-catalysed hydrolysis reactions adopt a loose transition state, without the formation of a metaphosphate intermediate. In the two dimensional reaction diagram (Figure 1.6), the reaction will proceed along the lower right corner where a metaphosphate-like transition state will occur. Several studies are present in the literature that support this thesis. The experimental data were obtained through LFERs and KIEs studies, which are the typical tools used for defining the nature of a transition state.

LFER or linear free energy relationship correlates the pK_a values and a series of nucleophiles or leaving groups with a specific $\log K$. pK_a is proportional to a standard free energy change and it refers to the nucleophile or leaving group considered; $\log K$ is a linear function of the free energy of activation, where K is the rate constant for the reaction. The slope of the LFER, called the Brønsted value (or β value), correlates with the amount of bond formation or bond breaking between the phosphorus and the nucleophile ($\beta_{\text{nucleophile}}$) and between the phosphorus and the leaving group ($\beta_{\text{leaving group}}$) (Lowry & Richardson, 1987; Jencks 1987; Williams A, 1992). A large $\beta_{\text{nucleophile}}$, which is suggestive of a substantial transition-state bond formation between the nucleophile and the phosphorus, together with a less negative $\beta_{\text{leaving group}}$, which is suggestive of small transition-state bond cleavage between phosphorus and the leaving group, identify a phosphoryl transfer reaction is more associative-like, thus a tight TS. The opposite trend (a small $\beta_{\text{nucleophile}}$ and a large-negative $\beta_{\text{leaving group}}$) denotes a more dissociative-like transition state or loose TS (Admiraal & Herschlag, 1995; Lassila *et al.*, 2011).

Brønsted values have been used for mapping the un-catalysed transition state of ATP, GTP and pyrophosphate dianion hydrolysis in solution. A series of primary alcohols of varying pK_a were used to investigate nucleophilic participation in this transition state. Alcohols are chemical homologs of biological nucleophiles such as sugars, water and the serine and threonine residues. For all three samples, small $\beta_{\text{nucleophile}}$ values were obtained (0.07 ± 0.08 for ATP; 0.05 ± 0.08 for GTP and 0.06 ± 0.06 for pyrophosphate dianion), indicating that a loose transition state is formed in all three reactions. The $\beta_{\text{leaving group}}$ was calculated too. The negative value obtained (-1.1 ± 0.2) indicated a large amount of negative charge on the leaving group, thus bond breaking. This outcome is a

proof of the presence of a more dissociative-like TS in solution (Benkovic and Schray 1978; Thatcher and Kluger, 1989; Admiraal & Herschlag, 1995). Similar results have been obtained in the study of other un-catalysed reactions on phosphomonoesters or dianion (Bourne & Williams, 1983; Bourne & Williams, 1984; Grzyska *et al.*, 2003; Herschlag & Jencks, 1987; Kirby & Jencks, 1965; Kirby & Varvoglis, 1967; Skoog & Jencks, 1983; Skoog & Jencks, 1984).

KIEs, or kinetic isotope effects, provide information on changes in bonding that occur in the transition state by measuring changes in reaction rates that occur when an atom in the compound of interest is substituted with a heavy isotope. The substitution can be made at the position of bond cleavage or formation (primary isotope effect) or at other positions that are not directly involved (secondary isotope effect). The substitutions increase the so called zero-point energy of each bond, that corresponds to its minimum vibrational energy state. When a bond is broken, the vibrational energy states for that bond are lost and there is no longer a difference in vibrational energy between the heavy and light isotopes. For the primary isotope effect the dominant vibration lost is the stretching mode, while for the secondary, significant effects are on the bending and torsional modes (change in geometry). Since the heavy-atom cleavage requires more energy, a difference in free energy can be observed at the ground state and the reaction is slower. This leads to a “normal” KIE ($k_{\text{light}}/k_{\text{heavy}} > 1$) (Hengge, 2002; Melander & Saunders, 1980). For bonds that are partially broken in the transition state, the KIE corresponds to the extent of bond cleavage in the transition state. In cases where bonding increases at the substituted position, for example as a result of protonation, an inverse isotope effect can be observed ($k_{\text{light}}/k_{\text{heavy}} < 1$). For monoester TS, the KIE is larger than for the leaving group, while it is under one for the non-bridging oxygen atom (Buchwald *et al.*, 1984; Buchwald *et al.*, 1982; Hengge *et al.*, 1994). Thus, the KIE data indicate bond cleavage between the phosphorus and the leaving group and a small bond formation between the nucleophile and the phosphorus. This is in agreement with LFERs studies, which described that reactions involving monoesters or phosphodiesters are characterised by a loose or dissociative-like transition state.

Eventually, in water, the bond to the incoming nucleophile starts to form before the bond to the leaving group has completely broken, avoiding the formation of a long-lived, freely diffusing metaphosphate intermediate. Stereochemical studies demonstrate

that hydrolysis proceeds with complete inversion of configuration (Buchwald *et al.*, 1984; Buchwald *et al.*, 1982), without any racemic product produced by the attacking of a water molecule to the metaphosphate intermediate.

1.4.3 Enzyme challenges in accommodating the un-catalysed loose transition state

As described above, phosphoryl transfer enzymes are involved in diverse biological processes. Remarkably, the reaction performed is the same (phosphoryl transfer), implying that the chemical basis should be similar among different enzymatic systems.

A hallmark of the active sites of phosphoryl-transfer enzymes is the presence of positively charged groups, in the form either of divalent metal ions and/or positively charged amino acid side chains (Cleland & Hengge, 2006; Knowles, 1980). These elements are used by the enzyme along the reaction for transferring protons, as described in **General Acid Base Catalysis (GABC)**. Many biochemical reactions, such as phosphoryl transfer, involve the formation of unstable charged intermediates that tend to break down more rapidly to their constituent reactant species instead of going to the product. Enzymes stabilise these intermediates by the transfer of protons to or from the substrate or intermediate to form a species that breaks down more readily to products. In GABC the proton transfer is mediated by residues and water molecules trapped in the active site. Water molecules and weak organic acids can supplement water as proton donors, while weak organic bases can serve as proton acceptors. Some phosphoryl-transfer enzymes use **metal-catalysis** in addition to GABC. A metal ion may serve as an electrophilic catalyst, stabilising a negative charge on a reaction intermediate; or it can generate a nucleophile by increasing the acidity of a nearby molecule. Finally, the metal ion may bind to substrate, increasing the number of interactions with the enzyme and thus the binding energy (Berg & Tymoczko, 2002; Nelson *et al.*, 2008).

These considerations raise the question of whether the enzymatic environment could accommodate the loose transition state characterised in solution.

The dissociative transition-state charge estimates for the un-catalysed hydrolysis of ATP in figure 1.7, show the challenges needed to be overcome by the enzyme for carrying out the reaction (Admiraal & Herschlag, 1995).

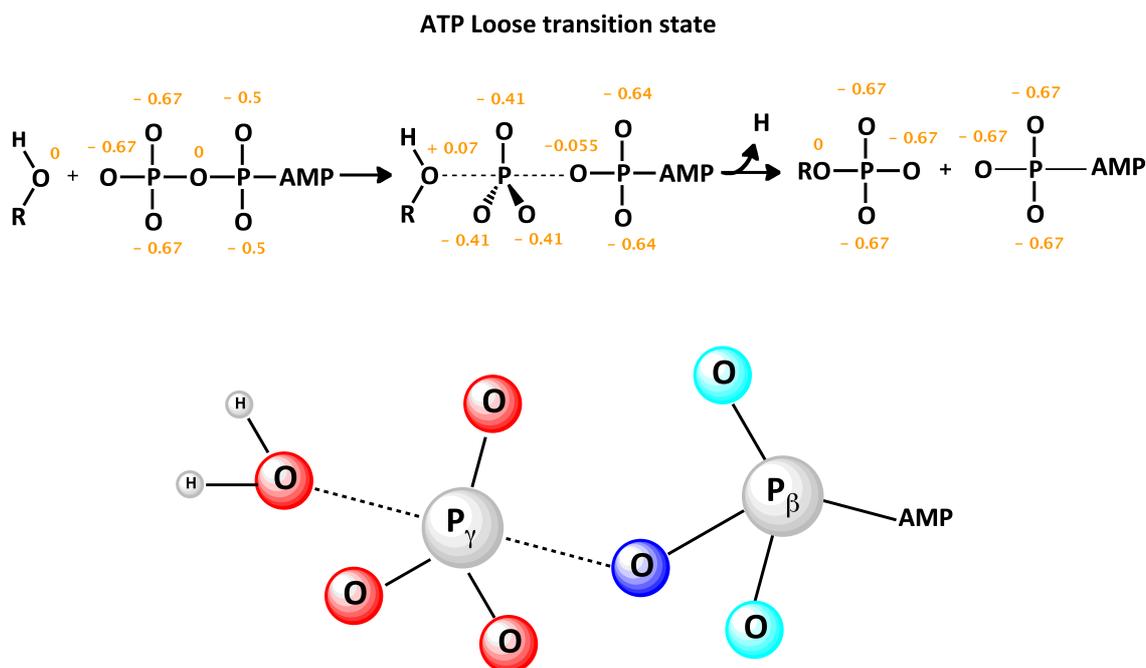


Figure 1.7 Transition state charge estimates for the un-catalysed hydrolysis of ATP. Charges on oxygen atoms of the reactants, transition state and products for ATP hydrolysis estimated from LFER. The schematic representation shows the change in charge in going from the ground state to the transition state, which has a dipolar character. The transition-state geometry is depicted, with the phosphoryl group undergoing transfer separated from the nucleophilic water and the ADP leaving group by dashed lines. In *blue-cyan* is shown the increase in negative charge. The nonbridging γ -phosphoryl oxygen shows a decreasing charge (in *red*), while the β - γ bridging oxygen atoms undergo a large negative charge increase (in *blue*). The β -nonbridging oxygen (in *cyan*) atom increases in negative charge too, even though it is a small change (modified from Admiraal & Herschlag, 1995).

As mentioned above, one of the effects of the binding energy is the alignment of enzymatic catalytic functional groups in the enzymes. This should result in: the fixing of the activated nucleophile with the γ -phosphate and the stabilisation of the transition state charge distribution.

The positioning of the nucleophile for facilitating catalysis is one of the most important features for enzymes. The nucleophile must be aligned with the phosphorus atom and the leaving group for in-line attack at phosphorus (Admiraal & Herschlag, 1995). Moreover, the nucleophile is required to be activated, to act as a proton donor in GABC. These two aspects are relevant in both loose and tight TS.

The change in charge from the ground state to the transition state has a dipolar character that leads to negative charge accumulation on the leaving group: in a loose TS the non

bridging γ -phosphoryl oxygen atom has a small negative charge, the β - γ bridging oxygen atoms undergo a negative charge increase. Moreover, the phosphorus group goes through substantial changes in geometry from tetrahedral coordination to a trigonal bipyramidal.

1.4.4 Enzymatic TS are studied through metallofluoride complexes

The first investigations on how loose transition states could be, or could not be, accommodated in a enzymatic environment, were performed through LFERs and KIEs calculations. However, these techniques have two limitations. First, they are applied to a much more complex system with regard to un-catalysed phosphorylation (Lassila *et al.*, 2011). Second, and more relevant, the data obtained do not provide a direct observation of the transition state. The transition state is a high-energy and short-lived species. This implies that no physical or spectroscopic method can be used to observe it directly. However, in the case of enzymes performing phosphoryl transfer, an alternative way has been found for trapping the transition state. Inorganic metallofluoride molecules, such as aluminium and magnesium fluoride, are used for investigating structural features of the phosphoryl transfer reaction. Their ability to inhibit the phosphoryl transfer reaction is used in X-ray crystallography and ^{19}F -NMR for trapping the protein structure in the catalytically active state.

The most known and used metallofluoride molecule is aluminium fluoride. AlF_x is a small inorganic molecule that mimics the chemical structure of a phosphate (Bigay *et al.*, 1987). Aluminium fluoride entered the scene with the accidental finding by Sutherland and co-workers in 1958 that adenylate cyclase is activated by fluorides (Rall & Sutherland, 1958). Twenty years later, Gilman and co-workers found that the target of the activation was a heterotrimeric G protein and that the active stimulatory agent was aluminium fluoride, which is present as a leached-out impurity in millimolar solutions of fluoride in glass (Sternweis & Gilman, 1982). Finally, in 1994, the three-dimensional structure of heterotrimeric G proteins was solved by X-ray crystallography, bound to GDP and aluminium fluoride (Coleman *et al.*, 1994; Schindelin *et al.*, 1997; Sondek *et al.*, 1994). Aluminium tetrafluoride (AlF_4^-) was found located in the γ -phosphate binding site of these proteins. Aluminium was found coordinated in

octahedral geometry with fluoride ligands in a square-planar coordination and two oxygen ligands at the apical positions. One oxygen ligand is a β -phosphate oxygen atom, the leaving group in the transfer reaction, whereas the other is the oxygen atom from the attacking water molecule. Thus, surprisingly, it was mimicking the transition state of the phosphoryl transfer reaction rather than the ground state (Figure 1.8). AlF_4^- is square planar and carries a net negative charge. Thus, it is an isoelectronic, but not isosteric mimic of the phosphoryl group in transfer.

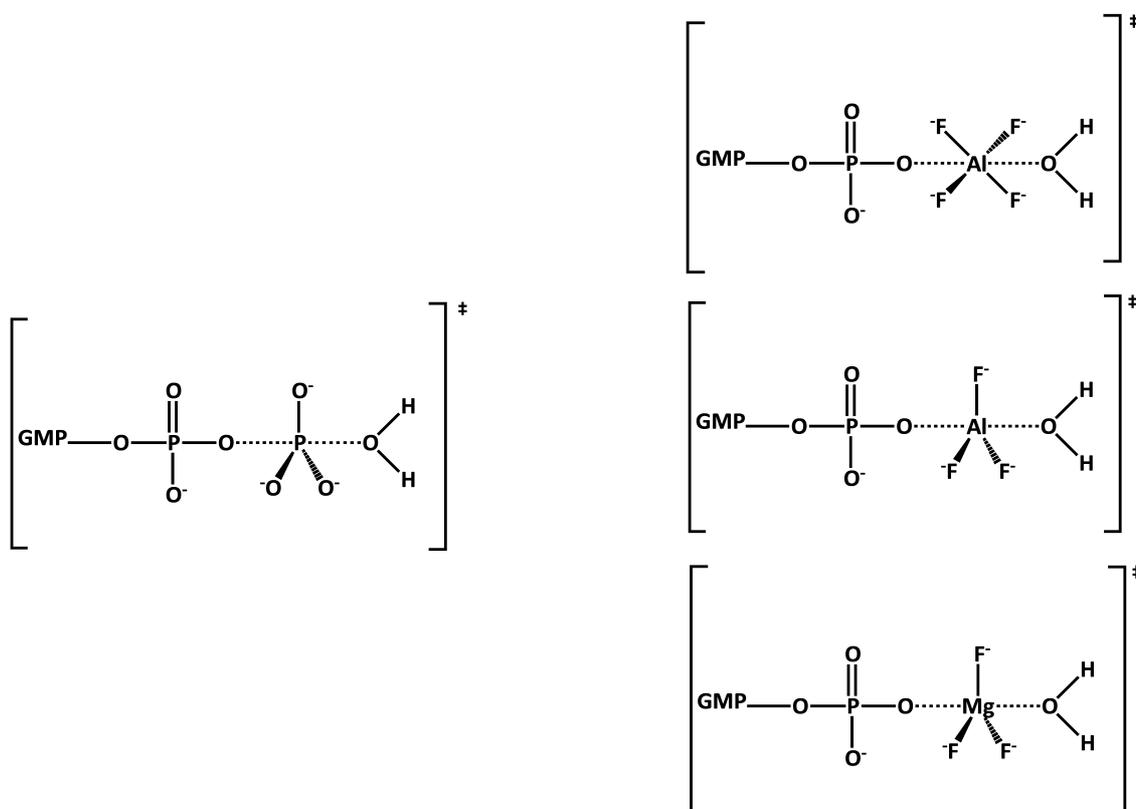


Figure 1.8 Transition state analogues of GTP-hydrolysis. The transition state analogues are compared with the phosphorus in flight. PO_3^- has a TBP geometry and is negative charged. AlF_4^- is an isoelectronic mimic of PO_3^- , AlF_3 is an isosteric mimic of PO_3^- , MgF_3^- is an isoelectronic and isosteric mimic of PO_3^- .

Aluminium fluoride was also observed as an aluminium trifluoride (AlF_3) species in other structures of small G proteins (Coleman *et al.*, 1994). Aluminium trifluoride is an isosteric but not isoelectronic mimic of the phosphoryl group (Figure 1.8). It adopts a trigonal planar arrangement at the active site but carries no net charge. The difference in

species was considered to be related to pH conditions. Crystallisation below pH 7.0 seems to favour AlF_4^- while higher pH results in binding of AlF_3 . Crystallisation of UMP/CMP kinase in complex with AlF_x at two different pH (4.5 and 8.5) showed an octahedral aluminium coordinated species at the acidic pH, while a TBP aluminium as observed at pH 8.5 (Schlichting & Reinstein, 1999).

In 1998, magnesium fluoride entered on the scene. In a paper by Gilman and co-workers, a comparison was made between the RhoA-GDP- AlF_4^- -RhoGAP-TSA structure and RhoA-GDP- MgF_3^- -RhoGAP-TSA fluoride structure (Graham *et al.*, 2002) (Figure 1.9).

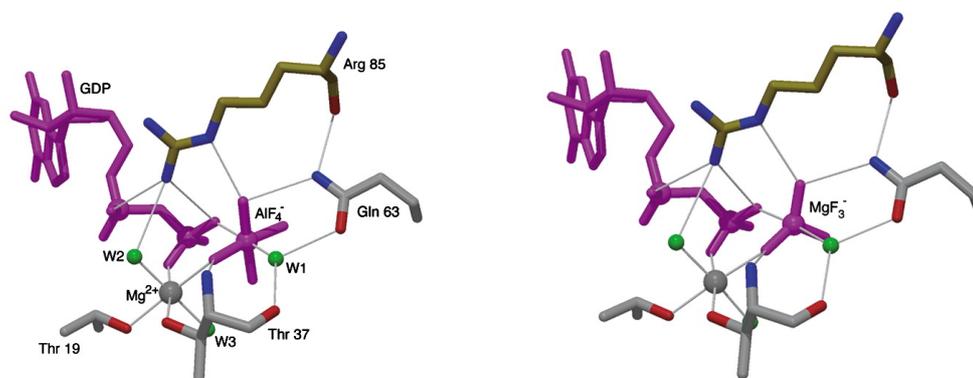


Figure 1.9 Aluminium and magnesium fluoride in RhoA-RhoGAP transition state structures. The left panel shows, in ball-and-stick representation, molecular details of the active site of RhoA-RhoGAP complex with AlF_4^- (PDB code 1TX4; Rittinger *et al.*, 1997), while the right panel shows similar view for the complex with MgF_3^- (PDB code 1WO3; Graham *et al.*, 2002). Arg85 from RhoGAP is colored *green* while the residues from RhoA are shown in *grey*. The GDP and metallofluoride moiety are shown in *magenta* while water molecules are in *green* (Graham *et al.*, 2002).

The structures showed that magnesium fluoride was coordinated in a trigonal bipyramidal geometry, similar both in geometry and charge to the one of the phosphorus group in transfer (Figure 1.8). Using proton-induced X-ray emission (PIXE) spectroscopy it was demonstrated that the magnesium fluoride species, rather than AlF_3 , was the correct one assigned to the electron density map. Moreover, in the crystallisation conditions, deferoxamine was added, which is an aluminium chelator. The comparison of the two structures showed that both metallofluorides are coordinated

in a similar way, pointing to the possibility to use magnesium as the reagent of choice for transition state studies. Indeed, they described for the first time the possibility to trap the TS using an isosteric and isoelectronic mimic of a phosphorus group.

The use of metallofluorides in structural biology has led to the understanding of the features of enzymatic transition state and, consequently, how catalysis is achieved in phosphoryl transfer enzymes. While X-ray crystallography describes the structural and geometrical features, fluoride NMR provides information on the electronic and protonic interactions of TSA formation in solution to the Van der Waals environment as well as local electrostatic fields deduced by the fluorine chemical shifts (Higashijima *et al.*, 1991; Hoffman *et al.*, 1998; Graham *et al.*, 1999; Praefcke *et al.*, 1999). The combination of the two techniques is a powerful tool for distinguishing minor differences and obtaining complementary information. For example, it is striking that the structural differences between the phosphorane intermediate and a slightly dissociative transition state are only about 0.5 Å between the entering and leaving groups within a similar TBP geometry (Williams, 2004) (Figure 1.10). Due to the high resolution of crystallographic data combined with fluorine solution NMR the problem can be overcome, and the TS can be defined as more dissociative-loose one, or more associative-tight one.

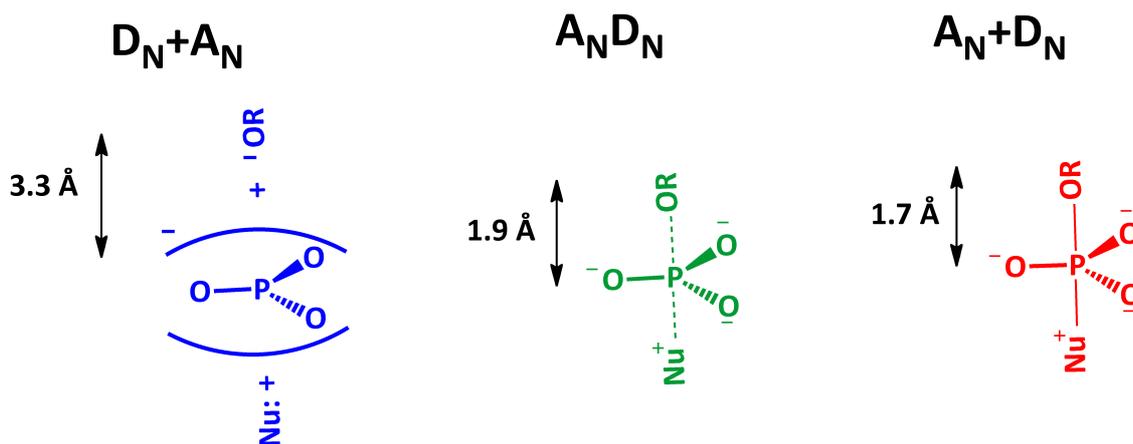


Figure 1.10 Approximate distances for differing degrees of bonding to nucleophile/leaving group in transition states for phosphoryl transfer. The left-hand structure represents a fully dissociative mechanism, with zero bond order to nucleophile and leaving group; the right-hand structure a fully associative mechanism with a phosphorane intermediate. As the character of the central concerted transition state moves towards either side, its character can be described as dissociative or associative (modified from Williams, 2004).

Furthermore, metallofluoride complexes provide insight into the enzyme catalysis, by allowing the definition of:

- how the nucleophile is activated;
- how the developing negative charge on the leaving group is stabilised;
- how charge and geometry of the phosphorus group in flight are accommodated in the active site.

1.4.5 Activation of the nucleophile

Nucleophile activation can be achieved in three different ways: positioning, increasing nucleophilicity and overcoming electrostatic repulsion.

1.4.5.1 Positioning

Positioning of the nucleophile, for facilitating catalysis, is one of the important features for enzymes. It is important for both associative and dissociative transition states. However, it is doubtful how significant this would be in a dissociative transition state,

because the transition state in the un-catalysed reaction for phosphate monoester hydrolysis has little nucleophile participation. The possible explanation is that even in a loose transition state there is some bond formation to the nucleophile, which requires positioning of the nucleophile with respect to the phosphoryl group (Lassila *et al.*, 2011).

Both residues and metal ions can participate in nucleophile positioning. In the small GTP proteins of the Ras superfamily, a conserved glutamine (Gln61 in Ras, Gln 63 in RhoA and Gln61 in CDC42 (Nassar *et al.*, 1998; Rittinger *et al.*, 1997; Scheffzek *et al.*, 1997) is positioning the attacking water molecule in line with the γ P of GTP. In kinases, the catalytic magnesium assists in the orientation of the γ -phosphate for “in line” with respect to the second substrate, creating the correct geometry to complete phosphoryl transfer (Lassila *et al.*, 2011).

1.4.5.2 Increasing the nucleophilicity

Increasing the nucleophilicity can be achieved by a number of mechanisms, including removal of a proton by a general base, activation by a metal ion, or by changing the identity of the nucleophile (Lassila *et al.*, 2011). Specific residues can work as general base: an aspartate is general base in *yeast* hexokinase, in glycerokinase and galactokinase of *Pirococcus furiosus* (Aleshin *et al.*, 2000; Anderson *et al.*, 1978; Golicnik *et al.*, 2009; Tari *et al.*, 1997; Tari *et al.*, 1996); in human creatine kinase a glutamate is the general base (Ostanin & Van Etten, 1993). Activation by metal ion occurs in *E. coli* alkaline phosphatase, where magnesium has been suggested to be the general base (Cleland & Hengge, 2006) that deprotonates the Ser102. Further, it was described that in phosphoenolpyruvate carboxykinase, magnesium coordination orients the β - and γ -phosphates in a high energy eclipsed conformation resulting in increased electrostatic repulsion between the phosphoryl groups, which may activate ATP for catalysis (Stec *et al.*, 2000).

Interestingly, in the un-catalysed reaction, metals such as magnesium or calcium do not show any significant improvement in reaction rate or change in the loose TS (Smith & Rayment, 1996). However, from the examples described, the relevance of metal ions in some phosphoryl transfer enzymes is unquestionable.

1.4.5.3 Overcome the electrostatic repulsion

A third way of activating the nucleophile is to overcome the electrostatic repulsion between the nucleophile and the negatively charged phosphoryl group. However, this has only small effects on un-catalysed reaction rates. For example, anionic nucleophiles react ~5-fold faster, whereas reactions of neutral nucleophiles are not significantly affected (Herschlag & Jencks, 1987; Lahiri *et al.*, 2002). The mechanism could be relevant in reactions that involve highly charged phosphate esters, such as nucleoside triphosphates. Malarial adenylate kinase needs to overcome the charge repulsion barrier in order to equilibrate ATP + AMP with 2 ADPs (Figure 1.11). It achieves this important reaction with a turnover of $k_{\text{cat}} = 35 \text{ s}^{-1}$ (Cassano *et al.*, 2002).

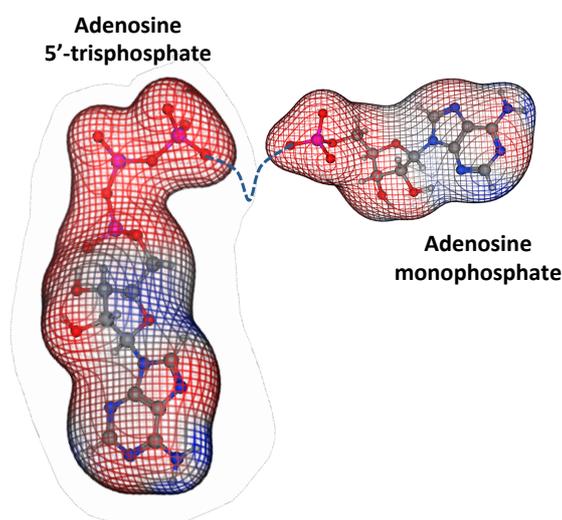


Figure 1.11 Electrostatic surfaces for the ATP tetra-anion (left) and the AMP dianion (right). The adenylate kinase reaction transfers the γ P from ATP to AMP by in-line attack resulting in cleavage of γ -P to β -O. Negative polarity is in red; positive polarity is in blue (modified from Bowler *et al.*, 2010a).

1.4.6 Stabilisation of negative charge development on the leaving group

As described in ATP hydrolysis there is a substantial negative charge that develops in the loose transition state relative to the ground state for phosphate monoester dianion reactions. Stabilisation of developing negative charge on the leaving group and on the other charge changes can contribute significantly to catalysis of monoester reactions.

In modelling studies for serine/threonine phosphate (PP1) it has been shown that if Mg coordinates the α and β phosphates instead of the β and γ phosphates of ATP; this results in transition state stabilisation and reaction acceleration of 3000 fold (Williams 2000).

A remarkable example of an amino acid that stabilises the charge development on the leaving group is the Arginine finger in the Ras-RasGAP system. An example is the RhoA-RhoGAP system, one of the subjects of this thesis. The GAP proteins enhance GTP hydrolysis by presenting an arginine, called the “Arginine finger”. The guanidinium group of this residue interacts with negative charged phosphate nucleotide region, decreasing the energy barrier for hydrolysis (Kim & Wyckoff, 1991; O'Brien & Herschlag, 2002; Reid & Wilson, 1971).

1.4.7 Stabilisation of the charge and geometry of the phosphorus group in flight: the charge balance hypothesis

The change in charge and geometry of the phosphorus group in flight needs to be accommodated for efficient catalysis. However, due to the difference in charge and geometry of the metallofluoride analogues used, it is challenging to clarify how this is occurring.

Studies of the interactions between enzyme active sites and transition state analogues have highlighted the important consideration that charge balance for phosphoryl transfer enzymes rather than the native geometry is essential for TS stabilisation during catalysis. The proposal of a fundamental role of charge balance raises an important question about the possible transition state species that could be accommodated inside the active site. While AlF_4^- and MgF_3^- carry the same charge of a scissile phosphate, AlF_3 only conserves a similar geometry to the phosphorus group in transfer. Interestingly the pH relationship between the two aluminium fluoride species is similar to the one between aluminium tetrafluoride and magnesium trifluoride. Most of the structures that contain AlF_3 , were crystallised in the presence of magnesium, pointing out the doubt that the electron density with TBP moiety were not assigned correctly. Moreover, no clear explanation has been provided for the switching between the two forms of aluminium fluoride, leaving unexplained how an enzyme such as

phosphoserine phosphatase could coordinate both species at the same time (Schlichting 1999). Fluorine NMR pH titration experiments have successfully shown that with increasing pH AlF_4^- is progressively displaced by MgF_3^- (Baxter *et al.*, 2008; Jin, 2012). Indeed, the increased pH causes the concentration of aluminium fluoride to fall below that of the enzyme. This is due to the displacement of fluoride ions by hydroxide ion binding and the consecutive change in geometry in solution of the aluminium fluoride species (Mesak & Dahl, 2000; Tremblay *et al.*, 2005; Webster, 2004). Thus, it was proposed that the charge on the metal fluoride species dominates its geometry in determining the preferred mode of binding. This hypothesis was termed the charge balance hypothesis.

One of the main consequences of this hypothesis is that the enzyme's activity ensures that the number of positively and negatively charged groups are exactly balanced within the transition state (Baxter, 2008). Thus, any charge perturbation of the TSA should be compensated in some way from the enzyme active site and *vice versa*, without any change in the final net-charge. A first proof of this proposal has been tested on human phosphoglycerate kinase (HsPGK). HsPGK catalyses the transfer of phosphate from 1,3-bisphosphoglycerate to ADP in the first energy generating step of ATP hydrolysis (Opperdoes, 1987). The protein has been co-crystallised with an octahedral tetrafluoroaluminate TSA in the active site, in agreement with the ^{19}F NMR data (Cliff *et al.*, 2010). The structure showed that Lys219 coordinates the aluminium fluoride in the transition state. In order to establish the role of charge balance, this positive residue was mutated to an alanine. The NMR data showed a change from aluminium tetrafluoride to trifluoride with the mutation. As expected, this caused a reduction in charge of the TSA species, producing a trifluoride aluminium observed in the crystal structure. Importantly, aluminium is still coordinated in octahedral geometry, with a fluoride atom replaced by a water molecule (Figure 1.13).

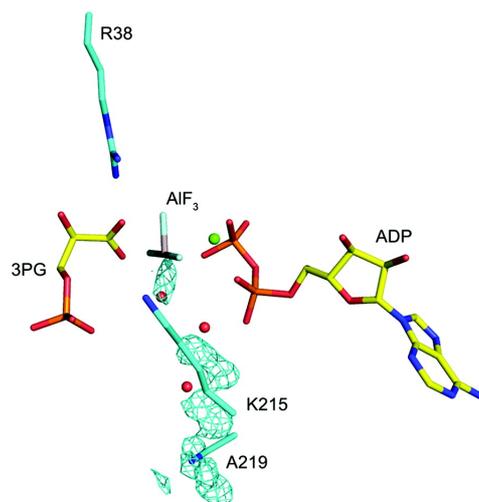


Figure 1.13 Octahedral moiety of AlF_3 in the $\text{HsPGK}_{\text{K219A}}\text{-3PG-AlF}_3\text{-ADP-TSA}$ complex. The $F_o - F_o$ electron density difference Fourier map between the HsPGK-3PG-AlF_4^- -ADP-TSA and the $\text{HsPGK}_{\text{K219A}}\text{-3PG-AlF}_3\text{-ADP-TSA}$ complexes is shown as a *cyan* mesh contoured at 2σ , with the 3PG and ADP ligands and the side chains coordinating the AlF_3 moiety in the $\text{HsPGK}_{\text{K219A}}\text{-3PG-AlF}_3\text{-ADP}$ active site shown as sticks. The water and magnesium are shown as *red* and *green* spheres, respectively (PDB code 2WZD, Cliff *et al.*, 2010).

Thus, no evidence of aluminium trifluoride with a TBP geometry was found, suggesting again that the reported AlF_3 structures in the protein data bank are in fact MgF_3^- structures. Similar results have been obtained in other systems, reported in Dr. Y Jin's thesis and further papers (Jin, 2012; Jin *et al.*, 2012; Xiaoxia *et al.*, 2011).

Expanding the charge balance rule to the entire transition state, the roles of positive residues and metal-ions in stabilising the transition state charge is clearly defined. In transition states complexes where it has been possible to calculate the final net charge of the TS, it was defined as zero net for the first coordination shell of the transferring phosphoryl group and usually for the second shell also. This implies that beyond the interactions with the nucleophile and the leaving group described, further positive residues and metal ions are also involved in the general stabilisation of the transition state. In enzymes coordinating phosphoanhydrides, positive protein charges are involved in neutralising the negative phosphoanhydride charges. One example is Arg166 in alkaline phosphatase described above (Cleland & Hengge, 2006) (Figure 1.12); a second example, is the structure of the metal fluoride transition state complex of UMP/CMP kinase with ADP and CMP (PDB code 3UKD, Schlichting & Reinstein,

1997), where the six negative charges of the nucleotide are exactly neutralised by 8 positive charges, one of which is coming from the catalytic magnesium (Bowler *et al.*, 2010a).

1.4.8 Is the enzymatic transition state dissociative or associative?

After having described how a phosphoryl transfer enzyme, which performs GABC or metal-catalysis accommodates the transition state, the question “*Is the enzymatic transition state dissociative or associative?*” can be addressed.

There is no clear answer to this question. It is still not obvious how an enzymatic active site can stabilise a loose, metaphosphate-like transition state. Structures of kinases show that active site elements interact with the non-bridging oxygen atoms, which are negatively charged in the dissociative transition state. In contrast, positive residues interact with the phosphoryl group in flight, that is expected to be more negatively charged in the associative TS than the dissociative (Maegley *et al.*, 1996; Rittinger *et al.*, 1997). However, the phosphoryl group inverts during phosphoryl transfer, and this geometric change has been shown in computational studies of protein tyrosine phosphatases to result, by means of stronger hydrogen-bonding interactions, in preferential enzymatic stabilisation of a loose transition state (Cleland & Hengge, 2006). Inverse secondary O¹⁸ isotope effects on alkaline phosphatase, nucleoside diphosphate kinase and hexokinase are consistent with a loose transition state (Alhambra *et al.*, 1998; Asthagiri *et al.*, 2002).

Despite this, a pentacovalent phosphorane intermediate has been claimed to be observed in β PGM phosphoryl transfer in X-ray crystal structures (Lahiri *et al.*, 2003). However, it was demonstrated by a combination of fluoride NMR studies and X-ray anomalous dispersion on the enzyme-TSA crystal that the intermediate was a magnesium fluoride, not a reaction intermediate (Baxter *et al.*, 2010) which is anyway consistent with a tight rather than a loose TS. A tight transition state has been suggested for pyruvate kinase (Baxter *et al.*, 2010) and for UMP/CMP kinase (Hassett *et al.*, 1982). In the case of this enzyme, the separation of the two apical oxygen atoms in the transition state structures is 4.2 Å, less than required for a monomeric metaphosphate intermediate (~6.5 Å) and rather towards the associative reaction distance (≥ 3.6 Å), suggesting the formation of a

tight transition state. Similarly, phosphoserine phosphatase (PSP) shows an in line tight mechanism, by a separation of 4.24 Å between the two apical oxygen atoms (Holtz *et al.*, 1999).

Further examples could be listed for both kinds of transition state. The main conclusion is that while a clear dissociative-like transition state is present in un-catalysed phosphor mono ester reactions, the scenario becomes more complex when the reaction is catalysed, ruling out the possibility of a definitive identification between the two TS species.

1.5 Subjects of this thesis

Despite the fact that the enzyme transition state has not unequivocally been defined as loose or tight, the structural studies performed on metallofluoride transition state complexes has improved remarkably the understanding of the biophysical mechanism of enzyme catalysis (Bowler *et al.*, 2010a). Metallofluoride complexes are considered the best model of phosphoryl transfer (Wittinghofer, 1997) and their used in both fluorine NMR and X-ray crystallography allows the study of the catalysis at a molecular level.

In this thesis metallofluoride complexes have been used for investigating enzyme catalysis in three different systems, already described above:

- the bacterial mutase β Phosphoglucomutase from *L. lactis* (β PGM);
- the human small G protein RhoA, with its GTPase activating protein (RhoGAP):
- the human MAPK p38 α , with one of its MAPKK activators (MEK6).

While similar methods were used, different questions were addressed for each subject.

1.5.1 β -Phosphoglucomutase from *L. lactis*

The mutase protein β Phosphoglucomutase (β PGM) catalyses the interconversion of β -D-glucose-1-phosphate (β G1P) and D-glucose-6-phosphate (G6P). Since β PGM is expressed only in certain bacteria and protists, it is a possible pharmaceutical target in infectious diseases. Therefore, our research was performed in the direction of a

complete dissection of the isomerisation reactions and preliminary drug design studies. Two main topics will be discussed in this thesis:

- Fluorine NMR spectra of β PGM-G6P-MgF₃⁻ transition state analogue (β PGM-G6P-MgF₃⁻-TSA) showed that β PGM can coordinate the magnesium fluoride molecule in two different ways: the major and the minor form, where the names are related to the percentage of the protein population corresponding to each group (Baxter *et al.*, 2010). In Chapter 3 it will be shown how, through the successful combination of ¹⁹F-NMR and room temperature crystallography experiments, the structural details of the major and minor form have been revealed.
- The absence in the literature of structural characterisation of a β PGM- β G1P-TSA structure and the necessity to design a possible β PGM inhibitor, led to the design and synthesis of phosphonate and fluorophosphonate analogues of β G1P. In Chapter 4 it will be shown how through the successful combination of ¹⁹F-NMR, binding assays and X-ray crystallography the best candidate in mimicking β G1P was selected. The resulting β PGM-TSA structures provide information on β G1P catalysis and a possible lead compound for combating bacterial infections.

1.5.2 RhoA GTP-hydrolysis

RhoA belongs to the Rho family of small G proteins, which cycle between an ‘on’ and ‘off’ state that bind either GTP or GDP respectively. Rho proteins are involved in signalling networks that regulate actin, cell cycle progression, cell motility, cell behaviour and gene expression. In cancer, the Rho family is up-regulated and genetic mutations are involved in cardiovascular, pulmonary hypertension, neurodegenerative and immune disorders.

RhoGAP (GTPase activating factor proteins) promotes the inactive conformation of RhoA by stimulating its intrinsic GTPase activity. It has been proposed that in small G proteins Rho, Ras, Cdc42 and Rab, GAPs contribute directly to the catalysis, using a conserved arginine residue, termed the “Arginine finger”. Despite this, there are GAPs (RanGAP and RapGAP) that enhance the reaction through residues other than arginine.

Moreover, RhoA and RhoGAP_{R85A} conserve an intrinsic and residual activity respectively. In Chapter 5, through the structural characterisation of transition state of RhoA-GDP and RhoA-GDP-RhoGAP_{R85A}, two unsolved questions will be investigated at the molecular level:

- How does RhoGAP contribute to catalysis in the absence of an arginine finger?
- How is charge balanced for achieving catalysis in RhoA and RhoA-RhoGAP_{R85A}?

Finally, the results obtained show novel insight into GAP mediated RhoA-GTP hydrolysis, by providing complementary information to the structures already known.

1.5.3 p38 α and MEK6

p38 α and MEK6 belong to one of most well studied MAPK cascades, where p38 α is the MAPK (Mitogen activated protein kinase) and MEK6 is its activator (MAPKK). p38 α is one of four MAPK p38 isoforms, which has a pivotal role in stress response and cytokine production. p38 α is under study for treatment of rheumatoid arthritis, pain relief, dementia and stroke. However, the drugs designed so far lack MAPK specificity. Therefore a better comprehension of how each MAPK interacts with upstream and downstream signalling protein is required.

It was proposed that the activation of p38 α by MEK6 proceeds through (Tanoue *et al.*, 2001) two steps: a docking interaction at a locus outside of the active site, and second, a transient enzyme-substrate interaction at the active site. In the second step MEK6 promotes the active p38 α conformation by double-phosphorylation at the conserved MAPK sequence TxY in the phosphorylation lip (P-Lip). In this thesis, structural and biochemical data have been produced in order to elucidate both steps:

- in Chapter 6, the docking interaction has been investigated by solving the structures of apo-p38 α K53R and p38 α K53R in complex with a MEK6 peptide containing the Docking sequence (D-motif). p38 α K53R is an inactive mutant of p38 α , whose apo-structure shows the ordered conformation of the usually highly flexible region such as the phosphorylation lip (P-Lip) and the Glycine-rich

loop. For the first time it will be shown how MEK6 recognises p38 α and what are the conformational changes induced. A model of the preliminary steps in p38 α activation will then be proposed.

- Unsuccessful attempts were performed to trap in the solid state the transition state of MEK6 phosphorylation reactions. The search for the correct approach for forming the TSA complex, led to the surprising discovery that MEK6 can use ADP as phosphate donor. Preliminary biochemical experiments, which suggest this non-canonical phosphorylation, will be shown. The possibility that MEK6 uses ADP as phosphate donor raises the question of whether MEK6 could employ one molecule of ATP for carrying out both phosphorylation reactions. This scenario appears compatible with the processive or quasi-processive phosphorylation mechanisms, where MAPKK-MAPK does not dissociate or dissociates quickly between the two phosphorylation steps. This project is still in progress and the potential of the novel findings need to be evaluated.

Chapter 2 Materials and Methods

2.1 General materials

All chemicals were purchased from Sigma-Aldrich Chemicals (St. Louis, MO, USA), Fluka Chemicals (Gillingham, UK), and Euromedex (Souffelweyersheim, Strasbourg, France) except for the following: Glycerol 99,5% in twice distilled water was purchased from VWR International (Merck House, Poole, Dorset, BH15 1TD), 2-mercaptoethanol from MP biomedical Europe (Illkirch, France), Acrylamide/bis acrylamide 37.5:1 30% (w/v) stock solution was purchased from National Diagnostic (Atlanta, GA, USA), Luria Bertani broth from Athena enzyme system (Baltimore, MD, USA). Deionised water was purified using a Milli-Q academic purification system (Millipore Ltd, Herts, UK) by reverse osmosis and microfiltration to produce ultra-pure water. This water was used to prepare all aqueous solutions.

2.2 Molecular Biology

2.2.1 Materials

Commercial kits from Qiagen were used for routine experiments, such as miniprep preparation or DNA gel extraction (Hombrechtikon, Switzerland). As expression system for plasmid DNA, homemade DH5 α cells were used. DH5 α contains useful mutations, which allow high plasmid expression without the risk of homologous recombination (*recA1* mutation) and possible endonuclease digestion (*endA1* mutation). For the MEK6-p38 α project, mutagenesis and DNA construct transfer from one plasmid to another were performed (Sections 6.3.1.1-6.3.1.2). For these mutagenesis experiments the following PCR machine was used: Mastercycler gradient from Eppendorf (Hamburg, Germany). Oligonucleotide synthesis and sequencing service were performed by Microgen company (London, UK). Commercial cell strains were used: XL1-blue (Stratagene, La Jolla, CA, USA) were used after mutagenesis experiments, while *E. Cloni* DH5 α cells (Lucigen corporation, Middleton, WI, USA) were used for transformation after ligation in the plasmid transformation experiment.

2.2.2 Transformation of competent cells

Transformation was carried out using the heat-shock method. Cells were mixed with the plasmid and submitted to one cycle of cold-hot-cold that allows the passage of DNA through the bacterial cell membrane.

DH5 α cells were regularly prepared each year by the laboratory technicians, in aliquots of 100 μ l using the calcium chloride method. Aliquots of 100 μ l of competent cells were thawed on ice and 1-4 μ l of plasmid was added (up to 200 ng of plasmid DNA). Cells were incubated on ice for 30 min. Then the tubes were placed in a preheated water bath at 42°C for exactly 50 s and immediately put back onto ice for 5 min. LB media (500 μ l) was added to the tubes and the culture was incubated at 37°C at 250 rpm for 1 h to allow the bacteria to recover. Then 50-100 μ l of transformed cells were plated out onto LB agar plates supplemented with 200 μ g/ml of ampicillin. After overnight growth at 37°C, the plate was stored for a maximum of 15 days at 4°C.

For the transformation of commercial cell strains, the guidelines proposed by the manufacturer were followed carefully. In particular, the time for the heat-shock was reduced to 40 seconds.

2.3 Gel electrophoresis

Gel electrophoresis is a method for the separation and analysis of macromolecules such as DNA, RNA and proteins and their fragments, based on their size and charge. Three different kinds of electrophoresis were used in this thesis: Agarose Gel Electrophoresis, Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and native PAGE.

2.3.1 Agarose gel electrophoresis

Electrophoresis agarose gels were used for detecting the integrity of DNA plasmids, the result of a specific reaction involving DNA, or for the purification of DNA through nucleic acid gel extraction.

1% agarose gels (0.5 g of agarose powder) were made up to 50 mL with 1 \times TBE buffer from a 10x stock solution (108g Tris base, Boric Acid 55g, EDTA 9.3g). The solution

was heated in a microwave until the powder was dissolved. Then, Syber safe DNA stain (Invitrogen, Saint Aubin, Essonne, France) was diluted one thousand times inside the gel solution. Finally, the agarose solution was poured into a Bio-Rad mini-sub DNA cell (Bio-Rad, Marnes-La-Coquette, Hauts-de-Seine, France), and left to cool for about 30 min until the gel had set.

Samples were prepared by mixing with a commercial agarose gel loading dye (Invitrogen, Saint Aubin, Ile de France, France) and centrifuged before loading. An appropriate DNA ladder (Euromedex, Souffelweyersheim, Strasbourg, France) was chosen according to the size of the sample.

Gels were usually run in TBE buffer for 20-30 minutes at 80 V. The final result was visualised by exciting the syber safe with UV irradiation. Images were recorded using a image detector from Thermo Fisher Biolab Scientific (Illkirch, Bas Rhin, France).

2.3.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse the protein content of samples. SDS-PAGE uses a denaturing polyacrylamide gel, where the separation of proteins is based on the size of the macromolecules. SDS-PAGE was used for analysing the protein purity during the purification process.

Mini-gels (10 cm x 10 cm) were run according to the conditions of Laemmli (Laemmli, 1970). Running or separating gel, was prepared with an acrylamide percentage up to 15% (w/v), while stacking gels were prepared at 4% (w/v). Gels were prepared from stock solutions in Milli-Q water. The SDS PAGE images reported in the thesis had an acrylamide percentage of 12% (w/v), for which the running (Tris HCl pH 8.8 390 mM, SDS 0.1%) and stacking gel (Tris HCl pH6.8 125 mM, SDS 0.1%) composition are indicated in Table 2.1.

Table 2.1 SDS PAGE 12% (w/v) composition

Stock solutions	1.5 M Tris HCl pH8.8	0.5 M Tris HCl pH6.8	30% Acrilamide	10% Ammonium Persulfate (APS)	10% SDS	TEMED	H ₂ O
RUNNING GEL (12%-1 gel)	1.3 ml	-	2 ml	0.05 ml	0.05 ml	0.02 ml	1.6 ml
STACKING GEL (4%)	-	1.25 ml	0.5 ml	0.050 ml	0.050 ml	0.02 ml	3.2 ml

Samples were mixed with 3X loading buffer [0.3 M Tris-HCl pH 6.8, 10% (w/v) SDS, 25% (v/v) glycerol, 5 mM β -mercaptoethanol and 0.015% (w/v) bromophenol blue] and boiled for 5 minutes at 100 °C. After a quick spin, they were loaded into wells. Markers Broad-range and Low-range Bio-Rad were used (Bio-Rad, Marnes-La-Coquette, Hauts-de-Seine, France). The first is a pre-stained marker, while the second was prepared by diluting ten fold in the loading buffer. Gels were run in running buffer (Solution 1L, 10X: 30.3 g Tris base, 14.4 g Glycine, 10 g SDS) in a protein gel electrophoresis unit (Bio-Rad, Marnes-La-Coquette, Hauts-de-Seine, France), at constant current of 30-45 mA.

Proteins were visualised by soaking gels in Coomassie Blue staining solution (40% (v/v) methanol, 10% (v/v) acetic acid and 0.25% comassie blue) for 5-15 min and by then transferring them into de-staining solution (75% (v/v) ethanol 25% (v/v) acetic acid).

2.3.3 Native PAGE gel electrophoresis

Native gels are useful for the analysis of proteins in non denaturing conditions. In this thesis they were used exclusively for the MEK6-p38 α project, therefore, the protocol is reported in Chapter 7, Section 7.3.2.

2.4 Macromolecule quantification

Two different methods were used in this thesis for following the recovery of DNA or protein: UV absorption, and the Bradford assay. The first technique is used to quantify both macromolecules, while the second one is only for protein quantification.

2.4.1 UV adsorption

The UV absorption method is based on the macromolecule's ability to adsorb UV radiation at specific wavelengths. Nitrogenous bases in nucleic acids absorb at 260 nm, while aromatic residues in proteins absorb at 280 nm. The ratio between the two absorbance values is further used for characterising the amount of nucleic acid or nucleotide contamination in a protein sample, and *vice versa*.

The concentration was calculated using the Lambert Beer Law:

$$A = l\varepsilon C$$

where A is the absorbance value at 260 or 280 nm, l is the path length (cm), ε is the molar extinction coefficient ($M^{-1}cm^{-1}$) and C is the nucleic acid or protein concentration. A Thermo Scientific NanoDrop 1000 was used to measure the UV absorbance of samples. This instrument allows the analysis of 0.5-2.0 μ l samples, without the need for cuvettes or capillaries. Thus, l was always corresponded to 0.1 cm.

The molecular extinction coefficient is the measure of how strongly a substance absorbs light at a particular wavelength. The ε of nucleic acids is automatically calculated by the Nano drop control software, by indicating the nature of nucleic acid in the sample (ds DNA, ssDNA, RNA). In the case of proteins or peptides, the value was calculated using the Prot Param tool, from the ExPASy website (www.expasy.org), which provides two different numbers based on a protein's oxidation state. Since all the buffers were supplied with DTT, the ε in reduced conditions were used.

Practically, 2 μ l were usually used for each experiment. Each was preceded by a blank measurement using buffer, and was repeated three times in order to minimize operator error.

2.4.2 The Bradford assay

The Bradford protein assay is a colorimetric protein assay used to measure the concentration of total protein in a solution. It is based on an absorbance shift of the dye Coomassie Brilliant Blue G-250, (Bio-Rad, Marnes-La-Coquette, Hauts-de-Seine, France), which has a maximum of absorption at 595 nm. A Biophotometer

spectrophotometer (Eppendorf, Hamburg, Germany) was used.

The dye is converted from red into blue upon binding the protein being assayed. During the formation of this complex, two types of bond interaction take place: the red form of Coomassie dye first donates its free electron to the ionizable groups on the protein, which causes a disruption of the protein's native state, consequently exposing its hydrophobic pockets. These pockets on the protein's tertiary structure bind non-covalently to the non-polar region of the dye *via* Van Der Waals forces and further ionic interactions are also formed. The binding of the protein stabilizes the blue form of the Coomassie dye; thus the amount of the complex present in solution is proportional to the protein concentration, and can be estimated by the use of an absorbance reading.

The assay requires a standard curve, which was prepared and calculated by recording the 595 nm absorption of a series of protein samples with known concentration. Bovine albumin serum (BSA) samples were prepared in Milli-Q water in the concentration range 0.5-10 mg/ml and the final linear regression equation was calculated from the curve obtained.

Protein concentration was determined following the guidelines provided by the Bradford reagent producer. The original solution was diluted 1:5 in water and 1 ml was used for each experiment. From 1 to 5 μ l of protein sample were added in the cuvette containing the reagent and the concentration was calculated. A blank was always performed in order to avoid that the buffer could interfere with the final result.

2.5 Protein expression and purification

Expression and purification of β PGM and β PGM_{K145A} were performed in the laboratory of Prof. Jon Waltho in Sheffield. Protocols are described in the thesis of Dr. J Griffin and Dr. Y Jin (Griffin, 2011; Jin, 2012). The following paragraphs will refer only to the proteins expressed and purified in our laboratory.

The recombinant proteins purified in this thesis were inserted in two vectors: pGEX-2T and pET15b. Due to the similarity between the two plasmids, expression and purification of the recombinant proteins were performed in a similar manner. In the following paragraphs the general guidelines followed for expression and purification will be described, while greater details are reported in specific chapters.

2.5.1 Plasmid description

Human RhoA and human RhoGAP_{R85A} were expressed and purified in a recombinant form in the pGEX-2T vector. Human MEK6 and truncated or mutated forms of this protein were expressed and purified using the pET15b vector. Human p38 α , and mutated forms of this protein were expressed and purified using the pET15b vector pET-BS (+). Protein expression is regulated by the *lac* and T7 RNA polymerase promoters, which confer inducible and high protein expression. Isopropyl- β -D-thiogalactopyranoside (IPTG) is used to induce protein expression. Both expression vectors provide ampicillin resistance to the bacteria and add a cleavable N-terminal GST (glutathione-S-transferase) tag or an N-terminal His₆ (6 histidine residues) tag respectively to the N-terminus of the protein when it is expressed. A thrombin cleavage site was coded in order to remove the affinity tags.

2.5.2 Transformation of competent cells

All the recombinant proteins, except β PGM, described in this thesis have human origin. Thus, they were expressed in competent Rosetta 2 BL21 (DE3) *E. coli*. The Rosetta 2 host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These strains supply tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on a compatible chloramphenicol-resistant plasmid.

Transformation was carried out in a similar way for *E. coli* strains used for amplifying the DNA plasmids (see Section 2.2.2). After transformation, 50-100 μ l of transformed cells were transferred onto the LB agar plate supplemented with 200 μ g/ml of ampicillin and 68 μ g/ml of chloramphenicol in the case of Rosetta2. After overnight growth at 37°C, the plates were stored for a maximum of 15 days at 4°C.

2.5.3 Large scale protein expression and protein harvesting

Large scale protein expression (2-12 L) was carried out in similar conditions for all the proteins. To produce each protein, a pre-culture was prepared, picking one colony from the bacterial plate and adding it to 25-100 ml of LB supplemented with ampicillin and

chloramphenicol and grown at 37°C overnight. Ampicillin and chloramphenicol were always added in the ratio 4:1.

Then, 12 ml of preculture was used to inoculate 750 ml LB-ampicillin-chloramphenicol. Cells were grown at 37 °C with shaking at 140 rpm until an optical density at 600 nm (OD₆₀₀) of approximately 0.6 had been reached. Protein expression was then induced with 1mM of IPTG and the cells incubated overnight at 20°C. They were then harvested by centrifugation at 4°C, at 4000 g for 15 minutes. Cell pellets were directly used for purification or they were flash-frozen in liquid nitrogen and stored at -80°C.

2.5.4 Cell lysis

The purpose of the lysis step is to disrupt the cells in order to release the recombinant protein into an aqueous “extract” which is the first stage of protein purification.

The cell pellet was disrupted by both enzymatic and mechanical lysis. Cell pellets were gently resuspended in lysis buffer (LyB). LyB was always supplied with the protease inhibitor cocktail COMPLETE (ROCHE, Basel, Switzerland), which reduces protein degradation. 10 ml of LyB was used for each flask of 750 ml of culture. Following the addition of deoxyribonuclease I (ROCHE, Basel, Switzerland) to break down contaminating DNA, cells were lysed mechanically using a sonicator (Thermo Fisher biolab Scientific, Illkirch, Bas Rhin, France). Sonication was performed using a maximum of 3 cycles of 50 sec at 60% power. During all these steps the sample was kept on ice. Finally, in order to remove cell debris and insoluble recombinant protein, the sample was centrifuged at 4°C, at 20000g for 30 minutes and the supernatant saved for the next purification step.

2.5.5 Protein purification

All the proteins contained a N-terminal tag. Resins that specifically bind GST or the His tags were used as the first step for purifying the recombinant proteins (affinity chromatography step). After tag cleavage, each protein was always further purified using size exclusion chromatography and occasionally ion exchange chromatography. The level of purification achieved was judged by a combination of different techniques.

SDS-PAGE was used to analyse protein purity. The elution profile recorded at two different wavelengths (280 and 260) provided an indication of possible nucleic acid or nucleotide contamination. The elution profile in combination with data produced by DLS analysis (dynamic light scattering), was used to assess protein poly-dispersion before crystallisation experiments and for comparing different protein batches. The final yield was calculated by using one of the methods described for quantification of protein concentration.

2.5.5.1 Affinity chromatography and Thrombin cleavage

Two pre-packed columns were used: a GST-trap column and a His-trap column (GE Healthcare).

In the GST-trap column the glutathione ligand is coupled via a 10-carbon linker to highly crosslinked 4% agarose. The coupling is optimized to give high binding capacity for GST-tagged proteins and other glutathione binding proteins. The His-trap column is composed of Ni-NTA Sepharose. Nickel ions are coupled by NTA (tetradentate chelating adsorbent nitrilotriacetic acid), which occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the His₆-tag. Elution of the tagged protein is performed by competition: glutathione in the case of GST-trap and imidazole for the His-trap.

Columns were connected to an AKTA prime system (GE healthcare, Buc Cedex, France) and pre-equilibrated with Binding Buffer (BB). The soluble fraction derived from lysis step was applied at low flow rate (0.5-1 ml/min) and the column was washed extensively until UV absorbance was stable.

Tag cleavage was performed differently for the two columns. For proteins purified on the GST-trap column, tag cleavage was performed on column. For the His-tag proteins, the cleavage was performed after elution with elution buffer. Thrombin (100U for each 20 mg of protein) was applied to the column or added to the protein sample. Thrombin is a pancreatic enzyme that recognises a specific protein sequence, inserted between the tag sequence and the first methionine of the recombinant protein. Cleavage was performed overnight at 4°C, with gentle shaking. In order to separate the cleaved protein from the un-cleaved and from active thrombin, a Benzamidinne-FF trap column

was connected (GE Healthcare, Buc Cedex, France) at the bottom of the GST-trap or the His-trap column. The cleaved protein was then eluted and purified in one step from un-cleaved protein and thrombin. Further details of this step will be described in Chapters 5 and 6.

2.5.5.2 Size exclusion chromatography

The sample obtained after thrombin cleavage was further purified by size exclusion chromatography.

Size exclusion chromatography separates molecules according to differences in their hydrodynamic size as they pass through a gel filtration medium packed in a column. The medium is a porous matrix in the form of spherical beads with a well-defined range of pore sizes. Molecules that are small enough can fit inside all the pores in the beads and elute in the last part of the column volume. Molecules that are larger than the pore size pass through the beads avoiding the pores and elute first.

Several columns were used for this step, based on protein size and purification yield: HiLoad 16/60 Superdex 200 or 75 pg or a Superdex 200 or 75 10/300 GL (GE Healthcare, Buc Cedex, France). Columns were connected to an AKTA purifier system at RT or to an AKTA explorer system at 4°C (GE Healthcare, Buc Cedex, France). Columns were equilibrated with at least one column volume of buffer (SEC buffer). In order to achieve good peak separation, the sample volume never exceeded 1% of column bed volume. Usually, a Centricon (Millipore Ltd, Herts, UK) concentrator was used to concentrate the sample to the desired volume before loading. The UV absorbance profile was recorded during the chromatography experiment with two different wavelengths (280 and 260 nm). Finally, the purity of fractions collected was analysed on SDS-PAGE. The fractions were then pooled and, after flash freezing in liquid nitrogen, stored at -20° C.

2.5.5.3 Anion exchange chromatography

In the case of p38 α and p38 α mutants, the protein was further purified using a MonoQ anion exchange column. In these experiments, the separation of particles depends on the

reversible adsorption of charged molecules to immobilised ion exchange groups of the opposite charge. A MonoQ 5/50 GL column (GE Healthcare, Buc Cedex, France) was connected to the AKTA purifier and equilibrated according to manufacturer's guidelines. Samples in low buffer salt were applied at low flow rate (0.3 ml/min). After extensive washing, a slow NaCl gradient at 0.5-1 ml/min was applied. Increasing the NaCl concentration changes the ionic strength, which allows the elution of the bound proteins at different NaCl concentrations. Analysis of the results obtained is described in Chapter 6, paragraph 6.3.2.

2.5.5.4 Buffer exchange

During the purification protocols described buffer exchange was often useful. Two different tools were used: a dialysis membrane and PD-10 columns.

Dialysis was always performed at 4°C ON, with a typical ratio of 1:200 of sample to dialysis buffer. Dialysis membranes with a molecular cut off of 7 kDa were used (Spectra/Por, Breda, Netherlands).

PD-10 columns were used when rapid RT buffer exchange was required. PD-10 columns are Sephadex-G25 columns (GE Healthcare, Buc Cedex, France), which allow the separation of proteins from small molecules, such as salt, as in a SEC chromatography experiment. The protocol provided by the manufacturer was followed.

2.5.6 Dynamic light scattering

Dynamic light scattering (DLS, Zetasizer Nano, Malvern Instruments Ltd., Worcestershire, UK) is used to analyse the quaternary structure of a protein, to evaluate its aggregation state and its polydispersity. In DLS experiments, photons are scattered by the randomly oriented molecules in the sample analysed. The radius of these molecules (r) is calculated by the Stokes-Einstein relation:

$$r = k_b T (6\pi\eta D_T)^{-1}$$

where k_b is the Boltzman's constant, T is the absolute temperature (in Kelvin), η is the

solvent viscosity and D_T is the diffusion coefficient. The DLS measures the time dependent fluctuations resulting from the Brownian motions in the scattering intensity to determine the Dt and subsequently the hydrodynamic radius. r is the radius of a hypothetical hard sphere that diffuses with the same strength as the particle under examination.

From the intensity, the reordered volume distribution is calculated with a polydispersity index (PDI). The PDI is a number between 0 and 1. A PDI value of 1 indicates that the sample has a very broad size distribution and may contain large particles or aggregates that could be slowly sedimenting. Values under 0.3 (monodisperse or low polydispersity) were accepted as an indication that a sample could be suitable for crystallisation experiments.

Protein samples purified for crystallisation were usually analysed by DLS before being used. After SEC chromatography, fractions corresponding to the central elution peak were pooled and concentrated. 50 μ l of the sample were centrifuged and used for the DLS analysis.

2.6 X-ray Crystallography

2.6.1 Crystallisation

Three-dimensional protein crystals are grown by forming a super-saturated solution and causing precipitation sufficiently slowly, and in such a manner, that ordered arrays of the protein are formed. This crystallisation process consists of two major events: nucleation and crystal growth. Nucleation is the step where the solute molecules dispersed in the solvent start to gather into clusters, which can re-dissolve or become stable nuclei. From these nuclei, crystal growth can continue. Nucleation and growth occur simultaneously while the sample remains supersaturated (Figure 2.1).

To supersaturate a protein solution there are two main possibilities: increasing the protein concentration, or adding a second reagent that reduces solubility. Reagents, such as salts or polyethylene glycols (PEGs), that ‘compete’ for the water in the solution, bring the protein out of solution.

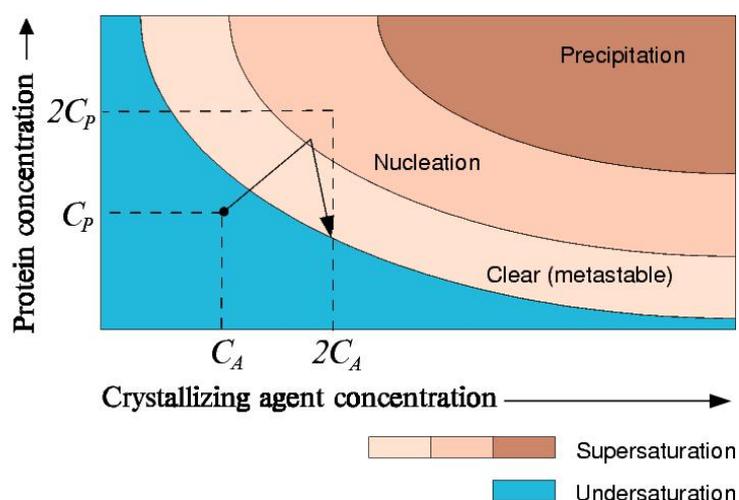


Figure 2.1 Schematic view of solubility curve of a protein. The supersaturation state is divided into three zones: Clear, Nucleation and Precipitation. Only in the nucleation zone do the protein crystals nucleate and grow, while in the first zone the solution may not nucleate for a long time, even if preformed crystals growth would be sustained, and in the precipitation zone, proteins do not nucleate but precipitate out of solution (<http://www-structmed.cimr.cam.ac.uk>).

Proteins or protein complexes described in this thesis were crystallised using the sitting drop vapour diffusion method. In this technique, a drop composed of a mixture of sample and reagent, is placed in a vapour equilibration chamber with a liquid reservoir composed of the reagent, so water vapour slowly leaves the drop to achieve equilibrium, eventually ending up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir. In every individual experiment, the drop was prepared by mixing 1-2 μl of complex solution with 1-2 μl of precipitant solution. For each precipitant, 500-700 μl of buffer reservoir was used in each well. Sitting drop plates were covered with adhesive tape (Hampton Research, Aliso Viejo, CA, USA) (Figure 2.2).

For some of the experiments described in this thesis, the microbatch technique was also used. The principle is that the precipitating reagent is instantaneously added to a protein solution, suddenly bringing the solution to a state of high supersaturation. The drop is then covered by oil and the superstauration is achieved by slow evaporation (Figure

2.2).

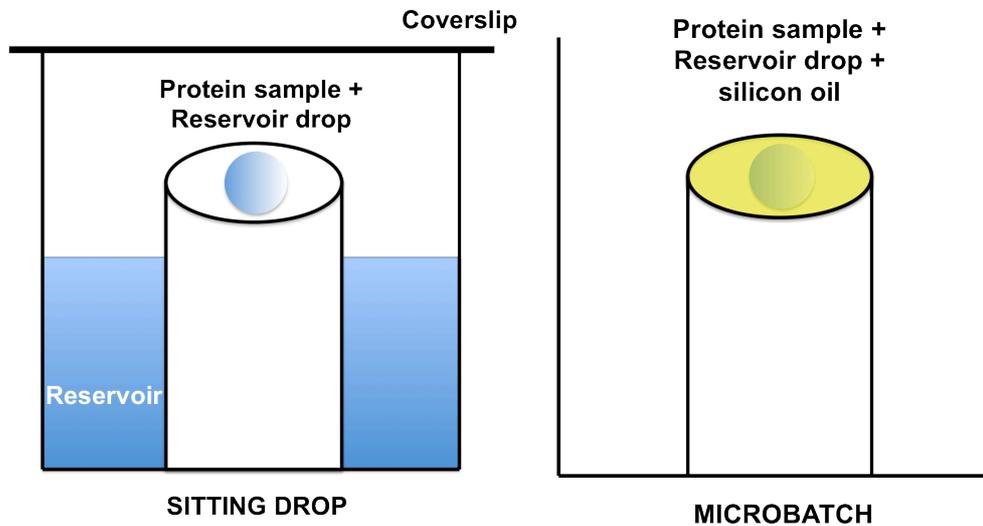


Figure 2.2 The sitting drop and microbatch methods. The same kind of plate was adapted to both methods.

In the case where crystallisation conditions are unknown, several Hampton commercial screens (Laguna Niguel, CA, USA) were tested at the EMBL High Throughput Crystallization Laboratory (HTX laboratory), EMBL, Grenoble, France. Sitting drop plates for high throughput crystallisation are also used by the HTX laboratory. All the experiments were carried out in similar way: the drop was prepared by mixing 1nl of complex solution with 1 nl of precipitant solution. For each precipitant, 100 nl was used in each well. Experiments were run at RT, except for Δ MEK6DD, which was performed at 4°C.

Optimisation of initial conditions was performed by changing the pH of the solution, solute concentrations or temperature. For the crystallisation of the RhoA-GDP-MgF_x/AlF_x-RhoGAP_{R85A} complex, seeding was also useful. Seeding is a technique used to transfer stable nuclei or micro/macro crystals in a similar or different precipitant-protein drop where crystal nucleation is difficult but crystal growth is favourable.

2.6.2 Cryocooling

Protein crystals are extremely sensitive to damage by X-rays caused by free radicals, which can seriously compromise the quality of the data collected. Thus, before being exposed to X-rays crystals required to be properly cryo-cooled. Cooling the crystals to 100 K dramatically reduces the damage to the crystals by preventing the chain reaction of radical damage (Garman, 1999). This eliminates the radical's diffusion in the crystal. Typically, the cooling procedure consists of transferring a crystal from the mother liquor to a cryo solution; then the crystal is quickly plunged into liquid nitrogen.

The presence of a cryoprotectant in the cryosolution is necessary to prevent the formation of ice crystals that can damage the crystal lattice (Garman & Schneider, 1997).

Common cryo-protectants are glycerol, PEG400, ethylene glycol, oil and MPD. Cryosolutions contained protein buffer conditions, reservoir conditions and a cryoprotectant at 20-25%. The preparation was carried out in two different ways. A quick protocol consisted of mixing all the components in an eppendorf tube. After equilibrating the solution obtained at the same temperature as crystal growth, the cryo-protectant was ready to use. A longer, but more efficient protocol, was then formulated. Cryo-protectant was prepared by repeating the protein crystallisation experiment by using the protein buffer supplied with the cryo-protectant instead of the protein. After 24 hours equilibration the cryo-solution was ready to use. MicroMount (MiTeGen, Ithaca, NY, USA) loops were then used for mounting the crystals in the cryo-cooling procedure.

During this thesis, a new cryo-protectant protocol was developed. In this protocol, crystals are cryo-cooled without adding any cryo-protectant. Instead, crystals are mounted using a MicroMesh loop (MiTeGen., Ithaca, NY, USA) and mother liquor is removed by touching the reverse of the mesh to laboratory tissue paper (Kimtech KimWipes, Kimberly Clark, Irving, Texas, USA) or by leaving them to dry out for a few seconds. The removal of the mother liquor limits the formation of ice by reducing the amount of liquid surrounding the crystal. In these conditions the protein itself is probably acting as the cryo-protectant by inhibiting the formation of crystalline ice within the solvent channel. Details of this protocol had been published (Pellegrini *et al.*, 2011).

2.6.3 X-ray crystallography analysis

2.6.3.1 Brief description of the beamlines used

Crystals were taken to the European Synchrotron Radiation Facility (ESRF, 6 Rue Jules Horowitz, Grenoble, France), where the structural biology Beamlines ID14-1, ID14-2, ID14-4, ID29, ID23-1 and ID23-2 were used. The choice of beamline was based on crystal size or on the experimental requirements. ID14-1, 14-2 and 14-4 were mostly used for screening or data collection from well ordered crystals. ID29 equipped with a Pilatus 6M detector was used when collection of several datasets with high quality was required. ID23-1 and 23-2 were chosen when a 5-10 μm beam was necessary. The essential features of these beam lines are shown in Table 2.2. The sample environment is comprised of a cryostream to maintain the sample at 100K, a diffractometer and a sample changer, which is an automatic crystal mounting tool. All the beamlines are controlled using the same software system, MxCuBE, which allows users to interact with beamline hardware and provides an intuitive interface for managing the entire X-ray experiment from sample loading and detection, to on-line data collection and X-ray emission spectra (Gabadinho *et al.*, 2010).

Long wavelength data collection on RhoA-GDP-MgF_x/AlF_x-RhoGAP_{R85A} crystals was performed at the BESSY Synchrotron (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung, Berlin, Germany), at MX beamline 14.2. The beamline has an energy range of 5-15.5 KeV, and a beam size of 180×70 μm^2 (H×V).

Table 2.2 Essential features of the ESRF beam lines

	ID14-1	ID14-2	ID14-4	ID23-1	ID23-2	ID29
Energy range	13.27 KeV	13.294 KeV	9-15 KeV	5-20 KeV	14.2 KeV	6-20 KeV
Beam size (H×V)	100 × 100 μm ²	100 × 100 μm ²	50 × 50 μm ²	20 × 30 to 10 × 10 μm ²	8 × 5 μm ²	50 × 50 to 10 × 10 μm ²
X-ray Detector	ADSC Q210CCD	ADSC Q4R	ADSC Q315r CCD	ADSC Q315R	Mar/Rayonix 3×3 Mosaic 225	Pilatus 6M (Dectris)
Flux	1 × 10 ¹¹ ph/sec	1 × 10 ¹¹ ph/sec	1 × 10 ¹² ph/sec	5 × 10 ¹² ph/sec	5 × 10 ¹¹ ph/sec	1 × 10 ¹³ ph/sec

2.6.3.2 Data collection

Collecting two images 90 degrees away from each other in ω orientation was used to assess the quality of the diffraction for each crystal. These images are named reference images; they provide information about the quality of the diffraction in different orientations, the resolution limit and the level of order in the crystal. At the ESRF, the collection of the two images is included in the **EDNA** flow process (**Enhanced automateD collectioN of data**). EDNA is a software program implemented in MxCuBE, which provides diffraction characterisation and data collection strategy calculation (Incardona *et al.*, 2009). After collection of the reference images, autoindexing is performed by EDNA through the program Mosflm (Leslie, 2006), which calculates crystal orientation, probable space group and unit cell parameters. Once the orientation of the crystal is determined, a strategy function is used to calculate the starting angle and minimum ϕ range needed to collect a complete data set. The tool used in this case is **BEST** (Popov & Bourenkov, 2003), which takes into account radiation damage effect using the program **Raddose** (Paithankar & Garman, 2010).

Usually, when a strategy at a reasonable resolution was proposed, datasets were collected using the suggested strategy. Searching for better diffracting volume within the crystal was performed by changing crystal orientation using the MiniKappa goniometer or by evaluating sample diffraction properties through diffraction cartography (Bowler *et al.*, 2010b). This automatic scan with micro beam has been used successfully in this thesis as tool for finding the best diffracting point of small p38 α

crystals.

At the BESSY synchrotron both indexing of reference images and strategy calculation were performed using iMosfilm. iMosfilm is the graphical interface of Mosfilm (Battye *et al.*, 2011) Searching for the better diffracting positions was performed manually.

2.6.3.3 Data processing

The goal of the data processing step is to produce from a set of diffraction images a set of indices (hkl) with their associated intensities. Data processing was performed with two programs: iMosfilm and XDS. Both programs go through fundamental steps: spot finding and determination of the most probable space group (indexing), cell refinement and intensity measurement (integration).

Mosfilm has been used from the graphical interface iMOSFLM (Battye *et al.*, 2011). In this program, Indexing is performed as default on two images 90 degrees apart. It then produces a list of possible space groups and unit cell parameters; generally, the option with the lowest penalty and highest symmetry space group is correct. The accuracy of the choice of unit cell is verified by placing boxes where spots are predicted from the selected space group: if the boxes overlay all observed reflections, the unit cell is correct. After space group solution choice, mosaic spread can be estimated (see Appendix A). Refinement of the unit cell and the mosaic spread is then performed, by measuring observed reflection in a certain number of images. Finally, images are integrated in blocks of 5-10 images. The integration is a two-step procedure: first the position in the image for each Bragg peak is predicted; second the intensity for each hkl is measured. In Mosfilm, mosaicity is continuously refined during integration. The final intensities are written to an *mtz* file.

XDS is a command line program, which run 8 different programs in sequence from the input file XDS.INP, which contains information about the experimental details (detector, beam position etc) and Jobs to run. First, image correction, background calculation and spot finding are run (programs: XYCORR, INIT and COLSPOT). Then the IDXREF program calculates the orientation, matrix and symmetry of the crystal lattice and refines all or a specified subset of parameters contained in the XDS.INP file. The job continues until 70% of the given spots can be accurately explained. If the

program fails further programs are required for better spot prediction. Finally, INTEGRATE and CORRECT are run. The first is the integration function, which follows the same principles described above for Mosflm. However, the method used is different. While Mosflm uses a two-dimensional integration, XDS uses a three-dimensional integration (Leslie, 2006). Thus, while in Mosflm the intensities of the different components of a partially recorded reflection are evaluated independently by two-dimensional profile fitting, in XDS they are assembled and a three-dimensional profile is used to evaluate the total intensity. CORRECT applies correction factors to the intensities and standard deviations of all reflections found, it determines the space group if unknown and refines the unit-cell parameters, it reports the quality and completeness of the data set and saves the final integrated intensities in the file XDS_ASCII.HKL (Kabsch, 2010). The file is converted to *mtz* format using the program Pointless (Evans, 2006).

2.6.3.4 Data reduction

In data reduction Structure Factors (F_s) are estimated from the intensities. Theoretically, the measured intensity I_h of a reflection h is proportional to the square of the underlying Structure Factor, with an associated measurement error, but systematic effects of the diffraction experiment break this proportionality. Such systematic effects include changes in the beam intensity, changes in the exposed volume of the crystal, radiation damage, bad areas of the detector and physical obstruction of the detector, such as the cryo-stream (Evans, 2011). Thus, before data reduction, the data are required to be on the same scale. First, **Pointless** was run to define the most probable Laue group after images integration. Intensity scaling and merging were performed through the program **Scala**. Finally, Structure Factors were calculated through **CTruncate**.

Indexing of the lattice in mosflm and XDS is based on lattice geometry, with no regard for the symmetry of the diffraction pattern, which can only be determined after integration. Thus Pointless (Evans, 2006) was used for scoring potential symmetry operators, looking at systematic absences and ranking the possible Laue group. Usually the Pointless solution was used for run a second time the integration step, by inserting the most probable space group in the XDS.INP file.

Then the data were scaled and merged. The program Scala tries to make symmetry-related and duplicate measurements of a reflection equal by modelling the diffraction experiment, principally as the function of the incident and diffracted beam directions in the crystal (Evans, 2011). Moreover, this step provides the best measure of the quality of the data. Several data-quality indicators are obtained, such as resolution limit, data completeness, multiplicity, Wilson B factor, R_{merge} and $I/\sigma(I)$ values. R_{merge} represents the discrepancy between the intensities of independently-measured symmetry related reflections, while $I/\sigma(I)$ represents the ratio between signal (I) and noise in the data. From these statistics decisions were made about elimination of problematic data or outliers, setting a new resolution cut off and possible third integration run with different parameters. The integrated and scaled data were then merged to form a single *mtz* containing the intensities for a complete data set.

The scaled and merged intensities were converted to structure amplitudes or (Structure Factors, F_s) by CTruncate (French & Wilson, 1978), which was run along with the Scala process. For perfect data, the module of F_s is the square root of the intensity. However, in presence of errors small intensities are wrongly estimated with this formula, with the final result that the negative intensities are set to zero and the contribution from weak reflections is underestimated. CTruncate estimates the “best” F_s based on the probability distribution of I . It outputs an *mtz* file containing the observed F_s and a Wilson B factor. The Wilson B factor plot should give a straight line which above 4 Å reflects how temperature factors and absolute scale of the intensities fall off with resolution. Deviation from the line indicates possible unreliable data. Moreover CTruncate provides statistical information about possible crystal problems, such as twinning, which indicates the presence of more than one lattice in the crystal (Evans, 2011).

At this stage a random 5% of the data were ‘flagged’ to be used in the *Free R_{factor}* calculation (see Appendix A).

2.6.3.5 Molecular replacement

For all the proteins described in this thesis a homologous structure is already known, the method for solving the phase problem was molecular replacement (MR) (see Appendix

A). The program **MolRep** from the CCP4i suite was used. The program requires as input the structure factors produced by CTruncate and a search model in the form of a PDB coordinate file. The goal of the molecular replacement program is to find the combination of rotation and translation functions, which allow the overlap of the Patterson map calculated from the model with the one derived from the reflection dataset (see Appendix A). MolRep goes through all the MR stages automatically. The rotation search is calculated for three orthogonal orientations of the model and averaged over them. Instead of the Rossman method, the more elegant fast rotation function from Crowther (Crowther, 1972) is used. The translation function (TF) is calculated by using the Patterson map and all the symmetry operators (Murshudov *et al.*, 1997) (see Appendix A). Structure factors are calculated for each orientation/position of the model by back calculation from the model coordinates. The TF is supported by a packing function, which eliminates the translation that allows the molecule to clash with each other. The TF concludes with a multi-domain rigid body refinement in space group P1 and provides as an output the position of the model in the unit cell. During the process a soft resolution cut off is applied to all calculation, which prevents the presence of systematic errors in the electron density, particularly at the surface of the molecule (Murshudov *et al.*, 1997). The output of the program is a list of translation function solutions together with the particular rotation function solution. The program outputs a contrast, which is the ratio of the top score to the mean score. At reasonable contrast (>2.5), the program provides a file of the model already incorporated in the unit cell.

2.6.3.6 Model refinement

In the final step in the process of solving a macromolecular crystal structure, refinement is carried out to maximise the agreement between the model and the X-ray data. There are several programs available that are either designed to perform refinement or offer refinement as an option. In this thesis **Refmac5** from the ccp4i suite and **Phenix.refine** from the Phenix suite were used (Adams *et al.*, 2010; Murshudov *et al.*, 1997). Moreover, programs such as Buccaneer, Arp/wARP and AutoSol were used for improving partial models by re-building or automatically adding water molecules (Cowtan, 2006; Perrakis *et al.*, 1999; Terwilliger *et al.*, 2009). The results of each

refinement cycle were validated visually using the graphical program **COOT**. COOT was also used for manual refinement (Emsley & Cowtan, 2004).

2.6.3.6.1 Refinement using Refmac5 and phenix.refine

The model produced by MolRep is refined against the observed data using the program Refmac5 (Murshudov *et al.*, 1997) and the program phenix.refine, which are likelihood based macromolecular refinement programs. Refinement programs differ in the target functions and optimisation techniques used to derive model parameters. The target function minimised in Refmac5 (f_{total}) comprises two components: a component utilising geometry (or prior geometry) and a component utilising experimental knowledge:

$$f_{total} = f_{geom} + wf_{xray}$$

Thus, refinement works through likelihood function which transfers optimal information from the data to the derived model. In this process f_{geom} ensures that the derived model is consistent with the presumed chemical and structural knowledge. In the target formula w is a weight between the relative contributions of the two components. In the refinement cycles with Refmac5 the value is varied in order to obtain the one that gives the lowest *Free* R_{factor} and best stereochemistry (Murshudov *et al.*, 2011) (see Appendix A).

The target function in phenix.refine is slightly different. F_{geom} is indicated as F_{model} , and it contains geometric and non geometric parameters such as bulk solvent contributions, twinning, crystal anisotropy and other parameters. Refinement goes through three main steps: model parameterisation, refinement of the target and optimisation method. The program optimises the target weight value automatically in each refinement cycle (Afonine *et al.*, 2012). Phenix.refine provides sophisticated algorithms, which are not present in Refmac5. In this thesis rotamer optimisation and simulating annealing (see Appendix A) were used for improving the final model.

Both programs have a dictionary of bond angles, lengths and other restraints for polypeptides and ligands to maintain correct stereochemistry within the model during

refinement. These dictionaries can be implemented by providing to the program *cif* files about the molecule of interest. Dictionaries for un-recognised molecules, such as magnesium fluoride, aluminium fluoride and fluorinated compounds were built using **eLBOW** or **PRODRG** (Moriarty *et al.*, 2009; Schuttelkopf & van Aalten, 2004).

The quality of the refinement was assessed by the R_{factor} values and root mean square deviations (RMS deviations) in bond angle and chirality (see Appendix A). In this thesis, Refmac5 was run in the early stages, while Phenix.refine was used later and during the validation step. Non-crystallographic symmetry (NCS) and Translation Libration Screw-motion (TLS) were refined with Phenix.refine (see Appendix A).

2.6.3.6.2 Manual rebuilding

After each refinement round, the result was assessed with **COOT**, a molecular modelling program (Emsley & Cowtan, 2004; Emsley *et al.*, 2010), which requires as input files the model and phases calculated with Refmac5 or phenix.refine. From the model and phases, COOT calculates two maps, the Fourier map and the difference Fourier map (see Appendix A). The program allows the maps and model to be displayed in three-dimensions and the rebuilding of the model to match the observed electron density. The maps were inspected initially for their quality and for novel features. Then manual rebuilding was carried out by searching for the highest peak in the difference Fourier map (either positive or negative) and interpreted by eye. This procedure was performed from the command ‘Difference map Peaks’, which shows the differences between the model and the observed data. The largest features in the difference maps were due to missing ligands, missing residues and missing ordered solvent. Additional positive and negative peaks were derived from non-fitting side-chain positions and disordered loop regions. Ligands (for example the magnesium fluoride) were not included in the model until further refinement had been performed, so as not to bias the maps. Real-space refinement, regularisation and rotate/translate zone from the ‘Model/Fit/Refine’ toolset were used for regularising the coordinates of the model and increasing the fitting to the map. Ramachandran restraints were added at resolution < 2 Å. After manual rebuilding, the coordinates were written out and refined against the observed X-ray data. This process was repeated until the difference density peaks

remaining could not be interpreted.

2.6.3.7 Structure validation

The final model was validated (see Appendix A), through the comparison against the electron density and by performing geometrical checks of protein structures. First validation was performed in COOT. COOT has a validate toolset which allows geometrical checks such as deviation from Ramachandran plot and Kleywegt plot, water geometry and rotamer analysis. Moreover, the ‘density fit analysis’ tool was used to validate the fit of the model to the map (Emsley *et al.*, 2010).

Further validation was performed using external tools, usually run within the phenix.refine cycle. **Molprobit** is a structure validation tool, which provides at the end of each cycle a list of geometrical errors and a summary of the observed and target values for assorted validation criteria such as, percentage of outliers in the Ramachandran plot, percentage of rotamer outliers and percentage of C-beta outliers (Chen *et al.*, 2010). The list of geometrical errors indicates deviations in bond length, bond angle, dihedral (torsion) angles, chiral centers and planar groups (such as aromatic rings). All-atom contact analysis is also run after adding hydrogen atoms to the model using the program phenix.reduce. Reduce will flag residues whose sidechains require flipping based on hydrogen-bonding geometry and clashes caused by newly added hydrogen atoms. These include asparagines, glutamines, and histidines, which are easily poorly fit, due to the apparent symmetry of the sidechain without hydrogen atoms (Adams *et al.*, 2010). Finally Molprobit performs a real-space correlation analysis, which provides similar results to the density fit analysis run with COOT (Chen *et al.*, 2010).

Polygon is a program used to compare model quality indicators to similar structures in the PDB databank. Pre-computed values for a selection of 1000 structures determined at similar resolution are plotted radially as one-dimensional histograms and the lines connecting these points form a complete polygon. For a high-quality, well-refined structure, the shape should be approximately symmetric and small (Urzhumtseva *et al.*, 2009).

Improving geometrical properties was judged by running phenix.refine. Small decreases

of *Free R_{factor}* and a decreased gap between *R_{factor}* and *Free R_{factor}* were positive indicators. Moreover, improvements in the Molprobit validation statistics and Polygon geometry after phenix.refine cycles were considered positive.

2.7 Figure preparation

All images of models, electron density maps and other structures were produced using the molecular graphics program **PyMOL** developed by DeLano, WL 2002 (www.pymol.org). The coordinates for solved structures, used to prepare figures, were downloaded from the Protein Data Bank [<http://www.pdb.org/>, (Berman *et al.*, 2000)]. Maps were calculated using the program **FFT** from ccp4i (Ten Eyck, 1985). Structures of small molecules and schematic views of protein structures were prepared with **CS Bio Draw Ultra**, (Cambridge soft).

Chapter 3 The Major and the Minor conformers in β PGM-G6P-MgF₃⁻-TSA

3.1 Introduction

3.1.1 β PGM is a pharmacological target

Phosphoglucomutases catalyse the interconversion of D-glucose-1-phosphate (G1P) and D-glucose-6-phosphate (G6P) (Bailey *et al.*, 1970). This reaction is central to energy metabolism in all cells (Lahiri *et al.*, 2002). Operating in the forward, G6P-forming direction, this reaction links polysaccharide phosphorolysis to glycolysis. In the reverse direction, the reaction provides G1P for the biosynthesis of cell wall polysaccharides in bacteria, in particular, exo-polysaccharides (Qian *et al.*, 1997; Qian *et al.*, 1994; Ramos *et al.*, 2001).

There are two classes of phosphoglucomutases, the α -phosphoglucomutases (α PGM), ubiquitous among eukaryotes and prokaryotes, and the β -phosphoglucomutases (β PGM), present only in certain bacteria and protists. Both carry out the same reaction, employing Mg²⁺ and a bisphosphate glucose intermediate as cofactors (Qian *et al.*, 1997; Qian *et al.*, 1994). In addition, both mutases are monomeric proteins. However, the two protein families are quite different in terms of structure and reaction specificity. Indeed, the two classes of mutases are distinguished by their protein-fold and their specificity for α - or β -D-glucose phosphates. These differences make β PGM an interesting pharmacological target. Therefore, a complete dissection of the isomerisation reaction it catalyses is a relevant goal in order to better understand its functional role.

Bacterial β PGM from *Lactococcus lactis* was the first β PGM crystallised and many of the studies on this family have been performed using this bacterium as a protein source. Following this path, a recombinant form of β PGM from *L. lactis* was used in this thesis as a model of the reaction. First, the role of this protein in *L. lactis* will be described, followed by structural features and the *state of the art* of the reaction. Differences between the α - and β -phosphoglucomutase will be also elucidated.

3.1.2 In *L. lactis* β PGM belongs to maltose and trehalose metabolic pathways

L. lactis is a Gram positive bacterium; it does not produce spores and is not mobile. It is used extensively in the production of buttermilk, yogurt and cheese due to its ability to produce lactic acid from lactose sources. The lactic acid produced by the bacterium curdles milk, which then separates to form curds, which are used to produce cheese and whey (Madigan *et al.*, 2006).

In *L. lactis*, β PGM is involved into two pathways: the maltose and trehalose catabolic pathways, which are linked through their metabolite, β -D-glucose-1-phosphate (β G1P). In other bacteria, β PGM is only involved in maltose degradation. Maltose is degraded by the concerted action of maltose phosphorylase and β PGM, whereas trehalose is assimilated by a novel pathway, including trehalose-6-phosphate phosphorylase, and β PGM (Andersson & Radstrom, 2002a) (Figure 3.1). Maltose is split into glucose and β G1P by a P_i -dependent reaction catalysed by maltose phosphorylase (Nilsson & Radstrom, 2001). The glucose formed then enters glycolysis via glucokinase while β G1P is converted to D-glucose-6-phosphate (G6P) by β PGM before entering glycolysis (Qian *et al.*, 1997). The maltose pathway is controlled by a transcriptional regulator, MalR, that regulates the expression of maltose transporters. Indeed, it regulates the maltose degradation pathway without affecting the main enzymes involved (Andersson & Radstrom, 2002b). In a similar way, trehalose is split into G6P and β G1P with same aims (Andersson & Radstrom, 2002a).

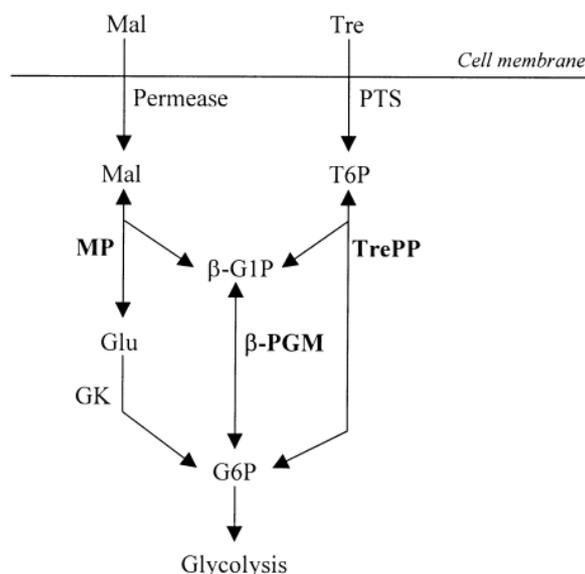


Figure 3.1 The closely connected maltose and trehalose metabolic pathways in *L. lactis*. Schematic view of the two pathways in *L. lactis*. Mal, maltose; Tre, trehalose; PTS, phosphotransferase system; MP, maltose phosphorylase; TrePP, trehalose 6-phosphate phosphorylase; β -PGM, β -phosphoglucomutase; GK, glucokinase; Glu, glucose; T6P, trehalose 6-phosphate; β -G1P, β -glucose 1-phosphate; G6P, glucose 6-phosphate (Andersson & Radstrom, 2002a).

3.1.3 Structural characterisation of β PGM

β PGM is a member of the haloacid dehalogenase (HAD) superfamily, which includes the sarcoplasmic Ca^{2+} -ATPase, phosphomannomutase and phosphoserine phosphatase. Indeed, the X-ray structure of β PGM shows a 2 domain α/β protein, the common fold shared with the entire HAD family. This fold is different from the α -phosphoglucomutase family, where α PGM (65KDa) is approximately twice the size of the β PGM (25 KDa) (Dai *et al.*, 1992) and the X-ray structure of α PGM reveals a 4-domain α/β protein. All four domains contribute residues to form a large active-site crevice (Lahiri *et al.*, 2002).

The first structure of β PGM was solved in the open conformation (with no sugar bound) at 2.3 Å resolution (PDB code 1LVH; Lahiri *et al.*, 2002). The protein was crystallised in its phosphorylated form (P-Asp8, β PGM^P). β PGM (221 residues in total) is a monomer that is composed of two distinct domains, a helical cap domain (residues 1-14 and 93-221) and the α/β domain (residues 15-92). Together, the core and the cap domain give rise to a kidney-bean shaped monomer, this suggested that the reaction is performed by conformational changes via cap domain closure (Figure 3.2). The

movement of the cap domain was visualised when the first closed structure was solved. β PGM has been crystallised in the closed conformation by two different groups. Karen Allen and co-workers crystallised it with bound a phosphorane intermediate (PDB code 1O03; Lahiri *et al.*, 2003). The second group demonstrated that the phosphorane group was in fact magnesium trifluoride, mimicking the transfer of the phosphate from the protein to the sugar molecule (Baxter *et al.*, 2010). Since the arguments reported in this and further papers (Webster, 2004), in this thesis only the closed structure with magnesium fluoride will be considered (PDB code 2WF5; Baxter *et al.*, 2010).

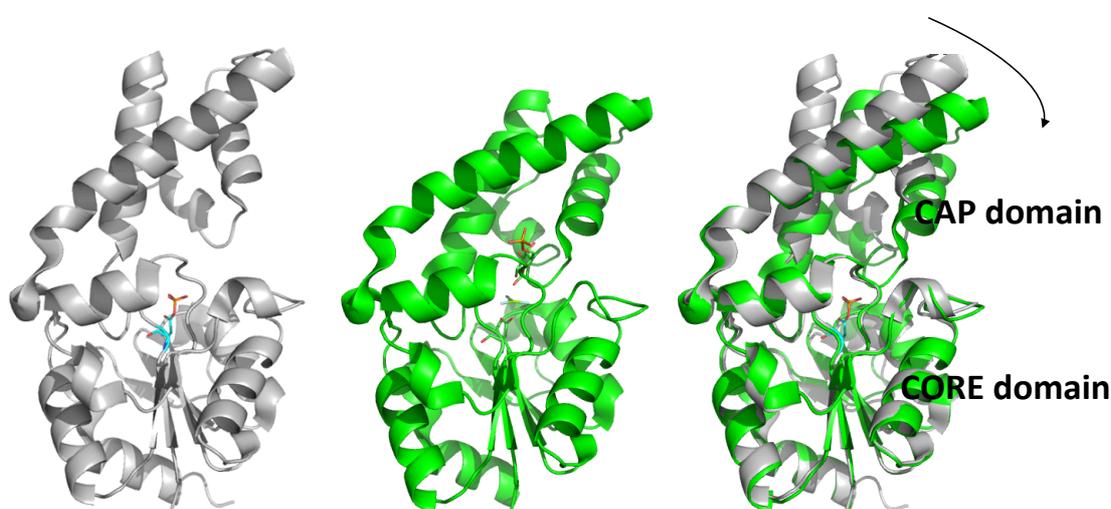


Figure 3.2 Ribbon representation of the cap domain movement. β PGM^P is in *grey*, with phosphorylated Asp8 represented in sticks (PDB code 1LVH; Lahiri *et al.*, 2002). β PGM-G6P-MgF₃⁻-TSA is in *green* (PDB code 2WF5; Baxter *et al.*, 2010). The arrow indicates the hypothetical movement direction.

The two structures showed that the active site of β PGM is located at the domain-domain interface. The exact location is identified by the phosphorylated side chain of Asp8 and by the Mg²⁺ cofactor. The active site is comprised of the cap domain side chains of Lys45 and the Thr16, and the side chains of Asp8, Asp10, Val12, Ser114, Lys145, Asp169 and Asp170 contributed by the core domain. Of these, Asp8 and Asp10, Ser114, Lys145, Asp169 and Asp170 are conserved among the HAD phosphotransferases. As further described in the closed conformation these residues are involved in magnesium coordination or in sugar binding. The magnesium is coordinated

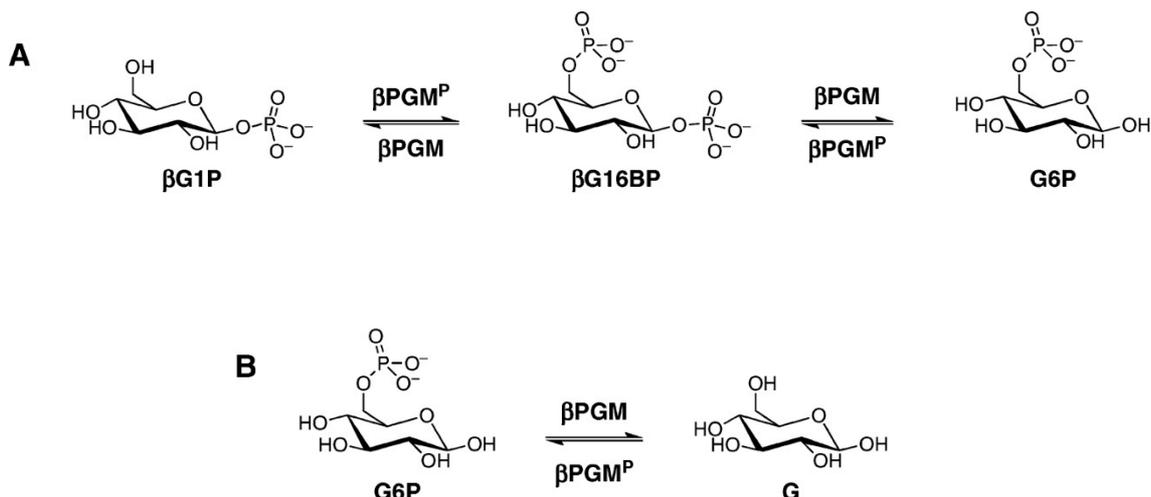


Figure 3.4 Reactions catalysed by β PGM. A) The mutase activity of β PGM catalyses the transfer of phosphate between the 1 and 6 positions of glucose. B) β PGM also catalyses the hydrolysis of G6P, the major species under equilibrium conditions, yielding glucose (Golicnik *et al.*, 2009).

The reaction proceeds through a general base catalysis system, which is comprised of two aspartates: Asp10 and Asp8. As shown by the structure of the transition state, Asp8 offers the nucleophile oxygen atom, while Asp10 is the proton acceptor or general base in the catalysis. The distances between the magnesium fluoride and the apical oxygen atoms are 2.09 and 2.18 Å, which defines the transition state as a tight one (Figure 3.5).

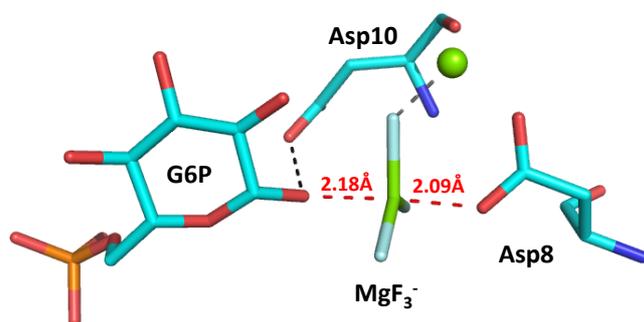


Figure 3.5 Residues involved in the General acid base catalysis of β PGM. The main characters of the reaction are shown, as described in the main text. The two aspartates involved are represented as stick. Magnesium is represented as *green* sphere.

3.1.4.1 β PGM activation

Presteady-state and steady-state kinetic analysis has shown that the amplitude of the initial burst is proportional to the concentration of β PGM^P. Depending on the cofactor that activates the protein, a lag period can be observed before a linear steady-state reaction is observed. This time is defined by the β PGM phosphorylation sources, such as β G16BP, α G16BP and α G1P.

β G16BP is the most efficient molecule in activating β PGM. Most phosphomutases require a sugar bisphosphate for optimal activity and in 1974 it was demonstrated that β G16BP is the one required for activating and maintaining the β PGM reaction (Marechal & Belocopitow, 1974). Indeed, no lag phase is observed when this molecule is used (Golicnik *et al.*, 2009). Furthermore, it was observed that a transient accumulation of [¹⁴C] β G16BP (12% at 0.1 s) occurs in a single turnover reaction performed with an excess of β PGM (40 μ M) and limiting [¹⁴C] β G1P (5 μ M) and β G16BP (5 μ M). This experiment supported the role of β G16BP as a reaction intermediate in the conversion of the β G1P to G6P (Dai *et al.*, 2006). In experimental conditions, a constant amount of this molecule remains in solution in order to generate new β PGM^P and any remaining after complete consumption of β G1P is slowly converted to G6P (Golicnik *et al.*, 2009).

Due to its general availability, α G16BP is used experimentally as enzyme activator. α G16BP requires further isomerisation or hydrolysis to be able to activate the protein. Thus, in this case where a lag period is required for starting properly the reaction, β PGM can isomerise the α anomer to the β , producing an amount of β G16BP necessary for maintaining the β PGM^P pool (Golicnik *et al.*, 2009).

Eventually, in a less efficient way, the protein can hydrolyse α G16BP to α G1P; α G1P as a further phosphate donor for β PGM (Golicnik *et al.*, 2009).

Due to the ability of water to hydrolyse P-Asp8, β PGM^P produced in this way has a half-life of 12 s. In this β PGM differs from α PGM, where a phospho-serine is involved in the reaction and the phosphoprotein is stable in water (Dai *et al.*, 2006; Zhang *et al.*, 2005).

3.1.4.2 Isomerisation induces through an intermediate

The ability of β PGM to convert both the monophosphosugars is related to the fact that β G16BP binds to the active site in two different orientations with roughly the same efficiency (Dai *et al.*, 2006). This means that β G16BP needs to be reoriented to produce either G6P or β G1P. For β PGM catalysis in the β G1P-forming direction, the β G16BP must be bound with the C(1)phosphate interacting with the cap domain and the C(6)phosphate with the Asp8. This orientation is termed E(β G16BP)-1. For β PGM catalysis in the G6P-forming direction, the β G16BP must be bound with the C(6)phosphate interacting with the cap domain and the C(1)phosphate with the Asp8. This orientation is termed E(β G16BP)-2. The reorientation of the bisphosphate has historically been studied in α PGM, where the α G16BP is reoriented by rotating 180° while still associated with the enzyme (Naught & Tipton, 2005). In the case of β PGM the same mechanism has been explored by kinetic methods and interpreted through the structure of the open and closed conformations. It has been postulated that the reorientation in the active site is achieved by dissociation into solvent and then binding in the opposite reorientation (Dai *et al.*, 2006). This is supported by the X-ray structure of the enzyme: in its catalytically active, fully closed conformation little room exists for ligand reorientation, while the apo-phosphorylated enzyme (Lahiri *et al.*, 2002) illustrates the open conformation in which the cap domain and core domain are dissociated and the active site is open to solvent (Baxter *et al.*, 2010; Lahiri *et al.*, 2002).

3.1.4.3 Possible inhibition of the β PGM reaction

In the pre-steady state the reaction can be inhibited in several ways. β G1P in either low or high concentrations decreases the reaction rate. Over a longer time scale, when the β G1P concentration is lower, water can compete with the glucose monophosphate species present to dephosphorylate β PGM^P and deplete the active components of the system. In the opposite direction, when β G1P is at high concentrations, substrate inhibition is observed. β G1P can bind to β PGM with the phosphate group in the distal binding site, presenting the 6-OH into the active site, thereby inhibiting the enzyme. Alternatively, if β G1P is bound with the phosphate group in the active site, it still

effectively forms an unproductive complex, as it only transfers the phosphoryl group to the enzyme very slowly relative to the other processes under these conditions (Golicnik *et al.*, 2009). Moreover, α G16BP can slow down the reaction rate, when the phosphate in position 1 is close to the β PGM active site.

All the steps described are summarised in figure 3.6. For each reaction k_{cat} and K_m are reported (Table 3.1). Eventually the limiting step of the reaction, is β PGM phosphorylation or activation.

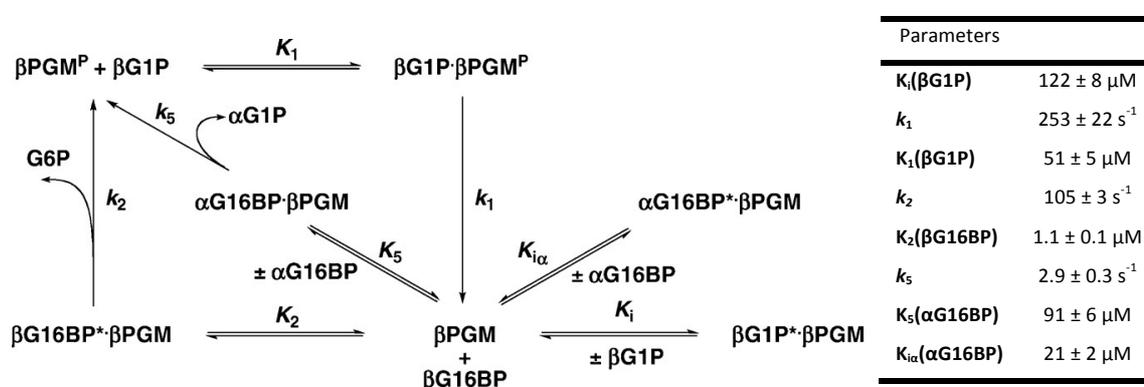


Figure 3.6-Table 3.1 Simplified reaction scheme of the reaction catalysed by β PGM in the presence of α G16BP as a cofactor. When the β G16BP intermediate binds to β PGM, either the 6-phosphate (β G16BP \cdot β PGM) or the 1-phosphate (β G16BP* \cdot β PGM) is placed at the phosphoryl transfer site. α G16BP can either bind productively (α G16BP \cdot β PGM, with the 6-phosphate at the transfer site) or unproductively (α G16BP* \cdot β PGM, with the 1-phosphate at the transfer site). The hydrolase activity is slow in comparison with the isomerase activity when α or β G16BP are present in the reaction solution as a cofactor. However, the removal of G6P by the coupled assay, used for measuring reaction parameters, is fast and irreversible (Golicnik *et al.*, 2009).

1.5 Challenges in studying the β PGM reaction

As described above, β PGM carries out a multistep reaction, making it a challenge to study. Of particular interest is how one active site can effectively catalyse a phosphoryl transfer reaction to two different substrates – this is notably the case with β PGM. X-ray crystallography is the most eligible technique that increases the reaction knowledge by providing reaction snapshots. Ideally, structures of β PGM in complex with each substrate (β G1P, β G6P and β G16BP) will provide a complete scenario for the whole

reaction. However, growing crystals of β PGM in complex with mono-phosphor sugar failed. This was because β G1P can be converted to G6P by activating β PGM, and G6P can slowly be converted to glucose and inorganic phosphate (Baxter *et al.*, 2006; Lahiri *et al.*, 2003). As described in the introduction of this thesis, the mechanism of phosphoryl transfer has been successfully studied by using metal fluoride compounds such as magnesium or aluminium fluoride. Fluoride inhibits phosphoryl transfer by blocking the reaction studied in the transition state, providing a chance to obtain a relevant reaction snapshot for understanding the transition state scenario. In the case of β PGM it has been demonstrated that increasing amounts of magnesium fluoride inhibit the hydrolase activity ($k_{\text{cat}} = 70 \text{ nM}$) and G6P is more tightly bound to β PGM (Golicnik *et al.*, 2009). Indeed, as described in Section 3.1.3, β PGM was successfully crystallised in complex with G6P and magnesium fluoride (Baxter *et al.*, 2010). Furthermore, ^{19}F -NMR has been demonstrated to be a tool for analysing the transition state formation and describing its energy environment in solution.

In this chapter and chapter 4, further details of this reaction will be analysed by the combination of ^{19}F -NMR and X-ray crystallography. Two main unsolved questions will be discussed. The first investigates the so called “*major-minor form*” of magnesium trifluoride. The second focus on the study of the first reaction step: the phosphorylation of β G1P. This second topic will be discussed in Chapter 4.

3.2 Aim of this project

The major-minor form refers to the particular NMR spectra recorded in the case of the β PGM-G6P-MgF₃⁻-TSA complex (PDB code 2WF5; Baxter *et al.*, 2010). The spectra showed that apparently β PGM could coordinate the magnesium fluoride molecule in two different ways: the major and the minor form, the names are related to the percentage of the protein population corresponding to each group. However, no evidence was observed for the minor conformation in the crystal structure, showing that the structure in solution could be different from the one in the solid state.

β PGM was mutated to remove one of the positively charged residues coordinating the transition state to a neutral one (β PGM_{K145A}). As expected from the charge balance hypothesis, the spectra collected on the β PGM_{K145A}-G6P-MgF₂-TSA recorded two peaks, instead of the usual three. Surprisingly, the chemical shifts of the two peaks were very similar to two of the minor form pattern. Thus, in order to characterise structurally both major and minor populations, it was decided to combine the new structural knowledge from the mutant with further methods that will allow the convergence of X-ray crystallographic with NMR data.

3.3 Experimental procedures

3.3.1 Cloning, protein expression and purification

Cloning, expression and purification experiments are described in the thesis of Dr. Joanna Griffin (Griffin, 2011).

3.3.2 Crystallisation of β PGM-G6P-MgF₃⁻-TSA and β PGM_{K145A}-G6P-MgF_x-TSA complexes

For crystallisation experiments, proteins were received purified in 50 mM HEPES pH 7.2, DTT 0.1 mM and MgCl₂ 5 mM and concentrated to 30 mg/ml. For crystallisation, the solution was adjusted to 15 mg/ml. 10 mM NH₄F and 6 mM G6P were added and the sample left for a few minutes on ice in order to promote complex formation. For crystallisation, the solution was mixed 1:1 with precipitant and placed in sitting drop at room temperature (RT), with a final volume of 4 μ l. Crystallisation conditions were slightly different for the complexes. In the case of the WT large plate crystals were obtained after one to two days in 27-32% (w/v) PEG 4000, 50-125 mM NaAc. In the case of the mutant, higher concentrations of PEG were required (30-36% (w/v)). As in the case of the mutant both closed (needles) and open conformation (plates) grew in the same conditions it was necessary to harvest the plate crystals in three days, before they were surrounded by needles (Figure 3.7).

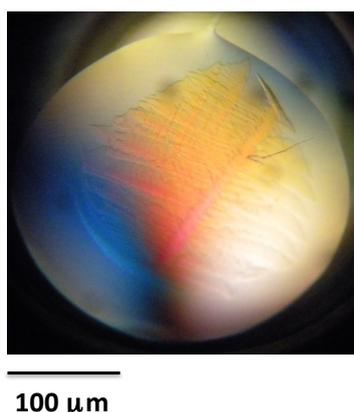


Figure 3.7 Crystal of β PGM_{K145A}-G6P-MgF_x-TSA

3.3.3 Data collection at 100 K, structure solution and refinement

All crystals were cryocooled as previously described (Pellegrini *et al.*, 2011). In the case of β PGM-G6P-MgF₃⁻-TSA, all data collection was performed at 100K. For structure analysis the structure deposited in the PDB as 2WF5 was used (Baxter *et al.*, 2010).

Data from β PGM_{K145A}-G6P-MgF_x-TSA crystals were collected to 1.6 - 1.4 Å resolution on an ADSC Q210 CCD detector on beamline ID14-2 at the European Synchrotron Radiation Facility, Grenoble, France (ESRF). Data were processed using XDS (Kabsch, 2010) and structures were solved by molecular replacement using MolRep (Vagin & Teplyakov, 2010). 2WF5 without ligands or water molecules was used as a search model. Refinement was carried out using REFMAC5 (Murshudov *et al.*, 1997) in combination with manual refinement in COOT (Emsley & Cowtan, 2004). Ligands were added in the last refinement cycles. Validation was performed with Phenix as described in Chapter 2 (Adams *et al.*, 2011).

3.3.4 Data collection at room temperature, structure solution and refinement

Data collection was performed at room temperature using the EMBL HC1b humidity control device. The HC1b allows dehydration experiments to be performed at synchrotron beam lines with minimal perturbation to the experiment configuration (Sanchez-Weatherby *et al.*, 2009). The experiment is thoroughly described in Russi *et al.*, 2011. Briefly, a typical dehydration experiment starts from the determination of the

relative humidity (RH) of the crystal mother liquor. This can be achieved by placing a drop of the mother liquor in a loop at a chosen RH and monitoring the changes in the defined region using a specific image processing software. No changes in size of the point of interest indicate the start RH. Then the crystal is mounted and images taken to evaluate the crystal quality.

To alleviate the problems of radiation damage a helical data collection protocol was used (Flot *et al.*, 2010). This allowed continuous exposure of fresh crystal volumes during data collection. Eventually, a data set at 1.47 Å resolution was collected with excellent data processing as statistics. Data processing was carried out as described in section 3.3.3 and Chapter 2.

3.4 Results and Discussion

3.4.1 ^{19}F -NMR spectra revealed an alternative way of coordinating magnesium fluoride in the $\beta\text{PGM-G6P-MgF}_3^-$ -TSA complex

In the case of the $\beta\text{PGM-G6P-MgF}_3^-$ -TSA complex (Baxter *et al.*, 2010), it has been demonstrated that the combination of solution NMR and high resolution X-ray crystallography can provide an independent validation of MgF-TSA complex formation. Indeed, it was shown that there is a remarkable relationship between NMR chemical shifts, primary isotope shifts, NOEs, hydrogen bond $\text{F}\cdots\text{H-X}$ scalar couplings, and the atomic positions determined from the high-resolution crystal structure. These data provided proof in solid and solution states of the ability of magnesium fluoride, rather than a pentacoordinated phosphorane intermediate, to trap the transition state and to be a sensitive probe of the electrostatic and hydrogen bonding distributions in a near-TS conformation. The first key probe of the interactions within the TSA was provided by ^{19}F -NMR. The spectra recorded in water and D_2O show three distinct peaks, each one corresponding to a different fluoride atom coordinating the Mg atom (F_A , F_B and F_C). The high resolution crystal structure of $\beta\text{PGM-G6P-MgF}_3^-$ -TSA (1.3 Å) showed that the magnesium is pentacoordinated in trigonal bipyramidal geometry with two oxygen atoms (protein D8-O and sugar C1-O) at the apical positions; while on the equatorial plane, the magnesium atom is coordinated by the three fluorides (bond lengths: Mg-F_A = 1.8 Å, Mg-F_B = 1.8 Å Mg-F_C = 1.9 Å) (PDB code 2WF5; Baxter *et al.*, 2010). Furthermore, isotope shifts combined with the structure characterisation provide information about the proton coordination of each atom. The three fluorine atoms in each of the MgF_3^- moieties are hydrogen bonded to multiple exchangeable donors of the protein, and a comparison of the spectra recorded in H_2O buffer and D_2O buffer allows the sum of the individual isotope shifts to be measured. Three protons, Leu9H^N , Asp10H^N , and Ser114H^O coordinate F_A , in a distorted tetrahedral arrangement, giving a sum isotope shift of 1.6 ppm. F_B and F_C have trigonal coordination involving two protons (Ala115H^N and Lys145H^N) and one proton (G6P-2'H^O), respectively, and have correspondingly smaller sum isotope shifts (1.4 ppm and 0.9 ppm, respectively) (Figure 3.8).

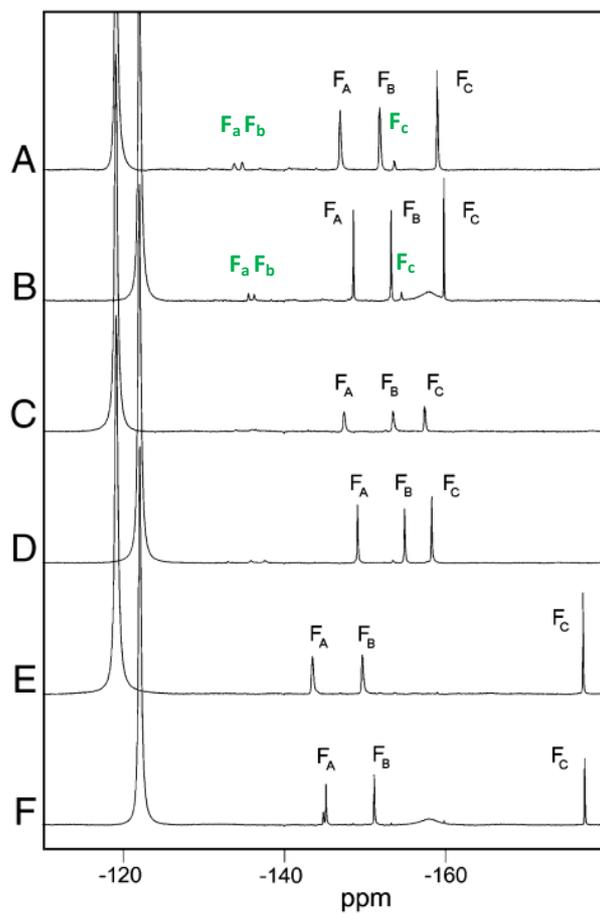


Figure 3.8 ^{19}F -NMR spectra of $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ complex. Spectra were recorded at 20°C in 50 mM K HEPES buffer at pH 7.2, in 100% H_2O or in 100% D_2O . Chemical shifts are given in ppm for each ^{19}F resonance in the complex. A) $\beta\text{PGM-MgF}_3\text{-G6P-TSA}$ in 100% H_2O buffer ($F_A = -147.0$, $F_B = -151.8$, $F_C = -159.0$). B) $\beta\text{PGM-MgF}_3\text{-G6P-TSA}$ in 100% D_2O buffer ($F_A = -148.6$, $F_B = -153.3$, $F_C = -159.8$). C-F) Not of interest for the topic discussed. The peaks indicated by *green* letters are part of the minor form pattern (Baxter *et al.*, 2010).

X-ray crystallography experiments failed in the assignment of all the peaks show in the NMR data. In the panels A-B further small peaks are present. Surprisingly, there are three of them and they are always present in $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ solution. The respective ppm are: $F_a = 133.8$, $F_b = 134.8$, $F_c = 153.6$ (in green in figure 3.8). Previous experiments have shown that the three peaks are related to each other. Indeed, repeating the same experiment in different conditions, such as different temperatures, the three peaks increased or decreased simultaneously. It was also reported that this minor conformer most likely exchanges with the major conformer more rapidly than the complex dissociates because the resonances were found to correlate via saturation

transfer with specific resonances of the major conformer (Baxter *et al.*, 2006; Griffin, 2011). These experiments, described in Dr. J Griffin's thesis, clearly show that in the sample two populations are present and each one is related to different pattern of peaks. Since in the first experiment the F_a , F_b , F_c peaks were observed at a low percentage of the entire population, they were termed the **minor form**; the F_A , F_B , F_C peaks were defined as the **major form**.

As the two populations should be different in terms of structure, two hypotheses were formulated. First, it was hypothesised that the protein was able to coordinate the magnesium fluoride using different residues from those described previously. No signals in the difference electron density map suggest a possible alternative conformation of the active site, which could accommodate the magnesium trifluoride differently in TBP geometry.

Second, it was hypothesised that these three peaks could indicate an alternative geometry of the coordinated magnesium, even if no proof supporting this hypothesis has been found in the literature. In vacuum and gas phase, the magnesium fluoride geometry depends on the number of fluoride atoms coordinating it or by the presence of solvent. Mg-F bonds are weakened and lengthened as the number of coordinating fluorine atoms is increased (Shibata *et al.*, 2011). All fluoride atoms tend to keep away from each other to reduce strong F-F Coulomb repulsion, and the bond angle is thus understandable in terms of the VSEPR model (Valence Shell Electron Pair Repulsion). In the case of magnesium trifluoride the averaged Mg-F bond lengths and F-Mg-F angles are respectively $1.845 \pm 0.04 \text{ \AA}$ and 119.90 ± 0.1 (intermediate values between the gas phase and aqueous solution) conferring trigonal planar geometry. In the case where solvent is present, the bond angles are significantly bent, without affecting the trigonal geometry (Shibata *et al.*, 2011) (Figure 3.9).

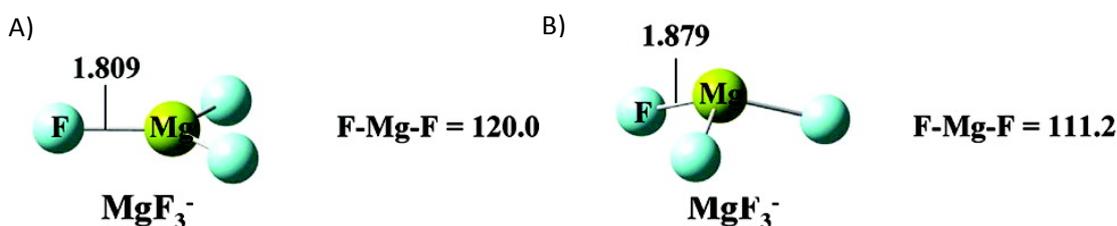


Figure 3.9 Optimised geometry of MgF_3^- in the gas phase (A) and in aqueous solution (B). Angles and bond lengths are given in degrees and Å (Shibata *et al.*, 2011).

Similarly, in crystal structures the bond lengths are 1.880 ± 0.15 Å and 108.80 ± 0.6 , which are much closer to the values in the gas phase (Shibata *et al.*, 2011). Furthermore, in the electron density of $\beta\text{PGM-G6P-MgF}_3^-$ -TSA there was no clear evidence supporting an alternative geometry. Thus further investigations were required.

3.4.2 The $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_x\text{-TSA}$ structure is related to the minor form

βPGM was mutated to remove one of the positively charged residues coordinating the transition state to a neutral one. Lys145 was mutated to an alanine, and the protein mixed with G6P and magnesium fluoride in order to observe any changes in the transition state analogue. Based on the charge balance hypothesis, the loss of a positive charge should bring the loss of a negative charge in the bound substrate in order to satisfy the zero charge required in the active site (Baxter *et al.*, 2008; Cliff *et al.*, 2010). In the case of PGK, it has been found that one of the fluoride atoms of AlF_4^- was substituted by a water molecule, without any change in the protein structure or in the metal fluoride features (Cliff *et al.*, 2010). The ^{19}F -NMR experiment on the K145A mutant of βPGM recorded two peaks, instead of the usual three. This indicated that the complex is still formed and that the magnesium atom is coordinated by two fluoride atoms instead of three, satisfying the charge balance in the active site. Surprisingly, the chemical shifts of the two peaks were very similar to two of the minor form pattern (F_b and F_c in figure 3.10, table 3.2), suggesting a structural relationship between the mutant and the minor form structure. Therefore crystallisation attempts of the mutant complex were performed.

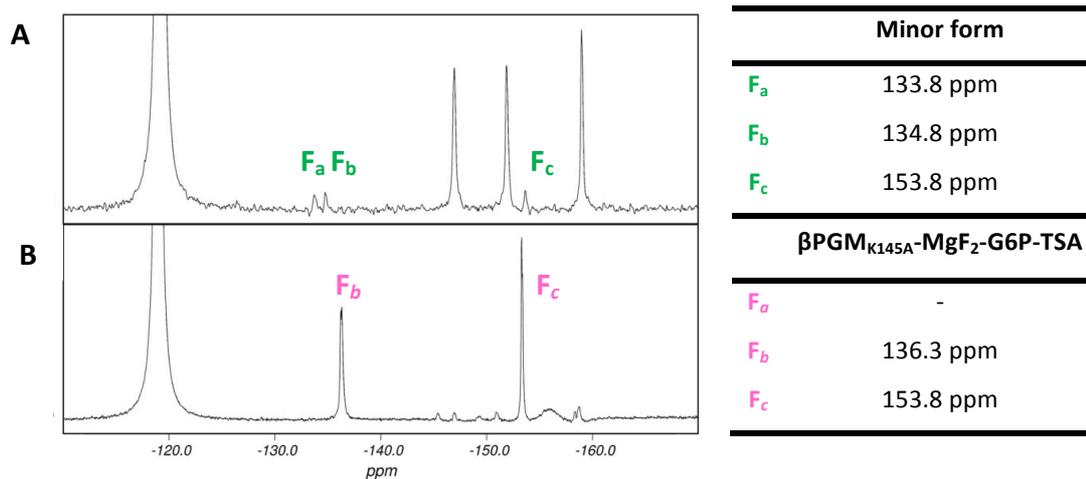


Figure 3.10-Table 3.2 ^{19}F -NMR spectra of $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ and $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_2\text{-TSA}$ complexes. Spectra were recorded at 20°C in 50 mM K HEPES buffer at pH 7.2 in 100% H_2O . In *green* the peaks/values of chemical shift of the minor form are shown. In *pink* the ones about the mutant complex are shown.

The mutant complex was crystallised successfully, in the same space group as the published wild type structure ($P2_12_12_1$). The asymmetric unit contained one molecule and the structure was solved to 1.4 Å resolution (Table 3.3). Macroscopically, the structure is very similar to $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ structure with an rmsd between Ca atoms of 0.26 Å (calculated with *Superpose*, Krissinel & Henrick, 2004). The difference Fourier map showed clear peaks for the expected ligands and the absence of electron density for Lys145, confirming that the protein crystallised was indeed the mutant.

Table 3.3 Data collection and refinement statistics

PGM_{K145A}-MgF₂(H₂O)-G6P-TSA	
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.933 Å
Unit cell dimension (Å) a,b,c	37.41,54.17,104.48
Resolution range (Å)¹	20(1.48-1.4)
Number of unique reflections	42,671
Multiplicity	6.4(6.4)
Completeness (%)	99.9 (100)
R_{merge}	0.075 (0.631)
<I/σ(I)>	19.0 (3.0)
Wilson B factor (Å²)	7.76
Water molecules	306
R_{factor} (%)	15.75
Free R_{factor} (%)	19.23
RMS deviations:	
Bonds (Å)	0.019
Angles (°)	1.785

¹Values for the higher resolution shell are in brackets

The most striking difference between the WT and the mutant was in the coordination of the magnesium ion in the metal fluoride. The density of mutant clearly showed octahedral coordination of the metal. This confirmed the NMR data, indicating that the mutation causes a change in the magnesium fluoride molecule. Due to the high resolution, it was possible to distinguish between peaks related to fluorine atoms and water molecules. The coordination distances between water and magnesium and that of fluoride are different (2.0 vs 1.8 Å); electron density peaks that can be assigned to fluorine atoms are usually continuous with density for the metal ion. In the case of water coordination this continuity is abolished; this difference can only be discerned at high resolution. As shown in Figure 3.11 the electron density corresponding to the magnesium fluoride was easily assigned to a magnesium difluoride, coordinated by an additional water molecule (W).

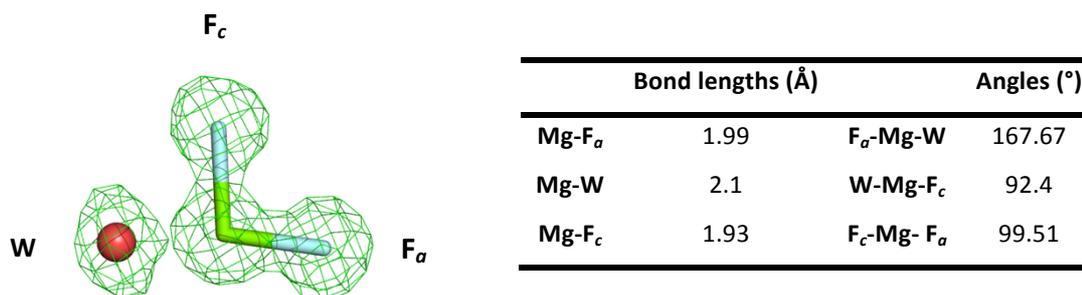


Figure 3.11-Table 3.4 Difference Fourier map of the magnesium fluoride molecule. The magnesium fluoride molecule is shown in sticks. The difference map, contoured at 5σ , is shown as *green* mesh. The table shows bond lengths and angles.

While the apical coordination has been conserved, the magnesium is now coordinated in octahedral geometry with the water molecule and two fluoride atoms symmetrically arranged around it. The other two vertices of the octahedron are occupied by the second fluoride molecule and by the carboxyl oxygen of Ser114 (Figure 3.12).

The fact that a protein can coordinate a magnesium fluoride species in octahedral geometry is a novelty. The fluoride-magnesium bonds are longer with regard to the

trifluoride magnesate. This has also been described for MgF_4^- in an aqueous solution, where the average bond length is 1.93 Å (Shibata *et al.*, 2011).

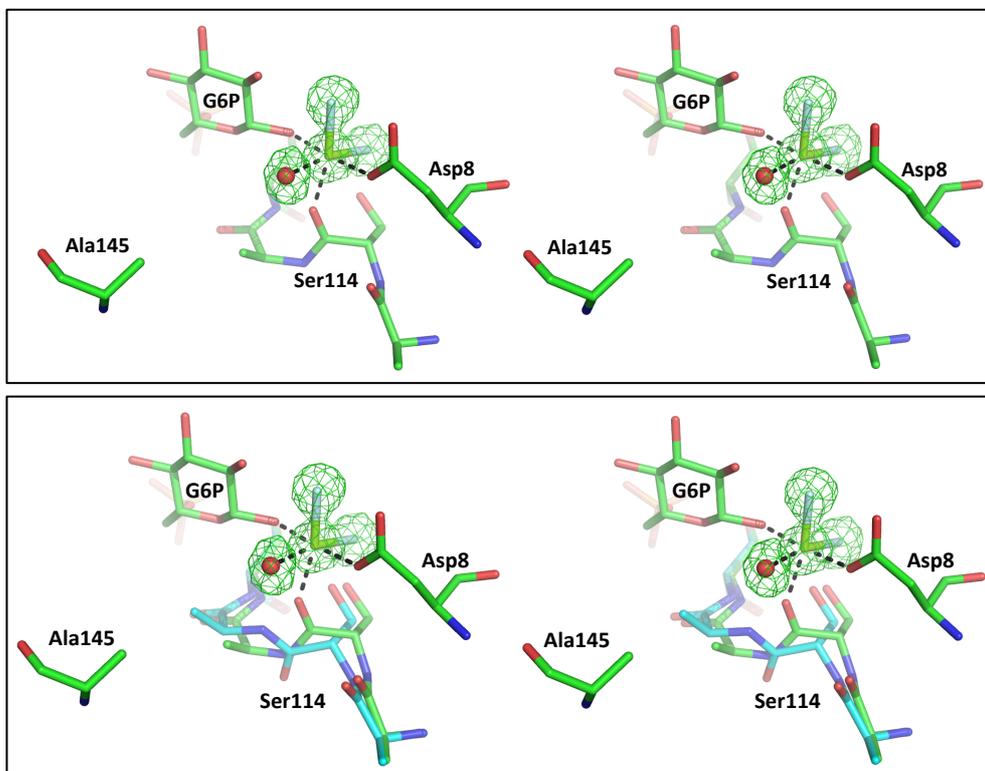


Figure 3.12 Magnesium is coordinated in octahedral geometry. Stereo-views of the molecules and residues in $\beta\text{PGM}_{\text{K145A}}$ involved in the octahedral coordination. On the bottom panel the overlay of $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_2(\text{H}_2\text{O})\text{-TSA}$ (in *green*) with $\beta\text{PGM-G6P-MgF}_3^-\text{-TSA}$ (in *cyan*) highlight the conformational change of Ser114 backbone for accommodating the octahedral specie.

The ability of the protein to accommodate this new species depends on the plasticity of the active site. This is clearly shown by the overlay between βPGM and $\beta\text{PGM}_{\text{K145A}}$ (Figure 3.12). The Ser114 backbone has flipped in order to coordinate the magnesium fluoride species to compensate for the loss of a fluoride due to the lack of Lys145. The Ser114 backbone is not only the 4th vertex of the octahedron, but is also coordinated to the water molecule coordinating the magnesium difluoride. Moreover, the interaction between the fluoride magnesium and the amine group of Ala115 has been substituted by two hydrogen bonds with further water molecules. These waters are absent in the $\beta\text{PGM-G6P-MgF}_3^-\text{-TSA}$ structure (Figure 3.13).

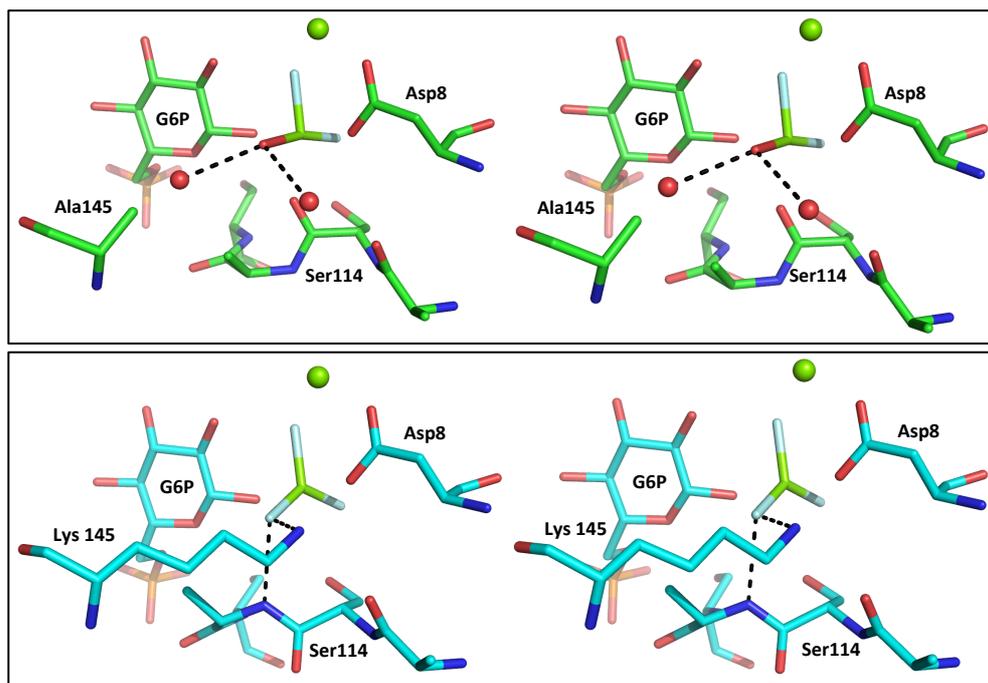


Figure 3.13 Stereo-view of the coordination of W-F_B. The hydrogen bonds are represented in black dashed lines. On the top panel $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_2(\text{H}_2\text{O})\text{-TSA}$ is in *green*, in the bottom panel $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ is in *cyan*.

In conclusion, the substitution of Lys145 with an alanine had three main consequences: a loss in charge of the magnesium fluoride, a rearrangement of the TSA species from TBP to octahedral geometry and protein conformational changes useful in the accommodation the new species.

Due to the correlation between the NMR spectrum of the minor form and mutant complex, it was proposed that the structural features observed could give structural information on the minor form. It was hypothesised that the minor form was related to the ability of βPGM to coordinate a magnesium fluoride molecule in octahedral geometry instead of the usual TBP.

First, the difference Fourier map of the $\beta\text{PGM-G6P-MgF}_3\text{-TSA}$ structure was inspected for features indicating alternative coordination. An overlay of the two magnesium fluoride molecules with the electron density map, highlights that the electron density corresponding to F_B or the water molecule has an elongated form, in which both molecules could be accommodated (Figure 3.14). Since the high resolution of the

structure, this detail had been assumed to be an indication of the minor form presence in the solid state. However, further X-ray data were required.

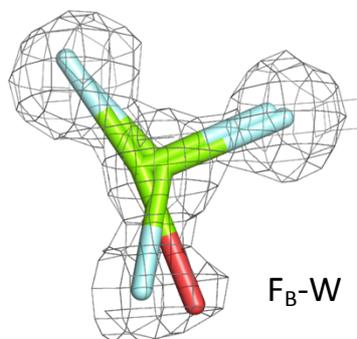


Figure 3.14 Overlay of the two magnesium fluoride structures on β PGM-G6P-MgF₃⁻-TSA Fourier map. The two magnesium fluoride molecules MgF₃⁻ from β PGM-G6P-MgF₃⁻-TSA and MgF₂(H₂O) from β PGM_{K145A}-G6P-MgF₂(H₂O)-TSA are represented in sticks. The Fourier map from β PGM-G6P-MgF₃⁻-TSA contoured at 3σ is shown as *grey* mesh.

3.4.3 Structure solved from room temperature data collection reveals the structure features of the minor form conformer

One of the main differences between the solid state and solution state is the amount of conformations a protein can assume. In solution, the mobility of protein regions is quite high, while in a crystal the protein is in an almost unique conformation due to crystal packing interactions. One of the factors affecting this phenomenon is the temperature. NMR experiments were run at 298K, while X-ray data collection was performed at 100K. The absence of alternative conformations, such as the minor form, could be caused by this difference.

Therefore it was decided to perform room temperature data collection, using the humidity control device HC1b on ID14-1, at the ESRF (Russi *et al.*, 2011). This system provides the possibility to collect a full dataset at a temperature around 293K, a temperature more similar to the NMR conditions. Combining dehydration with helical data collection (Flot *et al.*, 2010), a full dataset has been acquired.

The structure was successfully solved at 1.47 Å using as model the β PGM -G6P-MgF₃⁻-TSA structure. The rmsd of the C α atoms was 0.18 Å, indicating that the RT data collection did not affect the protein macro-conformation. The electron density for the

magnesium fluoride in TBP geometry was clearly evident. Surprisingly, in the difference Fourier map a positive peak of 3σ was present close to the Ser114 backbone and one more to the F_B electron density. This was a clear indication that the room temperature data collection affected the active site conformation by allowing a low percentage of the protein population to assume a conformation similar to the mutant complex (Figure 3.15).

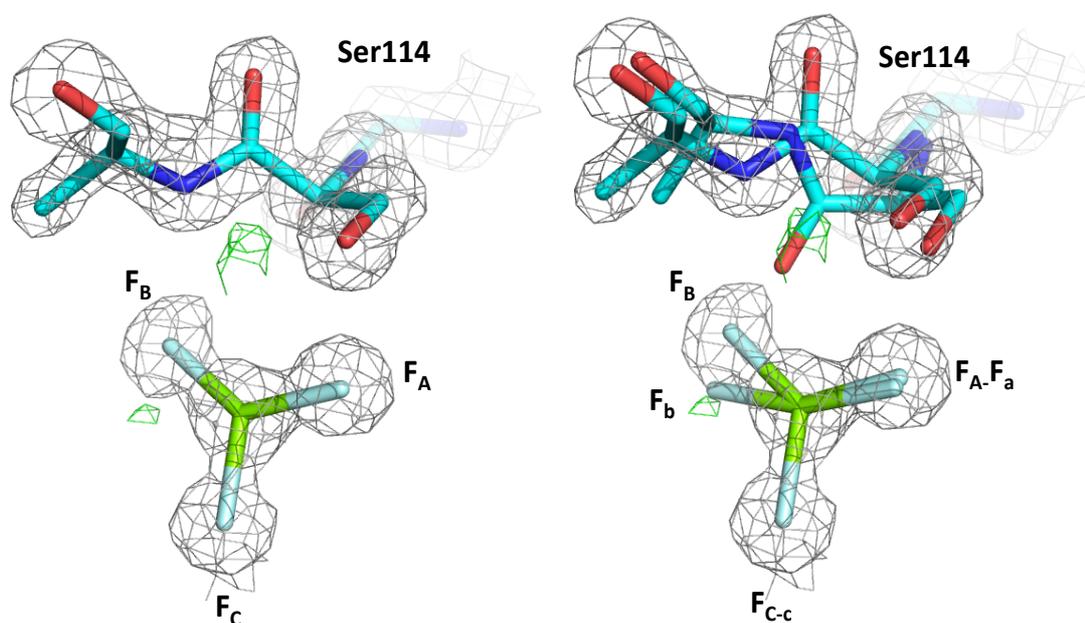


Figure 3.15 View of β PGM-TSA structure collected at room temperature. Overlay of the room temperature structure on the Fourier (grey mesh, countered level 2σ) and difference Fourier map (green mesh, countered at 3σ). On the left panel, the room temperature structure is shown before modelling the minor form. On the right panel, the final model is shown.

Thus, using the coordination of the octahedral magnesium fluoride molecule for the mutant and the Ala113-Ser114-Ala115 coordinates, an alternative conformer was modelled in the room temperature structure. As it was proposed that this alternative conformer was the minor form of the magnesium fluoride, instead of the water molecule a third fluoride atom was modelled in this position. The refinement statistics increased in quality, supporting the hypothesis that the peaks found were not noise, but a real signal coming from an alternative conformation (Table 3.5).

Table 3.5 Data collection and refinement statistics

	β PGM-G6P-MgF ₃ ⁻ TSA _{100K} (Baxter <i>et al.</i> , 2010)	β PGM _{K145A} -G6P- MgF ₂ (H ₂ O)-TSA	β PGM-G6P-MgF ₃ ⁻ TSA _{RT}
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.933	0.933	0.933
Temperature (K)	100	100	295
Unit cell dimension (Å) a,b,c	37.6, 54.6, 105.2	37.41, 54.17, 104.48	38.0, 54.6, 105.5
Resolution range (Å)¹	20(1.37-1.3)	20 (1.47-1.4)	24.3 (1.55-1.47)
Number of unique reflections	45,520	42,671	36,783
Multiplicity	2.5 (1.8)	6.4(6.4)	2.3 (2.2)
Completeness (%)	84.8 (47.2)	99.9 (100)	95.1 (94.3)
R_{merge}	0.06 (0.24)	0.075 (0.631)	0.052 (0.198)
<I/σ(I)>	8.5 (2.7)	19 (3)	12.1 (4.9)
Wilson B factor (Å²)	11.7	7.76	12.55
Water molecules	202	306	159
R_{factor} (%)	17.3	15.75	11.34
Free R_{factor} (%)	19.2	19.23	14.78
RMS deviations:			
Bonds (Å)	0.011	0.019	0.01
Angles (°)	1.501	1.785	1.39

¹Values for higher resolution shell are in brackets

The alternative conformer or minor form was modelled at 20% occupancy, with a trifluoride magnesium molecule coordinated octahedrally. The final geometry characterisation of the octahedral magnesium fluoride and trigonal bipyramidal species

are listed in table 3.6. In the case of the octahedral molecule, the bond lengths are shorter than those in $\beta\text{PGM}_{\text{K145A}}$. This was depending on the fact the magnesium fluoride molecule has been considered one entire molecule and it has been refined against dictionary containing specific restraints. Mg-F_a and Mg-F_b are shorter with regard to the $\beta\text{PGM-G6P-MgF}_3^-$ -TSA structure, while F_c has conserved the same position as F_C . However, the differences can be dependent on the fact that at 100K we cannot distinguish clearly between the two forms.

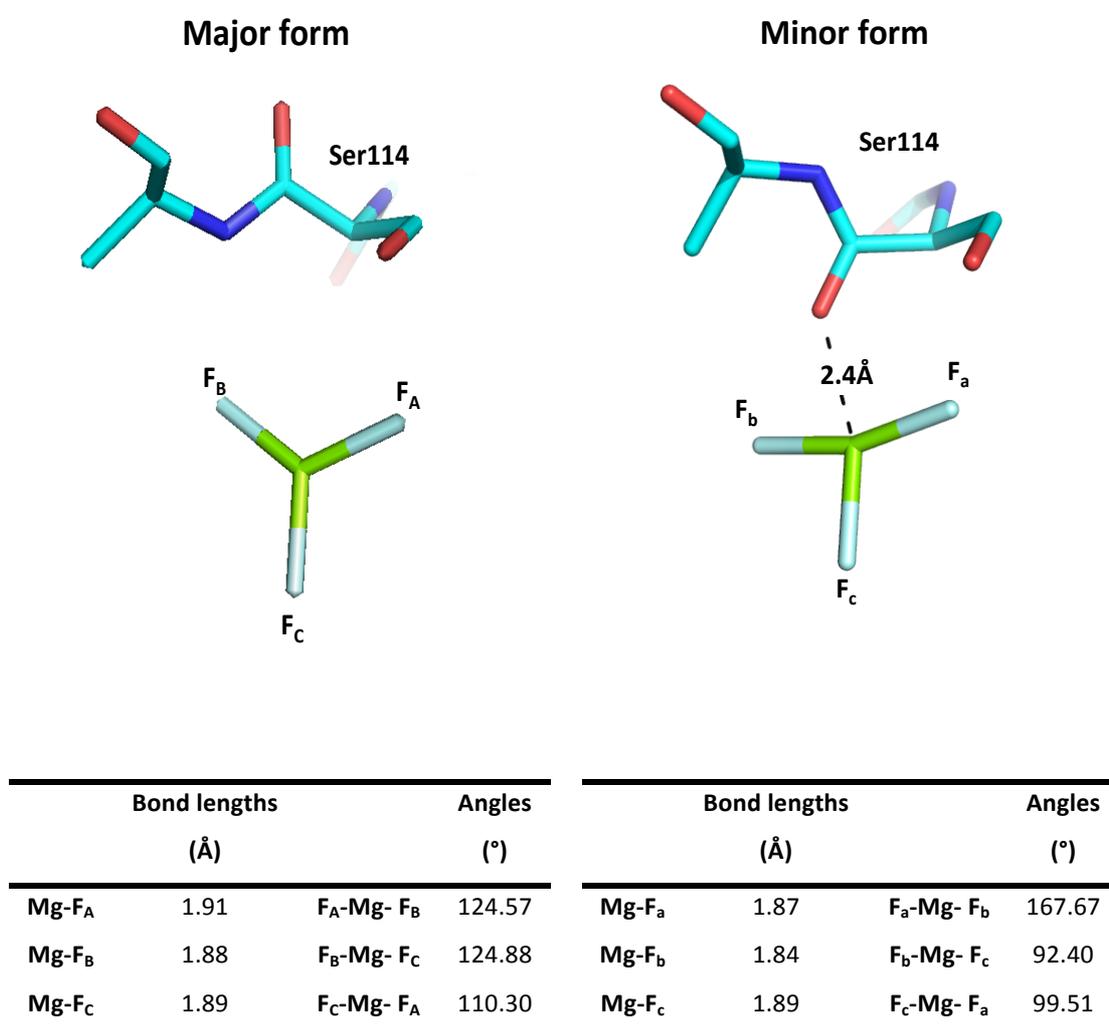


Figure 3.16-Table 3.6 Geometric details of the Major and Minor form. The minor and major form geometric features obtained from the RT structure, are shown.

3.5 Conclusions

The structures presented here define the structure of the minor form of MgF_3^- observed by ^{19}F -NMR in solution. In the minor conformer, the magnesium is still coordinated by three fluoride ions, but the coordination changes from TBP to octahedral and the backbone of Ser114 contributes to this coordination. The major/minor conformation data provide us with two major insights into the β PGM active site. The first is the novelty of the possibility for an active site to coordinate magnesium trifluoride in octahedral geometry. This opens the intriguing question of whether phosphorus could be similarly coordinated. The second is the plasticity of the A113-A115 protein region. Further experiments are required in order to understand the extent of this plasticity and how it is related to environment conditions. For example, it could be investigated whether such protein dynamicity is involved on the reorientation of β G16BP, a mechanism still not understood. The backbone flipping could be stabilised in one conformation by further mutations. Mutation of Ser114 or closer residues (such as Ala 115) to proline could decrease the backbone flexibility and lead to mutants, which bind preferentially β G16BP in one specific orientation. For future drug design projects, the results of these investigations will affect the design of a molecule. A practical example is discussed in the next chapter.

In conclusion, through the successful combination of ^{19}F -NMR and room temperature crystallographic experiments, new features of the β PGM active site have been determined. Moreover, this work is an excellent example of how two important structural biology techniques, in this case solution NMR and X-ray crystallography, are complementary to each other. The combined methods applied in this study could offer the possibility to study similar cases, in which alternative protein conformations are not visible at the usual 100K data collection (Lang *et al.*, 2010).

Chapter 4 Discrete analysis of both steps of the β PGM reaction using phosphonate analogues of Glucose 1- β -Phosphate

4.1 Introduction

4.1.1 Challenges in trapping β PGM- β G1P-TSA complex

Solid state attempts to trap a β PGM- β G1P-TSA complex failed when β G1P was used. This is due to the ability of β G1P to phosphorylate the protein, leading to the conversion of β G1P into G6P and resulting in the formation of a β PGM-G6P-TSA complex.

Searching for a method to successfully trap the β G1P complex is relevant for two reasons. First, the structure of β PGM- β G1P-TSA will provide a more complete understanding of the mutase enzyme mechanism. β PGM in a closed conformation bound to β G1P was only observed as a component of the dual occupancy β PGM-G6P-BeF₃ complex, where the complex is in a near attack conformation (Griffin *et al.*, 2012). However, this structure did not provide novel information into the catalysis in the phosphoryl transfer reaction on β G1P, as neither of the two β G1P conformers observed presents the 6-OH in line with the beryllium surrogate for the phosphorus centre of the Asp8 phosphate (Figure 4.1).

Secondly, a β PGM- β G1P-TSA complex will provide structural and chemical information, which are required for successful drug design.

To achieve both goals, it was decided to generate non-hydrolysable β G1P analogues by chemical synthesis. Since the design of these molecules called on the principle chemistry of α -fluorophosphonates, the following section will provide an overview of the use of phosphonates and α -fluorophosphonates, and the role of fluorine in medicinal chemistry.

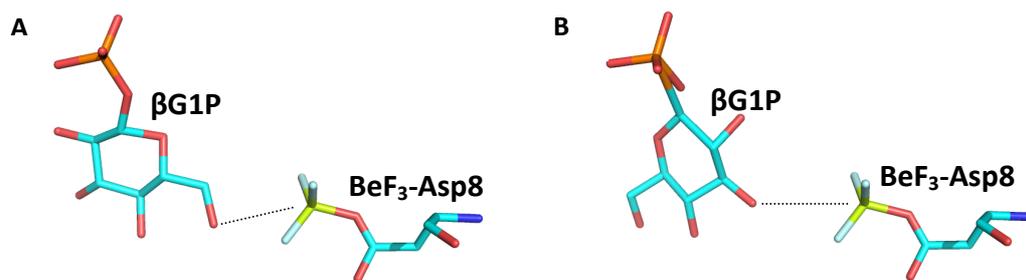


Figure 4.1 Structure of G1P in the β PGM-G6P-BeF₃⁻ active site. Structural details of the two β G1P conformers are shown. A) β G1P bound with the 6-OH close to the BeF₃⁻ complex. B) β G1P bound with the 3-OH close to the BeF₃⁻ complex. The donor-acceptor distance shown in round dots is 5.1 Å in each figure. β G1P was not added when the protein sample was prepared for these crystallization trials. In fact, the β G1P molecule observed results from significant mutase activity that persisted in the β PGM-G6P-BeF₃ complex (Griffin *et al.*, 2012).

4.1.2 α -fluorophosphonates and fluorine in medicinal chemistry

As in the case of β PGM, several enzymes catalyse the hydrolysis or the transfer of a phosphate group from a metabolite. Aside from prodrug applications, the production of non-hydrolysable metabolite, which has the same, or a higher, binding affinity as the natural substrate for the enzyme of interest, has become a key step in drug design and development. In the field of synthesis of phosphate ester analogues, the potential of phosphonates as phosphate mimics has been recognised for many years in drug development (Engel, 1977). Unlike a phosphate group, the phosphonate linkage is not readily hydrolysed in a biological environment, and this unique property has made these phosphate compounds attractive as phosphate analogues in numerous applications. Moreover, C-C-P and C-O-P bonds conserve a structural correspondence, which does not affect protein binding. However C-C-P and C-O-P, are significantly dissimilar in their chemical characteristics. Much more similar chemical characteristics have been achieved with the arrival of α -fluorophosphonate compounds, where the CH₂ group is substituted by CHF and CF₂ groups. CHF and CF₂ groups can both sterically and electronically mimic an oxygen atom, because the second dissociation protonation constant pK_a is much more similar to the natural phosphate. Indeed fluorine substitution on the phosphonate methylene group has been proved to lower the pK_a of phosphonate (Blackburn, 1981). This is particularly true for CHF, in which the second protonation

constant pK_a is 6.5, while the one for the corresponding phosphate is 6.4 (Romanenko & Kukhar, 2006). Moreover, since fluorine is a small atom (Van der Waals radius of 1.47 Å versus hydrogen atom Van der Waals radius of 1.2 Å, Bondi, 1964) the replacement of an oxidisable C-H group by a C-F group doesn't usually affect the binding affinity to the target (Barnette, 1984; Clader, 2004; Penning *et al.*, 1997; Van Heek *et al.*, 1997). However, fluorine has high electronegativity (Pauling, 1960), which could change the chemical properties of the molecule.

These fluorine features (size and electronegativity) are considered an advantage in medicinal chemistry. Fluorinated compounds are synthesised on a routine basis in pharmaceutical research (OHagan & Rzepa, 1997; Park *et al.*, 2001; Smart, 2001), with the remarkable result that 20 to 30% of all drugs on the market have fluorine in them (Muller *et al.*, 2007).

Usually, a carbon-bound fluorine, such as in the α -fluorophosphonates, is selected for introduction into a designed small molecule. This modification will alter the potential potency of the drug by altering pharmaceutical properties such as bioavailability and chemical biochemical properties such as protein target binding affinity.

4.1.2.1 How does a fluorine atom affect bioavailability?

Bioavailability is the percentage of the dose reaching the circulatory system and is denoted **F**. A drug's potency is directly proportional to its bioavailability, which depends on several factors such as metabolic stability and compound lipophilicity.

Metabolic stability is one of the key factors in determining the bioavailability of a compound. The introduction of a fluorine atom can block the rapid oxidative metabolism by liver enzymes (in particular the P450 cytochromes) (Figure 4.2). This is due in part to the high electronegativity of the atom, which increases the molecule's polarity.

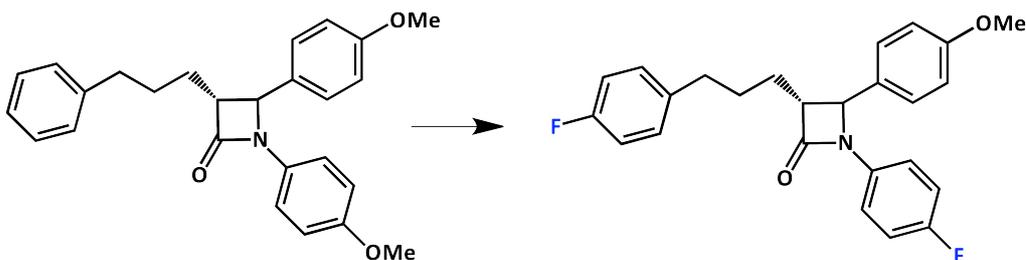


Figure 4.2 Development of Ezetimibe by optimization of a lead compound. An example of replacement of an oxidisable C-H group by a C-F group is the discovery of the cholesterol-absorption inhibitor Ezetimibe. The original compound contained two methoxy groups, which usually undergo metabolic demethylation. The introduction of a fluorine atom to replace one methoxy group and fluorination of the unsubstituted benzene ring prevented oxidation of the benzene rings to phenols. These changes increased the drug potency 400 fold, and enabled the use of lower doses due to improved metabolic stability *in vivo* (Clader, 2004; Van Heek *et al.*, 1997).

Bioavailability of orally administered drugs is also negatively affected by first-pass metabolism, poor absorption and inefficiency resulting from passage through the cell membrane (Purser *et al.*, 2008). Higher F values can be achieved by modifying the pK_a of the molecule of interest. As an electronegative atom, fluorine can change the physicochemical properties of a compound and a pK_a shift of several log units can be observed (Bohm *et al.*, 2004) (Table 4.1). This can affect both absorption, such as the stomach level (Figure 4.3), and lipophilicity (Bøhm & Schneider, 2003).

Table 4.1 Effect of fluorine substitution on pK_a and pK_b values

Carboxylic acid	pK_a	Alcohol	pK_a	Amine	pK_b
CH_3CO_2H	4.76	CH_3CH_2OH	15.9	$CH_3CH_2NH_2$	10.6
CH_2FCO_2H	2.59	CF_3CH_2OH	12.4	$CF_3CH_2NH_2$	5.7
CHF_2CO_2H	1.34	$(CH_3)_3COH$	19.2	$C_6H_5NH_2$	4.6
CF_3CO_2H	0.52	$(CF_3)_3COH$	5.1	$C_6F_5NH_2$	-0.36

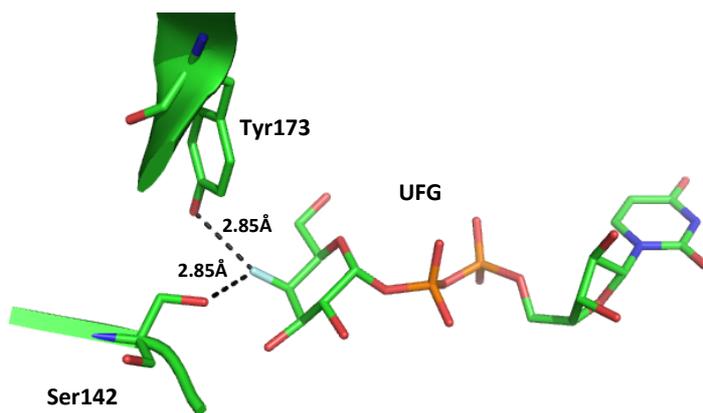


Figure 4.4 Structure of UDP-galactose-4'-epimerase from *Trypanosoma brucei* in complex with the substrate analogue UDP-4'-deoxy-4'-fluoro- α -D-galactose (UFG). Probable H-bonds between the ligand and with Tyr173 and Ser142 side chains are indicated as *black* dashed lines. The ligand and residues Ser142 and Tyr173 of the protein are drawn in sticks (PDB code 2CNB; Alpey *et al.*, 2006).

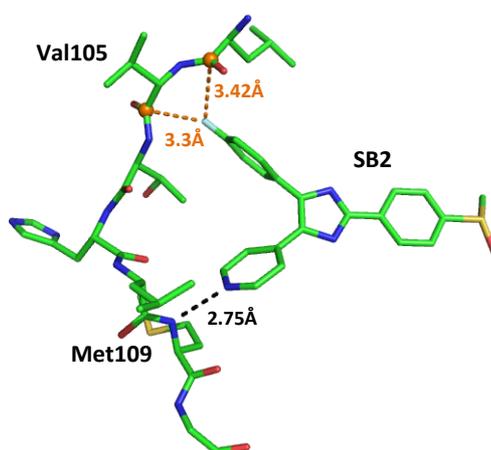


Figure 4.5 Binding of a fluorinated inhibitor to p38 kinase. Polar interactions have been observed in many fluorinated inhibitors of p38 kinase (Boehm *et al.*, 1996). One example is shown in this figure (PDB code 1AU9). The introduction of a fluorine atom forms two more polar interactions with the protein, increasing the binding affinity. Atoms involved and interactions are represented in *orange* balls. H-bonds are in *black* dashed lines. (Wang *et al.*, 1998).

It has been demonstrated that the kind of bond formed usually correlates with the electron density surrounding the fluorine atom. In this sense, Dalvit and Vulpetti (Dalvit & Vulpetti, 2011) proposed the “rule of shielding”, which is an empirical correlation between the fluorine isotropic chemical shift, measured by ^{19}F -NMR spectroscopy, and the type of fluorine–protein interactions observed in crystal structures in the Protein

Data Bank (PDB; Berman *et al.*, 2000).

Chemical shifts are a measure of the fluorine atom environment, which define how much the atom is shielded. In ^{19}F -NMR spectra, fluorine atoms rich in electron density are upfield, while the de-shielded atoms are downfield. Thus the shielding nature of the fluorine atom increases from the left to right along the chemical shift scale. The isotropic chemical shifts of compounds with groups CF_3 , CF_2 and CF were correlated with fluorine–protein interactions observed in crystal structures in the PDB. The results show that shielded fluorine atoms are observed primarily in close contact to hydrogen bond donors within the protein structure, suggesting the possibility of intermolecular hydrogen bond formation with the side chain nitrogen of histidine (78 % NE2 and 22 % ND1), tryptophan (NE1), glutamine (NE2), asparagine (ND2), lysine (NZ), arginine (NE, NH1, NH2), the oxygen atom of serine (OG), threonine (OG1), and tyrosine (OH). De-shielded fluorine atoms are predominantly found in close contact with hydrophobic side chains and with the carbon of carbonyl groups of the protein backbone.

Importantly, not all of the fluorine atoms in close contact with hydrogen bond donor nitrogen or oxygen atoms form hydrogen bonds. Investigations of the X-ray crystal structures of small molecules and fluorinated natural products listed in the Cambridge Structural Database (CSD; Allen, 2002) and available in the PDB reveal a general description of the $\text{F}\cdots\text{H}-\text{X}$ (where $\text{X}=\text{N}, \text{O}$) bond (Dalvit & Vulpetti, 2012; Vulpetti & Dalvit, 2012). As shown in figure 4.6, three measurements are considered: $\text{F}\cdots\text{H}$ distance (d), $\text{F}\cdots\text{H}-\text{X}$ angle (α_1) and $\text{C}-\text{F}\cdots\text{H}$ angle (α_2). The angle α_1 shows a linear correlation with distance d . For short contacts ($d < 2.13 \text{ \AA}$), the angle α_1 is restricted to values between 140° and 180° , whereas with larger distances the allowed range of angles becomes increasingly wide (110° – 180°). The difference in range is due to the repulsive electrostatic interaction between fluorine and oxygen, which has to be accommodated. Angle α_2 does not correlate with the distance d , and its range is between 100° and 170° .

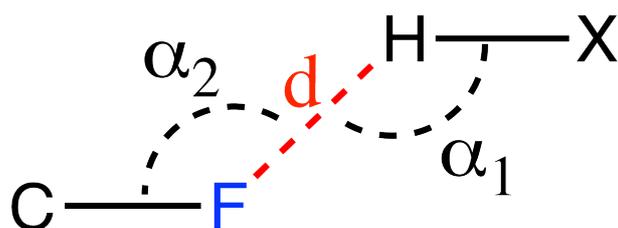


Figure 4.6 Schematic representation of C-F...H-X bond. d , α_1 and α_2 are represented in a C-F...H-X (where X=N, O) bond.

Short contacts C-F...H-X (where X = O, N), are also called dipolar interactions (Zhou *et al.*, 2009). The fluorine atom has three electron lone pairs that are held tightly by the nucleus, resulting in only weak hydrogen bond-accepting capability. Therefore, C-F...H-X looks much more like a dipolar interaction rather than a H-bond. Even if this observation appears to be a semantic discussion, it is important to remember that hydrogen bonding has a covalent component, whereas dipolar interactions by definition are purely electrostatic. The covalent component is anisotropic and provides distinct local geometric control whereas electrostatics do not.

In conclusion, the rule of shielding is a general guideline that will aid in the rational drug design process, providing guidance in the selection of fluorinated moieties for judicious incorporation into molecules in order to create favourable interactions with the protein target.

4.2 Aim of this project

The knowledge of the chemistry of fluorinated phosphonates has led to the synthesis of α -fluorinated phosphonocarbohydrates as a new class of phosphate mimics of natural metabolites that retain high affinity for targeted enzymatic phosphate binding pockets but are themselves resistant to phosphatase-mediated cleavage. In the last decade, several α -fluorinated phosphonate analogue compounds were synthesised: G6P, D-galactofuranosylphosphonate, difluoromethylphosphonate azasugars and fluorinated phosphonodithioacetate. All these compounds are potential inhibitors against specific proteins: D-galactofuranosylphosphonate could inhibit enzymes involved in D-Gal metabolism, difluoromethylphosphonate azasugars and fluorinated-phosphonodithioacetate are potential inhibitors for glycosyl transferase; finally the possibility to synthesise phosphatase resistance analogues as potential inhibitors of phosphatidylinositol 3-phosphate (PtdIns(3)P) has been also exploited (Alphey *et al.*, 2006; Behr, 1997; Berkowitz *et al.*, 2000; Kovensky *et al.*, 1999; Xu *et al.*, 2006).

The aim of this work is to apply this knowledge to the design of phosphonates and fluoro-phosphonate analogues of β G1P. Two phosphonates and four fluorine phosphonate analogues were produced. The evaluation of the best of these six candidates was performed through a combination of biochemical binding assay, ^{19}F -NMR and X-ray crystallographic structural studies. Since this work has been done in collaboration with Professor Dr. Jakeman (Dalhousie University, Nova Scotia, Canada), Prof. G.M. Blackburn and Dr. Jin Yi (University of Sheffield, UK), the non-crystallographic data is attributed to them.

4.3 Experimental procedures

4.3.1 Design and synthesis of the phosphonate analogues and ^{19}F -NMR experiments

βG1P analogues were designed by Prof. G.M. Blackburn (Sheffield University, UK) and synthesised by Dr. Debabrata Bhattasali and Professor D. Jakeman (Dalhousie University, Nova Scotia, Canada). The six phosphonates obtained were evaluated by fluorescence binding (section 4.3.2) and ^{19}F -NMR studies for the formation of metal fluoride TSAs with aluminium and magnesium fluoride. NMR experimental details are described in Dr Jin Yi's thesis (Jin, 2012).

4.3.2 Fluorescent titration binding assay

The binding assay experiments were performed by Dr. Y Jin (Sheffield University, UK). Dissociation constants (K_d) for the compounds prioritised by ^{19}F -NMR experiments were measured using a fluorescence based binding assay. This technique is based on the concept that the intrinsic fluorescent spectra of a protein changes when it binds a ligand. Fluorescence spectra were measured using a Varian Cary Eclipse Fluorescence Spectrophotometer. The excitation wavelength was 285 nm, which in turn gave an emission spectrum with a maximum at 450 nm.

A cuvette containing 0.5 mL of 20 μmol βPGM solution in a buffer of 50 mM HEPES pH 7.2, 5 mM MgCl_2 , 1 mM EDTA, 10 mM NH_4F was prepared and a fluorescence spectrum was run, giving a value for the apo-protein fluorescence. Trial compounds in stock solution were prepared at high concentration in the same protein buffer. For ligand titration, the compound solution was added to the protein, mixed by pipetting for 30 sec and a fluorescence spectrum determined. This gave a value for the fluorescence maximum at 336 nm, corresponding to the state where compound is fully bound to the protein in a TSA complex under the experimental conditions. This was repeated for a total of 25 aliquot additions. Data analysis was performed with Origin8 software (www.OriginLab.com).

4.3.3 Crystallisation of, β PGM-YO5-MgF₃⁻-TSA, β PGM-GRX-MgF₃⁻-TSA and β PGM-GRX-AlF₄⁻-TSA

Crystallisation of methylenephosphonate analogue (CH₂βG1P, also described as YO5) bound to βPGM-MgF₃⁻, and of the monofluoromethylenephosphonate analogue ((S)-βCHFG1P, also described as GRX) bound to βPGM as MgF₃⁻ or AlF₄⁻ TSAs, was carried out in a similar manner to that previously described (Baxter *et al.*, 2010). βPGM was expressed and purified as detailed previously, in the laboratory of Prof Jon Waltho in Sheffield university (Baxter *et al.*, 2006; Griffin, 2011). Buffer was freshly exchanged to 50 mM HEPES pH 7.2, 5 mM MgCl₂ and 0.1 mM DTT prior to crystallisation experiments. For each complex, 10 mM NH₄F and 5 mM βG1P analogue were added at a protein concentration of 15 mg/mL. For the βPGM-GRX-AlF₄⁻-TSA, a further 2 mM AlCl₃ was added to form the aluminium complex. For crystallisation, 2 μL of the complex solution defined above was mixed 1:1 with the precipitant (27-32% PEG 4000, 50-75 mM magnesium acetate) and placed in sitting drop crystallisation plates. Large plate crystals appeared between 1-4 days. They were mounted directly from the mother liquor using a mesh loop and cryo-cooled as described in Pellegrini *et al.*, 2011.

4.3.4 Data collection, structure solution and refinement

Diffraction data were collected from cryo-cooled crystals between 1.5 Å and 1.1 Å resolution on a PILATUS 6M detector on beamline ID29 or on a ADSC Q210 CCD detector on beamline ID14-2 at the ESRF, Grenoble, France. Data were processed with XDS (Kabsch, 2010). The βPGM-G6P-MgF₃⁻-TSA complex structure was used as a search model with Molrep (PDB code 2WF5; Baxter *et al.*, 2010) with all ligands and water molecules removed. Due to the high resolution, in the refinement protocol B-factors values were refined as anisotropic. Solvent molecules were built with the ARP/Waters function of ARP/wARP (Lamzin & Wilson, 1993). A refinement dictionary for each molecule was generated using ProDRG (Schuttelkopf & van Aalten, 2004). Refinement and validation were performed as described in Chapter 2.

4.4 Results and Discussion

4.4.1 Design and synthesis of β G1P phosphonate analogues

Three β G1P analogues, one phosphonate and two fluorinated phosphonates were designed and synthesised. The major modification was the replacement of the C(1)-O by a CH_2 or CHF group. The CH_2 phosphonate was synthesised in order to analyse the effects of the C-F group (**a1** in Figure 4.7, $\text{CH}_2\beta$ G1P). The fluorinated phosphonate compounds were separately prepared in both configurations *R* and *S*: (*R*)- β CHFG1P **b1** and (*S*)- β CHFG1P **c1** respectively (Figure 4.7). Analysis of both isomers of the CHF function has been found to be important. In studies on fluorinated phosphonate analogues of G6P by Berkowitz and co-workers, titration binding experiments showed that the *R* stereoisomer was a better-binding analogue than the *S* isomer. However, this difference did not correlate with the results of steady-state enzyme kinetic analysis, where the two CHF isomers showed a similar $k_{\text{cat}}/K_{\text{m}}$ values. These data clearly indicate that it is a challenge to identify the best substrate analogue *a priori*, without considering the protein environment (Berkowitz *et al.*, 2000).

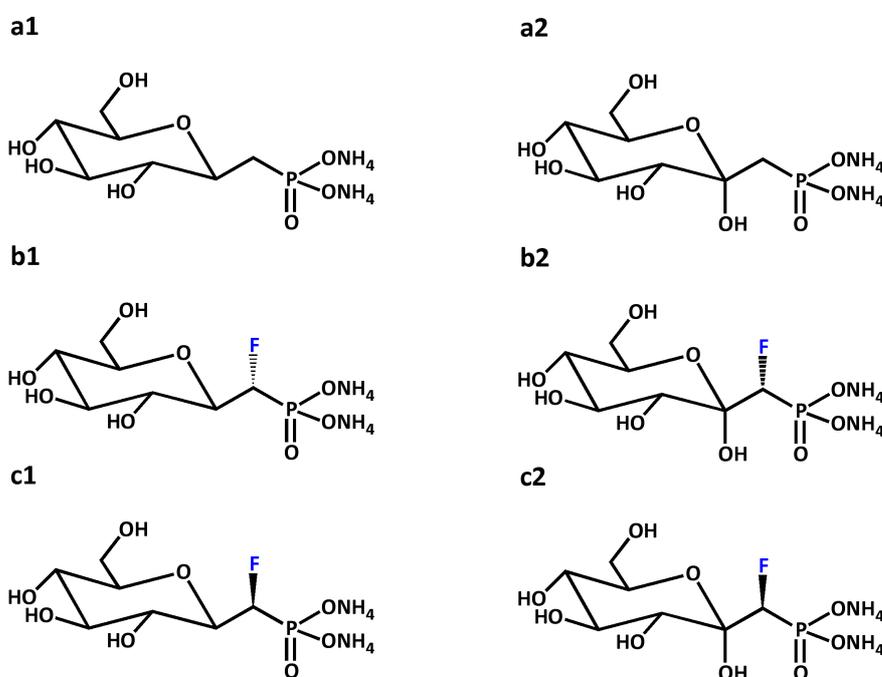


Figure 4.7 Chemical structures of β G1P phosphonate analogues.

In addition, the three compounds, **a2**, **b2**, and **c2** (Figure 4.7) were also produced. These compounds have an α -hydroxyl group on C(1) as they were intermediates in the synthetic pathway prior to reduction to the 1-deoxy counterparts **a1**, **b1**, and **c1**. Being readily available, they were employed to extend testing of the substrate steric discrimination of β PGM.

After synthesis, a combination of several techniques was used to judge the compound that best formed a transition state analogue complex. Without an activity test *in vivo*, the most powerful compound was defined as the one able to bind to the protein with the highest affinity. First the ability of each compound to form a β PGM- β G1P-TSA in presence of AlF_4^- or MgF_3^- , and thus inhibit the enzyme, was tested. Following successful TSA complex identification by ^{19}F -NMR, analogues giving the best results were used in crystallisation trials. Two crystal structures were used to analyse the transition state of phosphorylation of β G1P and to rationalise the relative differences in binding affinity, determined by the fluorescence assay for TSA complex formation.

4.4.2 ^{19}F -NMR, binding studies and X-ray crystallography identified two viable candidates

4.4.2.1 ^{19}F -NMR experiments and binding studies

All six compounds were tested by ^{19}F -NMR for their potential to form a trifluoromagnesate or tetrafluoroaluminate TSA complex with β PGM (Figure 4.8). The β -1-phosphonomethylene-D-glucopyranose (**a2**) and (*R*)-1- β -phosphonofluoromethylene-D-glucopyranose (**b2**) compounds showed no sign of TSA complex formation. (*S*)-1- β -Phosphonofluoromethylene-D-glucopyranose (**c2**) forms a very weak complex with aluminium fluoride but not with magnesium fluoride. The fluoromethylenephosphonate compound (*R*)- β CHFG1P (**b1**) did not form a complex with either magnesium or aluminium fluoride. On the other hand, β CH₂G1P (**a1** or YO5) and (*S*)- β CHFG1P (**c1** or GRX) formed very stable TSA complexes with both magnesium and aluminium fluorides (Jin, 2012). This contrasts with the Berkowitz results where both diastereoisomeric fluorophosphonates bind, albeit to different degrees.

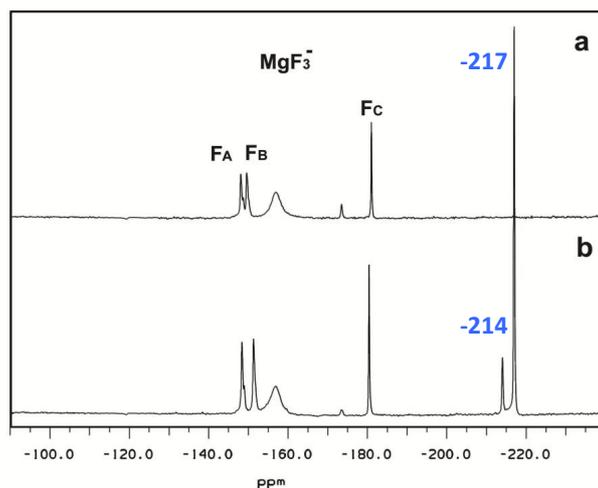


Figure 4.8 ^{19}F -NMR spectra of (a) $\beta\text{PGM}-\beta\text{CH}_2\text{G1P}-\text{MgF}_3^-$ -TSA complex and (b) $\beta\text{PGM}-(S)-\beta\text{CHFG1P}-\text{MgF}_3^-$ -TSA complex. The spectra were recorded at the same concentrations of fluoride, magnesium, and βG1P analogue, showing $(S)-\beta\text{CHFG1P}$ binding stronger than $\beta\text{CH}_2\text{G1P}$. Both samples contain 1 mM βPGM , 5 mM MgCl_2 , 10 mM NH_4F , 5 mM $(S)-\beta\text{CHFG1P}$ or $\text{CH}_2\text{G1P}$ at pH 7.3. The free and bound $(S)-\beta\text{CHFG1P}$ in (b) show peaks at -214.0 and -217.0 ppm, labelled in *blue* (Jin, 2012).

The strength of interaction between the protein and $\beta\text{CH}_2\text{G1P}$ or $(S)-\beta\text{CHFG1P}$ was calculated through a fluorescence based binding assay, performed by Dr. Y Jin. Results are shown in figure 4.9. $(S)-\beta\text{CHFG1P}$ binds twice as strongly as the corresponding non-fluorinated compound, showing that the *S*-fluorine atom has created favourable features to TSA binding.

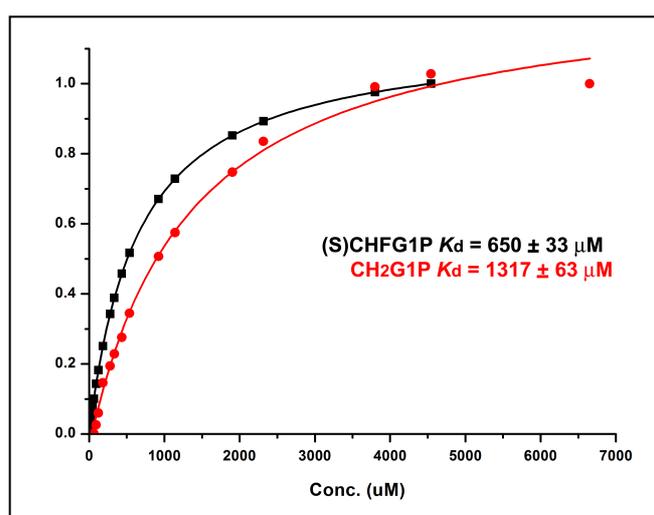


Figure 4.9 Fluorescence titration curve from a plot of binding fraction versus [ligand].

4.4.2.2 X-ray crystallography

Taken together, the NMR and binding studies clearly prioritised using compounds $\beta\text{CH}_2\text{G1P}$ and (*S*)- βCHFG1P for crystallisation trials, with very good results. The following three TSA complexes were crystallised successfully: $\beta\text{PGM-YO5-MgF}_3^-$ -TSA, $\beta\text{PGM-GRX-MgF}_3^-$ -TSA, and $\beta\text{PGM-GRX-AlF}_4^-$ -TSA, where YO5 is $\beta\text{CH}_2\text{G1P}$ and GRX is (*S*)- βCHFG1P . All complexes crystallised in the orthorhombic space group $P2_12_12_1$ with one molecule in the asymmetric unit. The structure of $\beta\text{PGM-MgF}_3^-$ -YO5 was solved at the highest resolution achieved for this protein (1.1 Å). It provided for the first time the features of the fully resolved $\beta\text{PGM-YO5-MgF}_3^-$ -TSA complex. The structures for the second and third complex were solved at 1.5 Å resolution. Because of the high resolution, the difference Fourier maps show the geometry of the ligands very clearly. In particular, the electron density for the sugar analogues is well defined, by contrast to that observed for G6P (PDB code 2WF5; Baxter *et al.*, 2010), implying tighter binding in the active site. The statistics are summarised in table 4.2. In the following discussion, the terminology $\beta\text{CH}_2\text{G1P}$ and (*S*)- βCHFG1P will be used instead of YO5 and GRX.

Table 4.2 Data collection and refinement statistics

Structures	β PGM-YO5-MgF ₃ ⁻ -TSA ⁻	β PGM-GRX-MgF ₃ ⁻ -TSA ⁻	β PGM-GRX-AlF ₄ ⁻ -TSA
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.976	0.933	0.933
Unit cell dimension (Å) a,b,c	38.01, 54.20, 105.41	37.54, 54.34, 104.35	37.19, 54.28, 104.5
Resolution range (Å) ¹	20 (1.16-1.1)	20 (1.58-1.5)	20 (1.58-1.5)
Number of unique reflections	82900	33872	34022
Multiplicity	4.7 (4.7)	3.4 (2.9)	4.2 (3.4)
Completeness (%)	93.1 (85.7)	97.1 (88.1)	98.1 (92.1)
R _{merge}	0.052 (0.352)	0.042 (0.338)	0.085 (0.595)
<I/σ (I)>	12.7 (4.9)	19.8 (3.2)	12.5 (2.1)
Wilson B factor (Å ²)	5.37	11.56	11.11
Water molecules	566	296	456
R _{factor} (%)	14.92	17.25	15.91
Free R _{factor} (%)	17.26	19.84	19.94
RMS deviations:			
Bonds (Å)	0.01	0.03	0.015
Angles (°)	1.36	1.42	1.39

¹Values for the higher resolution shell are in brackets

4.4.2.2.1 Structure characterisation of β PGM- β CH₂G1P-MgF₃⁻-TSA

Using β CH₂G1P in combination with aluminium or magnesium fluoride, a snapshot of the β G1P phosphorylation reaction was obtained with high precision for the first time. Experimentally, crystallisation trials were successful only for the magnesium fluoride sample. In the β PGM- β CH₂G1P-MgF₃⁻-TSA structure the nucleophilic OH (6-OH) is perfectly in line with the magnesium fluoride atom and the side-chain O δ from Asp8. The pentacoordinate magnesium fluoride is in the typical TBP geometry. Due to the high resolution of the data, the magnesium fluoride was refined without any restraints allowing the bond lengths to be determined with high precision. The resulting equatorial Mg-F bonds are 1.83 ± 0.02 Å, in accord with those previously described (Baxter *et al.*, 2010) (Figure 4.10).

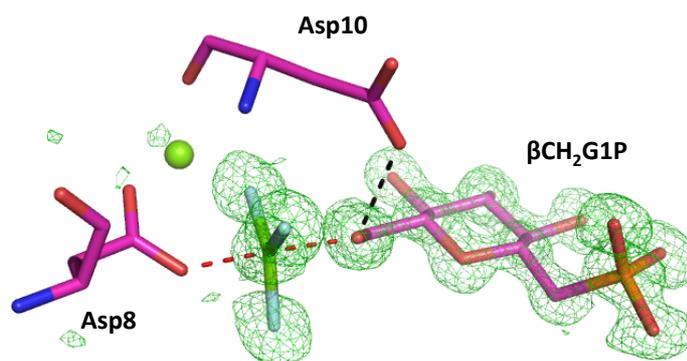


Figure 4.10 Active site of the β PGM- β CH₂G1P-MgF₃⁻-TSA complex. Difference Fourier electron density for β CH₂G1P and MgF₃⁻ is shown as *green* mesh contoured at 3σ . Hydrogen bonds are shown as *black* dashed lines. Interactions in red dashed lines of TSA are shown. The two aspartates involved in covalent and general base catalysis are shown as sticks.

The magnesium of the metal fluoride complex is almost equidistant from the two oxygen atoms ($6\text{OH-Mg} = 2.11$ Å; $\text{Mg-OAsp8} = 2.07$ Å), which suggest a tight transition state. As in the case of β PGM-G6P-MgF₃⁻-TSA structure, one of the O δ atoms of Asp10 forms a hydrogen bond to the nucleophilic 6-OH group and it is so positioned for general base catalysis of the nucleophile by accepting its proton. Due to the high resolution, all the Asp10 hydrogen atoms were easily modelled by the phenix.refine program (Afonine *et al.*, 2012). The hydrogen of the amide Asp10

backbone (Asp10H^N) is pointing to one of the equatorial fluorine atoms, which contributes to TSA stabilisation (Figure 4.11).

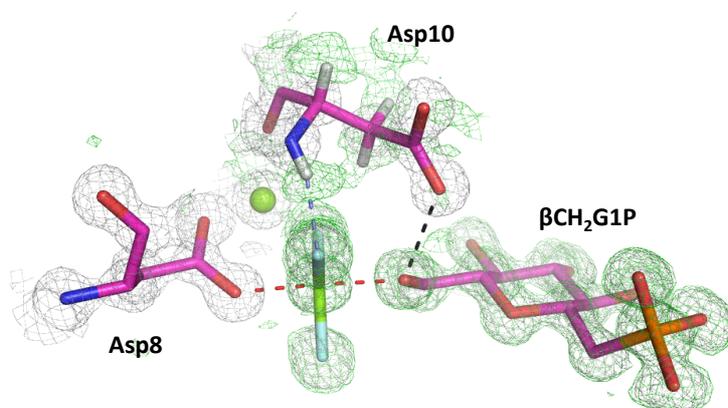


Figure 4.11 Asp10 in the β PGM- β CH₂G1P-MgF₃⁻-TSA complex. Difference Fourier electron density map and electron density map for the β PGM active site are represented in *green* mesh (contoured at 3σ) and *grey* mesh (contoured at 3σ) respectively. Hydrogen bonds are in *black* dashed lines. Interactions in *red* dashed lines of TSA are shown. The bonding between Asp10 and MgF₃⁻ is shown in *light blue*. The two aspartates involved in catalysis are shown as sticks.

4.4.2.2.2 Specificity for the recognition of β CH₂G1P and G6P

As in the β PGM-G6P-MgF₃⁻-TSA structure, residues from both the cap and the core domain contribute to the active site.

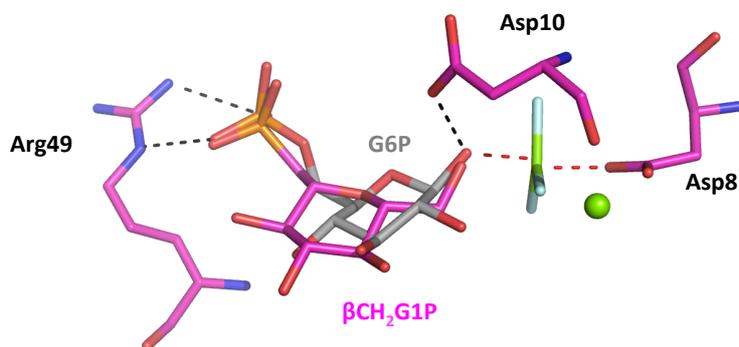


Figure 4.12 G6P and β CH₂G1P in β PGM active site. G6P (in grey) and β CH₂G1P (in magenta) are shown in the active site of β PGM of structure β PGM- β CH₂G1P-MgF₃⁻-TSA. Arg49 coordinates the phosphorus group of the sugar, while Asp10 and Asp8 interact with the 6-(OH) of β CH₂G1P or the 1-(OH) of G6P.

Despite the fact the two structures do not show any large differences in conformation (rmsd between Ca atoms of 0.3 Å), the sugar ring is coordinated differently (Figure 4.12).

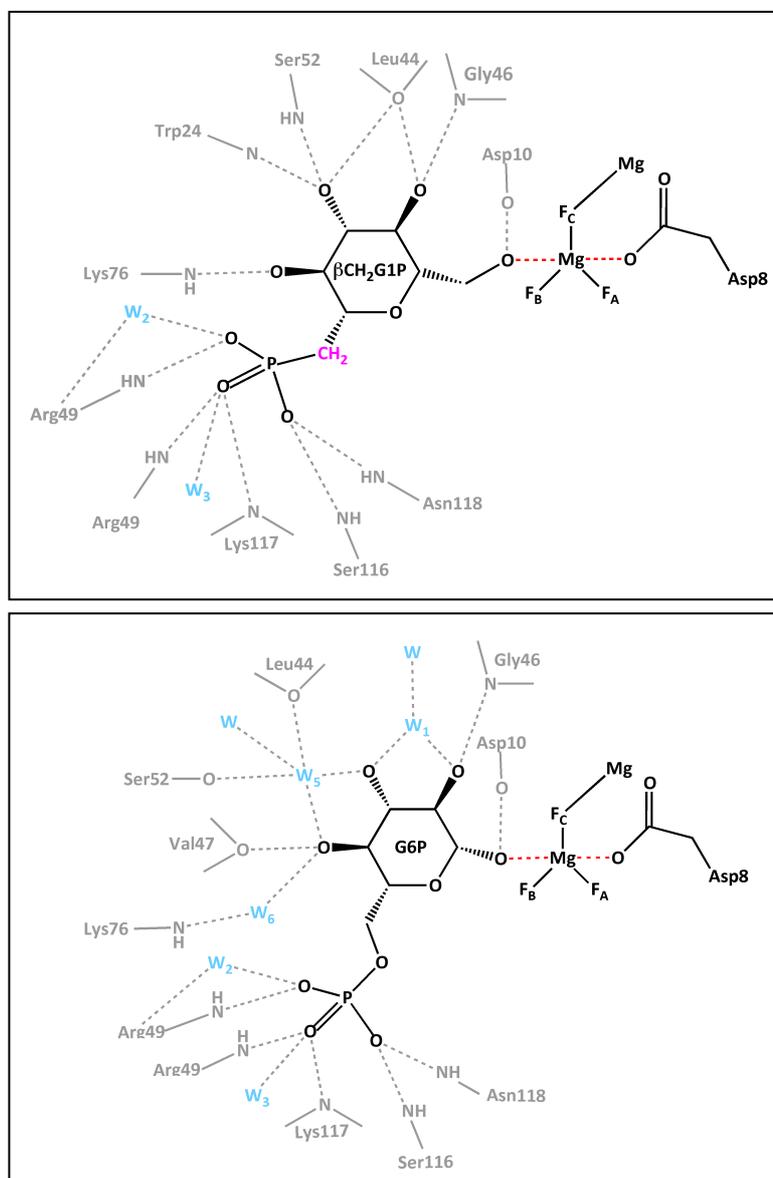


Figure 4.13 Schematic view of the β PGM- β CH₂G1P -MgF₃⁻-TSA and β PGM-G6P-MgF₃⁻-TSA active site. The overlay of hydrogen bonds are shown as *grey* dashed lines. “W” are water molecules. Oxygen atoms belong to the carboxylic backbone group and backbone amide are shown between two V shape lines. Groups belong to side chains are shown with a straight line.

In the case of β CH₂G1P, residues Trp24, Ser52 and Leu44 and Lys76 coordinate directly to the sugar ring. In contrast in G6P coordination Trp24 does not participate

while Val47 becomes involved. Moreover, the other residues (Leu44, Ser52, and Lys76) coordinate indirectly G6P, using water molecules as intermediates (Figures 4.12-4.13). Surprisingly, the structural conformation of β PGM region 136-144 is also different in the two TSA complexes. In β PGM-G6P-MgF₃⁻-TSA, Ser144 assumes two alternative conformations. In the structure with β CH₂G1P, the Ser144 side chain shows a third conformation that allows the residue to interact with the Ser48 backbone. Ser48 is part of a clamp loop that terminates with Arg49, a key residue in the coordination of the “static” sugar phosphate (Figure 4.14).

Overall, the β CH₂G1P TSA complex is more tightly bound than the corresponding G6P-TSA complex, because of the absence of water molecules coordinating the sugar hydroxyl groups and better stabilisation of the phosphate clamp region that traps the sugar inside the active site.

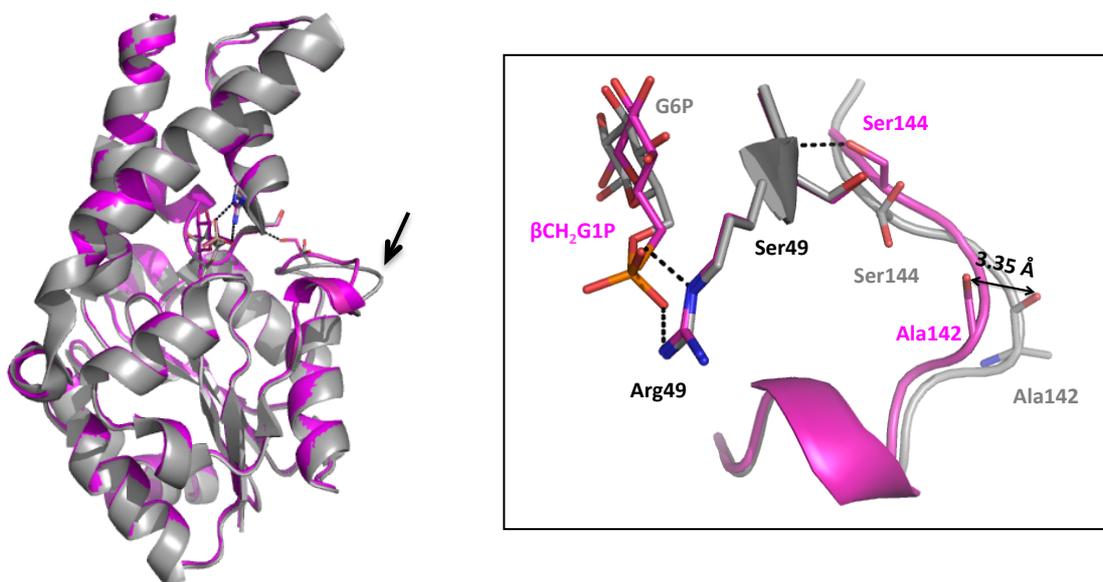


Figure 4.14 Structural details of the interaction between the region of residues 136-144 and the clamp loop. A *black* arrow indicates the protein region 136-144 and the clamp loop position in the overlay of β PGM- β CH₂G1P-MgF₃⁻-TSA (in *magenta*) onto β PGM-G6P-MgF₃⁻-TSA (in *grey*). The panel arrow represents the locus of maximum spatial difference between the two structures. Residues involved in the phosphate clamp stabilisation are represented as sticks. Hydrogen bonds are in *black* dashed lines.

4.4.2.3 Structural characterisation of β PGM-(*S*)- β CHFG1P-TSA structures

The β PGM-(*S*)- β CHFG1P-TSA complex was crystallised successfully with both AlF_4^- and MgF_3^- . The two structures are very similar (0.165 rmsd of $C\alpha$). Further comparisons in rmsd on $C\alpha$ atoms are reported in table 4.3. As in the case of the β PGM- β CH₂G1P - MgF_3^- -TSA complex no large conformational differences were observed (Figure 4.15).

Table 4.3 Rmsd of $C\alpha$ displacement among (*S*)- β CHFG1P, β CH₂G1P and G6P structures

	β PGM- β CH ₂ G1P - MgF_3^- -TSA	β PGM-G6P- MgF_3^- -TSA
β PGM-(<i>S</i>)- β CHFG1P - MgF_3^- -TSA	0.274	0.216
β PGM-(<i>S</i>)- β CHFG1P - AlF_4^- -TSA	0.292	0.224

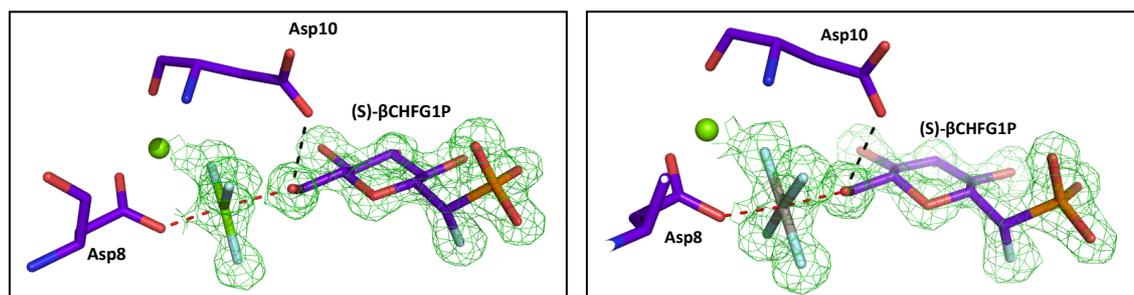


Figure 4.15 Active sites of the β PGM-(*S*)- β CHFG1P- MgF_3^- -TSA active site and β PGM-(*S*)- β CHFG1P - AlF_4^- -TSA complex. Difference Fourier density map for (*S*)- β CHFG1P and MgF_3^- is shown as *green* mesh, contoured at 4σ .

Comparing β PGM-(*S*)- β CHFG1P- MgF_3^- -TSA with β PGM- β CH₂G1P- MgF_3^- -TSA, differences can be seen in sugar coordination and in the structure of the clamp loop and β PGM region 136-144. The introduction of fluorine appears to result in a possible additional interaction between the fluorine and the side chain amide of Arg49 at a distance of 2.96 Å. This amide group is coordinated by a water molecule (W_2) in β PGM- β CH₂G1P- MgF_3^- -TSA, which is absent when the fluorophosphonate is bound (Figure 4.16).

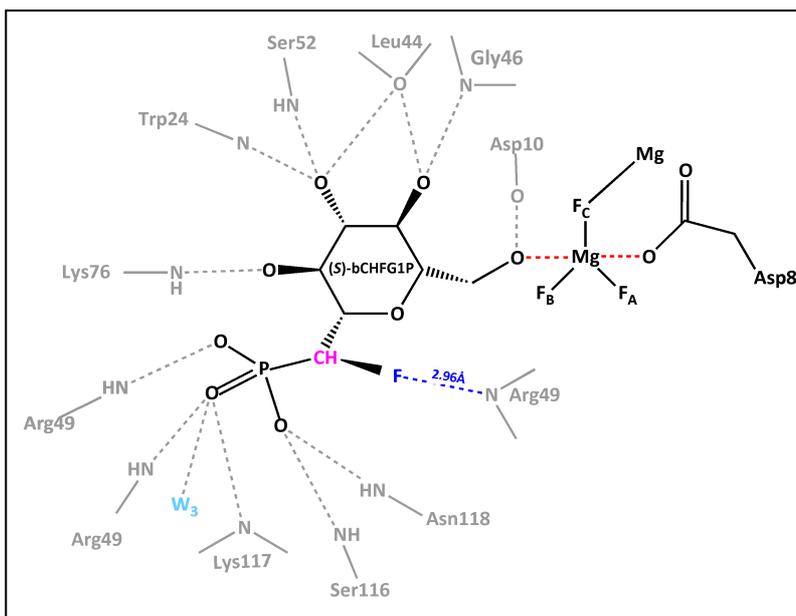


Figure 4.16 Schematic view of the β PGM- (S)- β CHFG1P - MgF_3^- . The figure legend is the same as figure 4.13. The distance between the fluorine and the Arg49 amine is shown as *blue* dashed line.

In addition, a new pattern of water molecules is observed in the region, located close to Arg49, the α -fluorine, and the sugar phosphate. Loop 136-144 is in a similar conformation to that in the β PGM-G6P- MgF_3^- -TSA, apart from the fact that the interaction with the clamp region is maintained as in β PGM- β CH₂G1P- MgF_3^- -TSA. This is due to water molecules coordinated by Ser49, Asp51 and Ser144 (Figure 4.17). Thus, it appears that a third conformation of this loop exists in addition to the ones described for β PGM-G6P- MgF_3^- -TSA and β PGM- β CH₂G1P- MgF_3^- -TSA. It is unclear whether this interaction is caused by the fluorophosphonate.

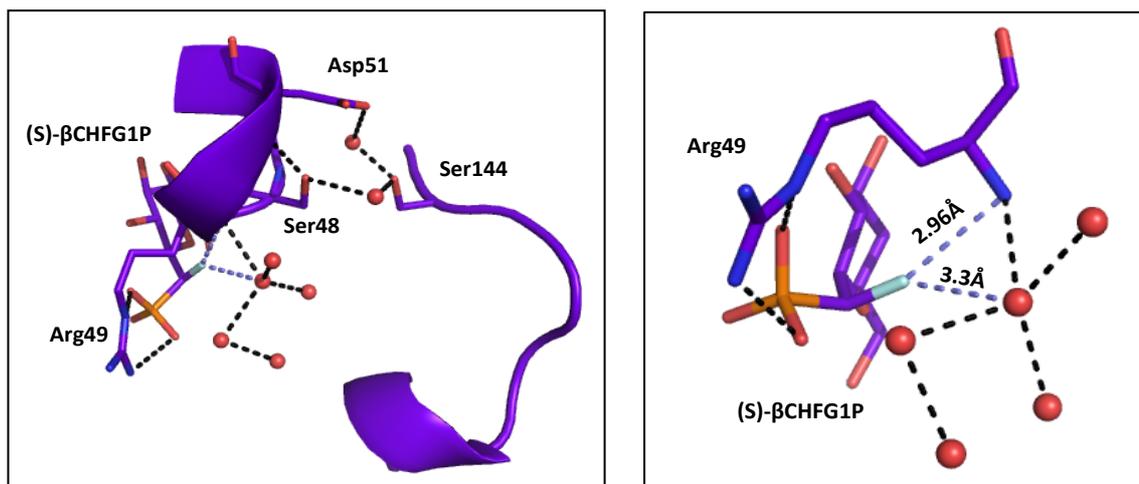


Figure 4.17 Structural details of the β PGM- β CH₂G1P-MgF₃⁻-TSA active site and β PGM region 136-144. Hydrogen bonds are in *black* dashed lines, water molecules are *red* spheres. Distances of C-fluorine to Arg49 and C-fluorine to a water molecule are shown in *light blue* dashed lines. The right panel is a magnification of the fluorine and Arg49 coordination.

4.4.3 (*S*)- β CHFG1P is the best ligand for TSA formation

From ¹⁹F-NMR, the binding affinity assay, and the structures solved, β CH₂G1P and (*S*)- β CHFG1P appear to be the best candidates for forming TSA complexes with the mutase. Moreover, the affinity assay showed a higher *K_d* for the (*S*)-fluoro-phosphonate, which appears to make a novel interaction with Arg49 in the structure.

Thus, these data provide answers to three specific questions:

- Why do phosphonates with 1- α -hydroxyl on D-glucopyranose not bind?
- Why does (*S*)- β CHFG1P bind stronger than β CH₂G1P to β PGM?
- Why is there specificity for the (*S*)-stereochemistry of the α -fluorine?

4.4.3.1 Why do phosphonates with a 1- α -hydroxyl phosphonate not bind?

¹⁹F-NMR has shown that 1- α -hydroxyl analogues 1b, 2b and 3c (Figure 4.7) do not form a TSA complex. The crystal structures show clearly that the 1- α -hydroxyl on D-glucopyranose is a negative factor for complex formation, as the distance between C(1) on (*S*)- β CHFG1P and the imidazole ring of His20 is too short to allow an α -hydroxyl substituent on C(1) (Figure 4.18).

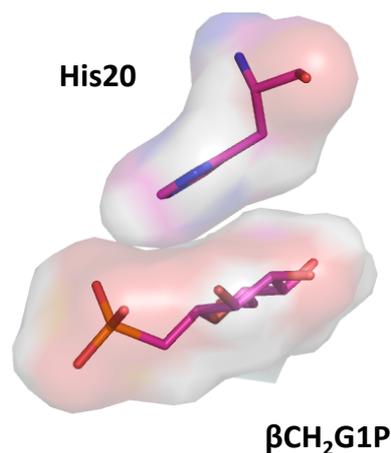


Figure 4.18 Surface representation of His₂₀ and β CHFG1P in β PGM-(*S*)- β CHFG1P-MgF₃⁻-TSA structure. The distance between the His₂₀ ring and the sugar ring is 3.7 Å. The insertion of an α -C1-hydroxyl group would clearly disrupt the hydrophobic interaction between the imidazole and the pyranose ring and it will prevent binding.

4.4.3.2 Why does (*S*)- β CHFG1P bind more strongly than β CH₂G1P to β PGM?

In section 4.2.3.3 it was shown that the α -fluorine is in close proximity to the amide backbone of Arg49. The shielding rules described in section 4.1.2.2 can now be applied to determine whether a favourable interaction exists.

NMR experiments have determined that the ¹⁹F chemical shift for the fluoro-sugar analogue is in the range of -214/-217 ppm, which suggests a shielded fluorine atom. Following the rule of shielding, if the atom is close to a H-bond donor, a H-bond or a dipolar interaction should be observed.

The program phenix.reduce was used to calculate the probable position of protons (Afonine *et al.*, 2012) (Figure 4.19). Fluorine is at short distance from the two protons of Arg49: H^N of the amide backbone (2.63 Å), and H^{C β} of Arg49 (2.68 Å). The first group is a H-bond donor, while the second one could make a polar interaction. Since the chemical shift is strongly negative, that rules out the possibility of a polar interaction.

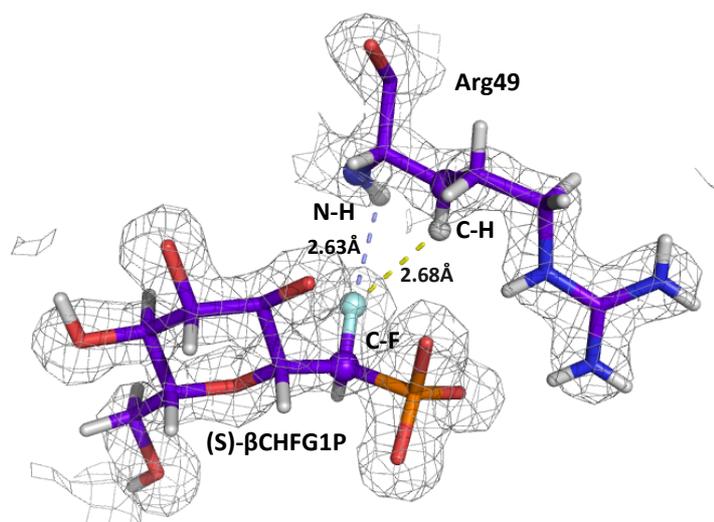


Figure 4.19 Protons in β PGM-(S)- β CHFG1P-MgF₃⁻-TSA structure. The electron density map contoured at 2σ is shown for (S)- β CHFG1P and Arg49. Distances between fluorine and protons of amide backbone and C _{β} of Arg49 are shown in *light blue* and *yellow* respectively.

For defining if a dipolar interaction or a H-bond is present, the shielding rules consider three measurements: F \cdots H distance (d), F \cdots H-X angles (α_1) and C-F \cdots H angle (α_2). F \cdots H is 2.63 Å, while α_1 and α_2 are respectively 104.36° and 172.3° (Figure 4.20). The long d should correlate with a α_1 in the range 110°–180°, therefore the measured α_1 is outside this range and should not support H-bond assignment. Similarly, angle α_2 is near the limit. However, because the two angles are close to the limits, the presence of a H-bond cannot be excluded. Unfortunately, even at this resolution, an X-ray structure cannot define the real position of protons. Thus, in conclusion, (S)- β CHFG1P binds tighter than CH₂G1P in the TSA complex because of a probable interaction between the fluorine and the backbone amide proton of Arg49. Since the fluorine atom is a poor H-bond acceptor, the bond is considered a dipolar interaction.

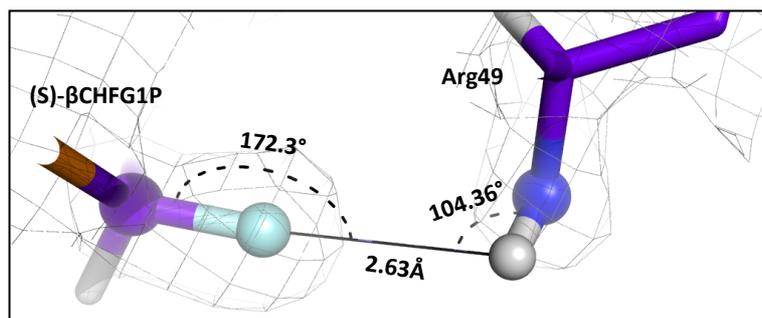


Figure 4.20 The C-F...H-N bond in β PGM-(*S*)- β CHFG1P-MgF₃⁻-TSA structure. The electron density map countered at 2σ is shown in *grey* for (*S*)- β CHFG1P and Arg49.

4.4.3.3 Why is there specificity for the fluorine (*S*)-stereochemistry?

¹⁹F-NMR data has shown that the stereoisomer (*R*)- β CHFG1P does not bind in a trifluoromagnesate complex and binds only weakly in a tetrafluoroaluminate complex. At first sight, there is no evident steric hindrance to explain this result. The reason has to be found in the possible protein interaction. Thus, using Pymol tools, the fluorine in *R* configuration was modelled and possible residues involved in adverse interactions defined. The model revealed that fluorine in the *R* configuration will be close to potential hydrogen bond donors, such as the amide groups of Ser116 or Lys117. Importantly, the *R* configuration will be too close to the pyranose oxygen, which implies that an unfavourable dipole-dipole interaction will be formed. Such interactions can drive rotation of both C1 and phosphate group, which will change the small molecule structure to a non-binding mode. Furthermore the *R* conformer is close to Ser114, whose backbone has the ability to flip inside out (see Chapter 3) making that region unstable and a poor candidate for binding (Figure 4.21).

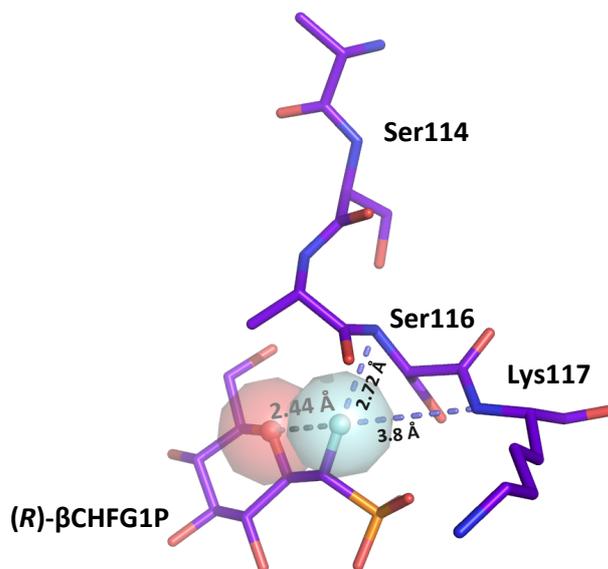


Figure 4.21 VdW clash in (R)-βCHFG1P. Distances between the fluorine and possible H-bond donors are shown in *light blue* dashed lines. Surface representation of fluorine and pyranose oxygen atom show where VdW spheres overlap. Clearly the (*R*)-fluorine eclipses the pyranose oxygen in the bound conformation of the βG1P. The VdW clash between F and O5 is at a distance of 2.44 Å shown as *grey* dashed line, while the sum should be 2.67 Å.

4.5 Conclusions

Several goals have been achieved in this study. First, two analogues of β G1P able to form TSA complexes with β -phosphoglucose mutase have been identified by analysis of transition state complex formation. Combinations of complementary techniques were essential for achieving this goal.

Second, high resolution analysis of these TSA complex structures for phosphorylation of β G1P has enabled the first direct comparison of the features of catalysis of both steps of a mutase reaction that bind a common hexose moiety in two different orientations. The overlay of structures reveals that the sugars are coordinated differently and also that β PGM region 136-144 is a flexible protein region, possibly involved in sugar binding/recognition, and which adopts different conformations based on which sugar mono-phosphate is bound. It appears that β PGM has developed to recognise both sugars: interacting more specifically with β G1P and adapting the same binding site for G6P. As expected, the more specific binding concerns the sugar that defines the α/β anomer specificity.

Third, the advantage of fluorination of a phosphonate analogue has been demonstrated to improve the physical properties relative to β G1P, and the dependence of strong binding identified with specific stereochemistry at the chiral C-CHF-P centre: the order being (*S*)-CHFP > CH₂P >> (*R*)-CFHP for the trifluoromagnesate transition state analogue complex. The higher affinity of the (*S*)-configuration CHF- over the CH₂-analogue may be due to an additional dipole interaction. A higher resolution structure or a neutron structure will be useful for better evaluating the nature of fluorine. The strong reduction of binding of the (*R*)-analogue depends on the fact that the VdW contacts between the fluorine in *R*-conformer and the pyranose oxygen will clash, which probably induces a non-binding conformational change in the sugar phosphate. Although the details of hydrogen bonding cannot be determined directly, the crystallographic results are in full accord with ¹⁹F-NMR and fluorescence binding studies, which suggests the (*S*)-CHFP β G1P analogue to be a possible lead compound for combating bacterial infections.

Chapter 5 Novel insight into RhoA GTP-hydrolysis

5.1 Introduction

5.1.1 RhoA is a small G protein

There are two distinct families of G proteins: heterotrimeric G proteins and small G proteins. Heterotrimeric G proteins, also called Large G proteins, are activated by G protein-coupled receptors and are made up of α , β , and γ subunits. Small G proteins are monomeric proteins homologous to the α -subunit found in the heterotrimers. Small G proteins (20 ~ 25kDa) belong to the *Ras superfamily of small GTPases*. As the α -subunits, most of them are localised to membranes. This is due to modified C-terminus containing prenyl groups (such as farnesyl and geranylgeranyl), which act as lipid anchors (Bos *et al.*, 2007).

All small G proteins are inefficient GTPases. The rate of spontaneous GTP-hydrolysis in water (10^{-9} s^{-1}) is accelerated to 10^{-4} s^{-1} by small G proteins. However, this is insufficient for biological signalling. A protein called GAP (GTPase Activating Protein) is required to enhance the reaction up to 10^5 -fold (Lamarche & Hall, 1994; Lancaster *et al.*, 1994). How GAPs stimulate the intrinsic GTPase activity of G proteins, is an intriguing question that has been largely investigated through biochemical and structural approaches. Structures of small G proteins in complex with GAP and aluminium fluoride transition state analogues have elucidated this mechanism and magnesium fluoride has been first used in this context (Graham *et al.*, 2002). However, some questions remain unanswered, which have been addressed in this thesis. Before introducing the aim of this work, the structure and functional role of the Ras superfamily proteins will be summarised. In particular, the *state of the art* of the Rho proteins family, to which RhoA belongs, will be discussed.

5.1.1.1 The Ras superfamily of monomeric G proteins

The Ras superfamily is divided into five major families: the Ras, Rho, Arf/Sar, Ran, and Rab families. Recent phylogenetic analysis has identified 167 proteins in the human Ras

superfamily: 39 Ras proteins, 30 Arfs, 22 Rhos, 65 Rabs, and 1 Ran family (Rojas *et al.*, 2012). All the members function as signalling nodes in eukaryotes that are activated by diverse extracellular stimuli and that regulate intracellular signalling. This signalling ultimately controls gene transcription, which in turn influences fundamental processes such as cell growth and differentiation. The **Ras family** is involved in transmitting signals from the cell surface to other parts of the cell, in particular the nucleus. The genes expressed regulate cell proliferation, differentiation, morphology, and apoptosis (Karnoub & Weinberg, 2008). The **Rab family** is involved in regulating the traffic of intracellular transport vesicles and the trafficking of proteins between different organelles *via* endocytotic and secretory pathways (Zerial & McBride, 2001). The **Rho family** is involved in signalling networks that regulate actin, cell cycle progression, cell motility, cell behaviour and gene expression (Heasman & Ridley, 2008; Mulloy *et al.*, 2010; Park & Bi, 2007; Schlessinger *et al.*, 2009). The **Ran proteins** are the most abundant in the cell and they are involved in nuclear transport. Finally, the **Arf family** of proteins comprises the most divergent proteins, which, like Rab family proteins, are involved in vesicle trafficking (Wennerberg *et al.*, 2005). These are referred to as “atypical” Rho/Rac GTPases because they are very different from the other GTPase subfamilies according to structural, regulatory and functional criteria (Bustelo *et al.*, 2007).

5.1.1.2 The Small G protein switching cycle and structure correlation

Despite the divergence in function, all Ras proteins have a conserved mechanism. They act as molecular switches, moving between an active GTP-bound state and an inactive GDP-bound state (Mackay & Hall, 1998). The interconversion between the two forms allows Ras proteins to act as molecular switches that regulate the intracellular signalling pathways described above (Ihara *et al.*, 1998). The entire cycle is highly regulated by two protein families: the Guanine nucleotide-exchange factors (GEFs) and the GTPase-activating proteins (GAPs). The GEFs reset the Ras switch, promoting the exchange of the nucleotide from GDP to GTP (Bourne *et al.*, 1990). The GAPs promote the inactive conformation by stimulating the intrinsic GTPase activity of the G protein (Boguski M. S. and McCormick F., 1993) (Figure 5.1). For a subset of small G proteins (mainly the

Rab and Rho proteins) guanine nucleotide-dissociation inhibitors (GDIs) provide an additional level of control. These proteins remove small G proteins from membranes by sequestration of their lipid tails (Bos *et al.*, 2007).

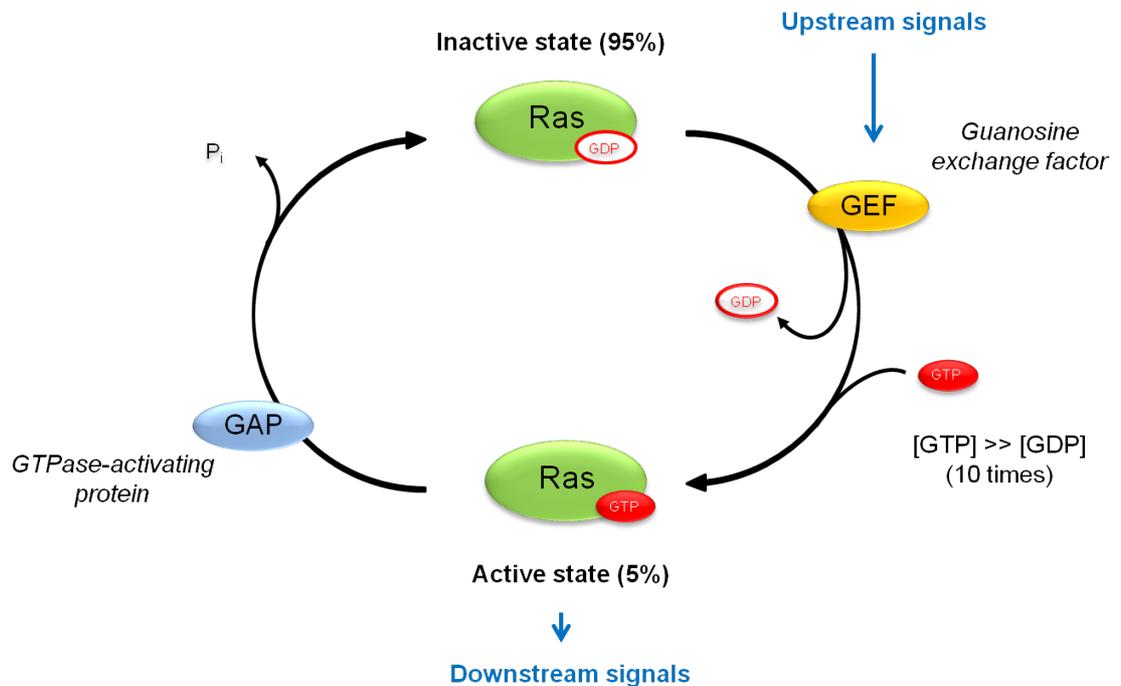


Figure 5.1 The Ras superfamily G protein cycle. The diagram shows the regulation of Ras proteins in the cell. On receiving an upstream signal, the GEF activates Ras by stimulating it to release its GDP; as the cellular GTP concentration is greater than the GDP concentration, Ras rapidly binds GTP once GDP has been ejected. The GAP stimulates GTP-hydrolysis, inactivating Ras. GAP maintains most of the Ras protein in unstimulated cells in an inactive GDP-bound form (modified from Alberts, 2008).

The Ras superfamily of small GTP-binding proteins maintains a structurally and mechanistically preserved GTP-binding core (Colicelli, 2004; Wennerberg *et al.*, 2005). Crystallographic and NMR analyses of some small G proteins, including Ha-Ras, N-Ras, Rap2A, RhoA, Rac1, Rab3A, Rab7, Arf1, and Ran, have revealed that all GDP/GTP-binding domains have a common topology (Geyer & Wittinghofer, 1997). By comparison of the structure in the GTP-bound conformation and the GDP-bound conformation, two highly flexible regions surrounding the γ -phosphate of GTP have been established: the Switch I region within loop L₂ and β_2 (the effector region) and the

Switch II region within loop L_4 and helix α_2 . Switch I (28-38 residues in RhoA) and II (61-78 residues in RhoA) regions are the small G protein regions mostly involved in the GTP-GDP transition (Figure 5.2).

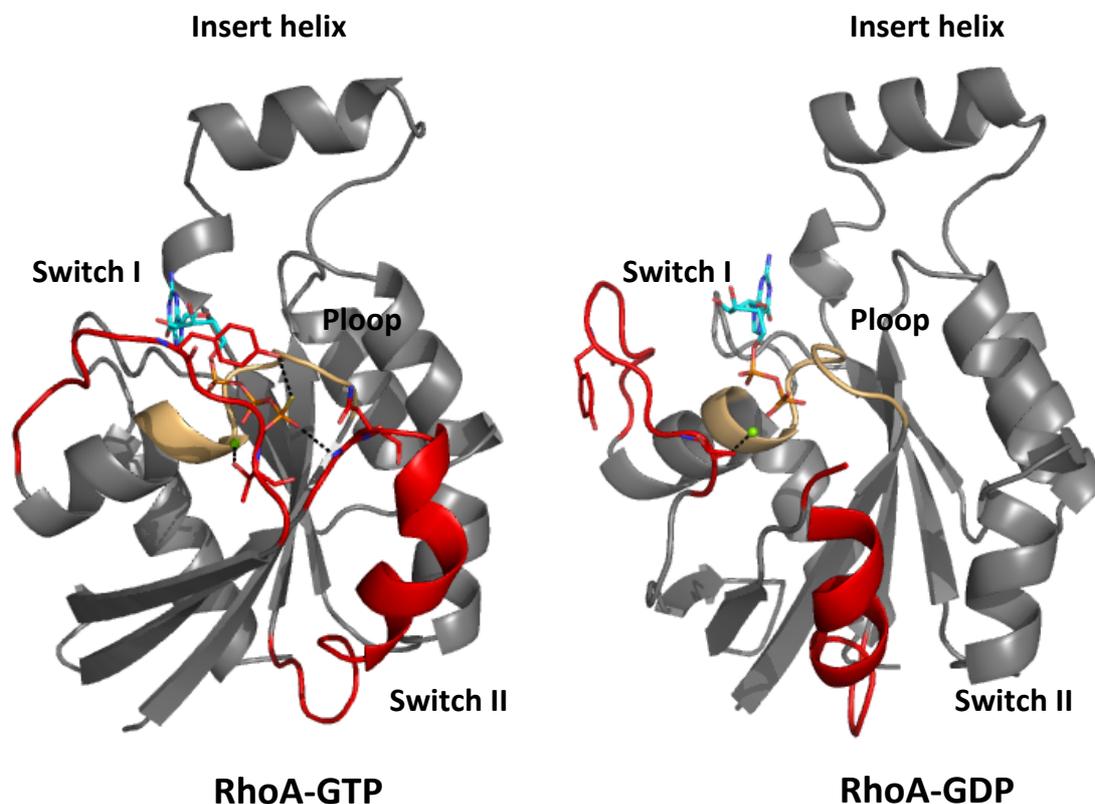


Figure 5.2 Structure of Switch I and II regions in RhoA_{G14V}-GTP γ S and RhoA-GDP structures. RhoA consists of a single domain with six stranded β -sheets, surrounded by α -helices connected with loops. In this ribbon representation of RhoA_{G14V}-GTP γ S on the left (PDB code 1A2B; Ihara *et al.*, 1998) and RhoA-GDP on the right (PDB code 1FTN; Wei *et al.*, 1997), Switch regions and other relevant regions are highlighted: Switch I (residues 27-36) and Switch II (residues 61-78) are in *red*, the catalytic magnesium ion is shown as a *green* sphere, GDP is in *cyan* sticks and the P-loop is in *brown*. The P-loop or phosphate binding loop (residues 13-20) wraps around the nucleotide so strongly that GTP and GDP bind to RhoA with dissociation constants in the nanomolar range. Tyr34, Thr37 from Switch I and Gly62, Gln63 from Switch II are in sticks. *Black* dashed lines represent hydrogen bonds.

The Ras superfamily proteins generally undergo an enzymatic cycle that involves the so-called loaded-spring mechanism, where release of the γ -phosphate after GTP-hydrolysis allows the two switch regions to relax into the GDP-specific

conformation (Ihara *et al.*, 1998; Milburn *et al.*, 1990; Vetter & Wittinghofer, 2001). In the GTP-bound state there are two hydrogen bonds from GTP γ -phosphate oxygen atoms to the main chain amine groups of the invariant Thr37 and Gly62 residues, in Switch I and II respectively. In the GDP-bound state, where the GTP γ -phosphate is already hydrolysed, these interactions are broken (Figure 5.2). Both GEFs and GAPs act on small GTP proteins by interacting with the Switch regions.

5.1.1.3 Physiological roles of mammalian Rho family

In the mammalian Rho GTPases family, twenty protein isoforms have been described: Rho (A, B and C), Cdc42, Rac (1, 2 and 3), TC10; TCL, Chp (1 and 2), RhoG, Rnd (1, 2 and 3), RhoBTB (1 and 2), RhoD, Rif and TTF (Etienne-Manneville & Hall, 2002). The human genome contains over 82 Rho family GEFs, over 67 Rho family GAPs, and only three Rho family GDIs (Hall, 2012; Meller *et al.*, 2005).

The three best characterised Rho family members are: Cdc42, Rho and Rac (Hall, 1998; Madaule & Axel, 1985). Historically, they have been described by their effects on actin polymerisation on serum-starved Swiss 3T3 fibroblast cells *in vitro*. Activation of the three proteins promotes fibroblast movement by formation of filipodias, lamellipodias, membrane ruffles, focal adhesion and stress fiber formation (Figure 5.3).

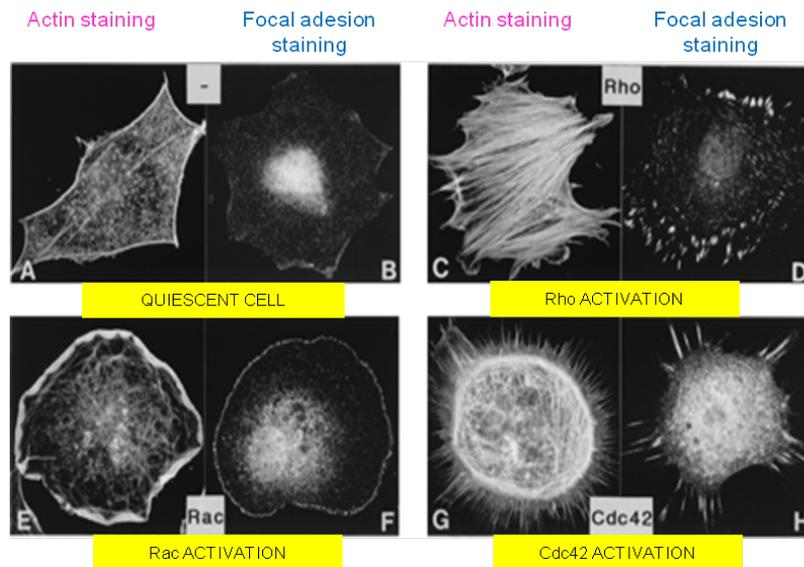


Figure 5.3 The effects of Rac, Rho, and Cdc42 on actin organisation in fibroblasts. In each case, the actin filaments have been labelled with fluorescent phalloidin and the focal contacts have been located with an antibody against vinculin. A-B) Quiescent, serum-starved Swiss 3T3 fibroblasts (-) contain very few organised actin filaments or vinculin containing integrin adhesion complexes. C-D) Addition of the growth factor lysophosphatidic acid activates Rho, which leads to stress fibres and focal adhesion formation. E-F) Microinjection of constitutively active Rac induces lamellipodia and associated adhesion complexes. G-H) Microinjection of FGD1, an exchange factor for Cdc42, leads to formation of filopodia and the associated adhesion complexes. Cdc42 activates Rac; hence, filopodias are intimately associated with lamellipodias, as shown in (G). Scale: 1 cm = 25 μm (modified from Hall, 1998).

In vivo, the regulation of the actin cytoskeleton and further functions, mediated by the Rho family, contributes to different aspect of cell biology, such as morphology, movement and behaviour of the cell.

All three proteins are involved in the morphogenesis of epithelial and neuronal cells, by contributing to **cell polarity** determination. In epithelial cells, the Rho GTPase pathway is involved in the formation of cell-cell interactions such as adherent junction, tight junction and cell-ECM (extra cellular matrix) interactions; these contacts contribute to the apical/basal polarity of epithelial cells (Ebnet *et al.*, 2001; Jacinto *et al.*, 2001; O'Brien *et al.*, 2001; Raich *et al.*, 1999; Vasioukhin *et al.*, 2000). In neurons, Rho proteins are determinants for axon or dendrite development. While Cdc42 and Rac promote the neurite outgrowth, Rho inhibits the extension (Lee *et al.*, 2000; Li *et al.*, 2000; Luo, 2000; Ozdinler & Erzurumlu, 2001; Wong *et al.*, 2000). Furthermore, the Cdc42 pathway is one of the main determinants in establishing cell polarity in yeast

budding and cell asymmetry in mammalian zygotes (Gotta *et al.*, 2001; Kay & Hunter, 2001; Pruyne & Bretscher, 2000).

In **single-cell migration**, Rho GTPases act on both actin and the microtubule cytoskeleton. Rac promotes the protrusive activity, by actin polymerisation and integrin adhesion complex assembly at the cell periphery (Condeelis, 2001; Kraynov *et al.*, 2000; Small *et al.*, 2002), while Rho is involved in the contraction and retraction of the rest of cell body by promoting the formation of stress fibers (Worthylake *et al.*, 2001). The isoform RhoA is particularly involved in this process (Ridley *et al.*, 1992). Cdc42 pathway stabilises the direction of the movement (Allen *et al.*, 1998). Recent studies suggest that Rho is also active at the leading edge protrusion, where, together with Rac, it promotes the protrusion (Machacek *et al.*, 2009). Moreover, the Rho proteins are involved in cell migration and cell stretching involved in the dorsure closure during morphogenesis (Brock *et al.*, 1996; Knust, 1997; Lu & Settleman, 1999; Stronach & Perrimon, 1999).

Rho, Rac and Cdc42 contribute to **cell behavior**, where contractile actin and myosin filaments are used by many cell types to induce rapid, reversible changes in shape. The Rho pathway promotes actin and myosin contraction in aortic smooth muscle cells for the regulation of blood pressure (Fukata *et al.*, 2001; Sakurada *et al.*, 2001). It also promotes destabilisation in endothelial cell-cell junctions to regulate extravasion of lymphocytes from circulating blood into surrounding tissues (Wojciak-Stothard *et al.*, 2001). All three GTPases are involved in several behaviour processes, such as phagocytosis (Bokoch, 2000; Caron & Hall, 1998), G1 cell cycle procession (Cantrell, 1998; Olson *et al.*, 1995; Olson *et al.*, 1998; Welsh *et al.*, 2001) and in regulating secretion of histamine and serotonin from mast cells (Pinxteren *et al.*, 2000).

Rho GTPases mediate their signalling by activating a huge number of effectors. 100 targets have been reported until now, which include 30 kinases and a large number of scaffold-adaptor-like proteins (Bishop & Hall, 2000; Hall, 2012). Well known effectors are ROCK and mDIA, which are both activated by Rho (RhoA for ROCK) and they lead to polymerisation of actin in linear filaments and activation of myosin II respectively (Kimura *et al.*, 1996; Watanabe *et al.*, 1999; Watanabe *et al.*, 1997) (Figure 5.4).

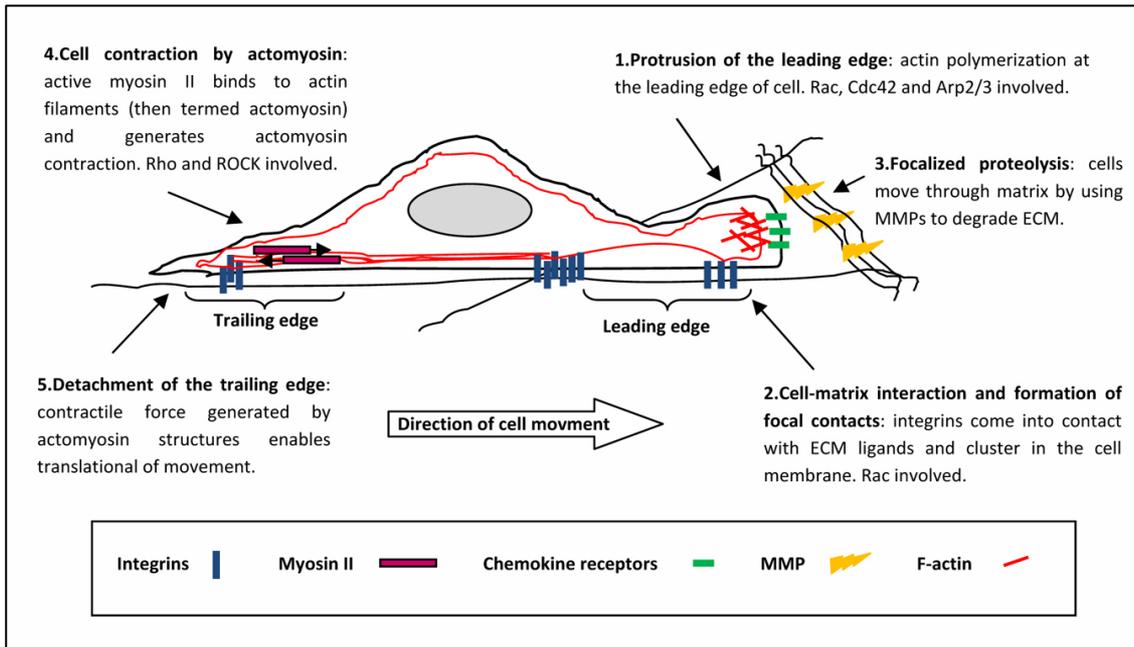


Figure 5.4 Rho family signalling in cell motility. The steps required in cell motility are shown with details of signalling molecules involved. Arp2/3 effectors are activated by Rac (Etienne-Manneville & Hall, 2002), while ROCK is activated by Rho. In step 3 MMPs, which are surface protein proteases, are produced. It has been suggested that Rho signalling is required for the activation of these proteases (Parri & Chiarugi, 2010).

5.1.1.4 The Rho family role in cancer and further diseases

Ras, the founder member of the Ras superfamily was intensely studied in the 1980s as it was found to be an oncogene in 30% of all human cancers (Hall, 2012). Further analysis revealed a single amino acid substitution resulting in a protein that was insensitive to GAPs and therefore constitutively active. Similar mutations have been reproduced *in vitro* in Rho family proteins. The Rho oncogenes are dominant active mutants (Cdc42V12, RacV12, and RhoV14), which corresponds to a permanently GTP-bound state, insensitive to GAP (Feig, 1999). Constitutively active forms of Rho family proteins in cancer cells were thought to promote the acquisition of migratory and invasive properties, a key event in the oncogenic progression of cells (Hanahan & Weinberg, 2000; Schmitz *et al.*, 2000). However, the enormous amount of sequencing information derived from human cancers has uncovered none of the mutations hypothesised *in vitro* in Rho GTPases. Further studies have shown that the role of the Rho family proteins in cancer is due to up-regulation of almost all the family, which

affect all the cancer steps, within the transformation step. The up-regulation induces uncontrolled proliferation, which increases survival signals and permits the tumour cells to escape from apoptosis (Vega & Ridley, 2008). For example, in the case of the Rho proteins, RhoA and RhoC are up-regulated, while RhoB is down-regulated. *In vitro* and *in vivo* studies have shown that RhoA has transforming activity, while RhoC up-regulation correlates to the metastasis step (Gomez del Pulgar *et al.*, 2005; Hakem *et al.*, 2005). RhoB down-regulation correlates with increasing tumour aggressiveness (Huang & Prendergast, 2006).

The possibility that the Rho family GEFs and GAPs might act as tumour suppressors has also been explored. Many of the Rho GEFs were originally identified as oncogenes in experimental tissue culture transformation assays. Despite this, there are just a few sporadic examples of genetic alterations in Rho GEFs in human cancer (Hall, 2012).

For GAPs, there is strong evidence supporting an oncogenic function for DLC1 (deleted in liver cancer 1), a GAP active on Rho (Xue *et al.*, 2008; Yuan *et al.*, 2003). Heterozygous deletions are found in approximately 50% of human cancers, and loss of DLC1 can promote tumorigenesis in mouse cancer models. However, homozygous deletions of DLC1 are not found in human cancer, whereas expression from the remaining allele appears to be attenuated in some cancers, definitive proof for a tumour-suppressor role has yet to be obtained (Yuan *et al.*, 2003).

Rho signalling is also involved in cardiovascular, pulmonary hypertension, neurodegenerative and immune disorders (Hall, 2012; Lu *et al.*, 2009; Nossaman & Kadowitz, 2009; Shimokawa, 2002; Shimokawa *et al.*, 2002). In these diseases genetic mutations of Rho signalling proteins have been found. Mutations in GAPs, GEFs and targets are found in a variety of mental retardation syndromes and it has been proposed that these lead to defects in spine morphogenesis and an alteration in the size and shape of dendritic spines associated with memory and learning (Nadif Kasri & Van Aelst, 2008). Genetic alterations are also associated with immunological disorders: Rac2 is mutated in LAD (leucocyte adhesion deficiency), while Cdc42 and cdc42 GEF FGD1 are respectively mutated in WASP (Wiskott–Aldrich syndrome), and developmental disorders such as faciogenital dysplasia syndrome (Kirchhausen & Rosen, 1996; Olson *et al.*, 1996; Pai *et al.*, 2010).

Finally, there is a great deal of interest in ROCK, one of the major Kinase effectors of Rho, as a pharmaceutical target. Aberrant actomyosin contraction has been implicated in a variety of human diseases, including hypertension, atherosclerosis and ischemia. A ROCK inhibitor is already in clinical use in Japan for cerebral ischemia and Phase II trials are underway in the US for treatment of glaucoma (Zhou *et al.*, 2011).

5.1.2 How do GAPs stimulate GTP-hydrolysis?

The role that the Rho family proteins have has led to the idea of Rho GTP-hydrolysis as a natural way to inhibit activity, in particular in cancer, where genetic alterations are not present and a much more efficient hydrolysis system is required. Thus, several biophysical and structural studies have revealed the mechanism by which GAPs stimulate the intrinsic GTP-hydrolysis of small GTP proteins. Effective catalysis of phosphoryl transfer by GAPs comprises several elements: the proper orientation of the attacking water molecule and its polarisation, occlusion of water from the active site, stabilisation of the leaving group and stabilisation of the transition state.

In the case of the Ras superfamily, it has been proposed that GAPs act either through residues that participate directly in the catalysis or through an allosteric mechanism (Wittinghofer, 1993). RhoGAP uses a conserved arginine residue, termed the “Arginine finger”, to stabilise the transition state and to promote the cleavage of the γ -phosphate (γ P) (Barrett *et al.*, 1997). Details of this reaction have been explored in the structures of the RhoA-GDP-AlF₄⁻-RhoGAP-TSA and RhoA-GDP-MgF₃⁻-RhoGAP-TSA complexes (Graham *et al.*, 2002; Rittinger *et al.*, 1997). These structures show how residues that participate directly in catalysis, such as Arg85_{RhoGAP}, are responsible for the hydrolysis of GTP. The Arginine finger acts by stabilising the negative charge on the GTP phosphate groups, as well as by coordinating Gln63_{RhoA}, a vital residue that properly positions a water molecule for nucleophilic attack (Rittinger *et al.*, 1997). Similar conclusions have been made for Ras and Cdc42 proteins (Nassar *et al.*, 1998; Scheffzek *et al.*, 1997). In Rab, both residues are supplied by the GAP, while Rab Gln92 is pointing away from the active site (Pan *et al.*, 2006). In ArfGAP/Sar system the Gap contributes to the reaction with the Arginine finger, while the glutamine is substituted by a histidine (Bi *et al.*, 2002). Thus, it has been proposed that GAPs promote

GTP-hydrolysis by contributing directly to catalysis. However, there is evidence, which suggest that an allosteric function upon protein binding is also relevant. For example, RhoGAP_{R85A} is an active mutant and it is able to increase the hydrolysis rate of Rho-GTP by 160-fold, without affecting protein binding (Graham *et al.*, 1999). Similarly, the mutant Cdc42GAP_{R305A} conserves 20% of its activity (Nassar *et al.*, 1998). Furthermore, it is known that small GTP proteins have an intrinsic GTPase activity, which is independent of the contribution from the Arginine finger. This implies that further contributions to activation are involved in catalysis.

Interestingly, there are GAP proteins, such as RapGAP and RanGAP, which lack an Arginine finger. In the case of RanGAP-Ran, RanGAP supplies an asparagine to stabilise the orientation of the Gln69 of Ran. A tyrosine (Tyr39_{Ran}) belonging to Ran, well conserved in all Ras proteins, interacts with the GTP- γ P instead of the Arginine finger (Seewald *et al.*, 2002). In the case of RapGAP-Rap the GAP contributes an asparagine, referred to as the “Asparagine thumb”, into the active site to substitute the missing glutamine (Scrima *et al.*, 2008). The same tyrosine described for Ran (Tyr32_{Rap}) also interacts with the γ P instead of the Arginine finger. This tyrosine is conserved in Rho, Ras, Cdc42, Ran and Rap: Tyr34_{Rho}, Tyr32_{Ras}, Tyr32_{Cdc42} Tyr39_{Ran} and Tyr32_{Rap} respectively. As shown in figure 5.5, Tyr34_{Rho} and Tyr32_{Ras} are in the so called open conformation, while Tyr32_{Cdc42} Tyr39_{Ran} and Tyr32_{Rap}, are in a close conformation, which correspond to the one coordinating the γ P. The closed conformation appears to clash with a potential Arginine side chain. Surprisingly, in the structure of Cdc42-GDP-AlF₃-Cdc42GAP_{R305A}-TSA, in which the Arginine finger has been mutated to an alanine, the side chain of the conserved Tyr32_{Cdc42} is in the same position as the Arginine finger, coordinating the γ P, in a closed conformation (Nassar *et al.*, 1998) (Figure 5.5). As no Cdc42-GTP structure is available, which is the physiological role of Tyr34_{Rho}, Tyr32_{Ras}, Tyr32_{Cdc42} in the Rho family switch cycle has not been yet defined.

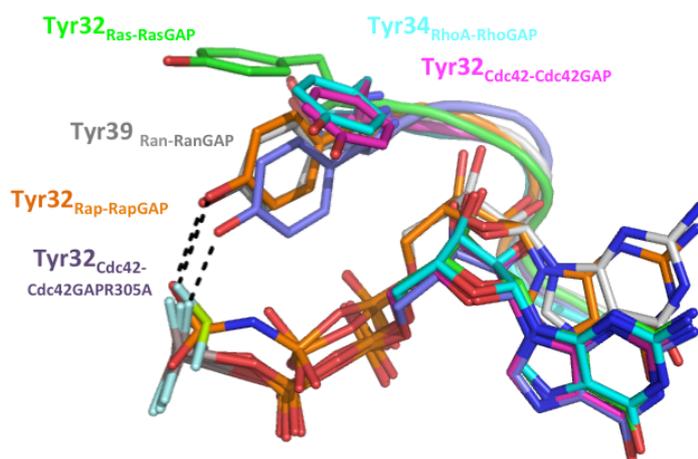


Figure 5.5 Structure conformation of conserved tyrosine. The superimposition of the active sites from various structures shows the two possible conformations assumed by the conserved tyrosine during catalysis. The tyrosine is in closed conformation in Ran-GPPNHP-RanBP1-RanGAP (PDB code 1K5D; Seewald *et al.*, 2002), in Rap-GDP-BeF₃⁻-RapGAP (PDB code 3BRW; Scrima *et al.*, 2008), and in Cdc42-GDP-AlF₃-Cdc42GAP_{R305A}-TSA (PDB code 2NRG; Nassar *et al.*, 1998). The tyrosine is in an open conformation in RhoA-GDP-MgF₃⁻-RhoGAP-TSA (PDB code 1OW3; Graham *et al.*, 2002), Cdc42-GDP-AlF₃-Cdc42GAP-TSA (PDB code 1NGR; Nassar *et al.*, 1998) and Ras-GDP-AlF₃-RasGAP-TSA (PDB code 1WQ1; Scheffzek *et al.*, 1997).

5.2 Aim of this project

The question of how GAPs stimulate the intrinsic GTPase activity of G-proteins is still open. The aim of this work is to investigate the structural and the catalytic consequences caused by the absence of the Arginine finger in the Rho-RhoGAP system, in order to define the relevant characteristics in the active site of Arg85_{RhoGAP} and Gln63_{Rho}. Among the three isoforms of Rho, RhoA was chosen as a model system. RhoA, RhoB and RhoC are highly homologous (83%), but appear to have divergent biological functions (Wennerberg & Der, 2004).

Two systems were studied: RhoA-GDP-AIF_x-TSA, through which the intrinsic RhoA activity will be investigated, and RhoA-GDP-AIF_x-RhoGAP_{R85A}-TSA or RhoA-GDP-MgF_x-RhoGAP_{R85A}-TSA, through which the Arginine finger independent events will be described. Since both RhoA and RhoA-RhoGAP_{R85A} have an intrinsic and residual activity respectively, the question of how charge balance in these catalytic systems is achieved will be discussed.

Finally the results obtained bring novel insight into GAP mediated RhoA GTP-hydrolysis by providing complementary information to the structures already known.

5.3 Experimental procedures

5.3.1 RhoA-GDP-AIF_x-TSA and RhoA-GDP

5.3.1.1 RhoA expression and purification

Human RhoA was expressed and purified recombinantly from pGEX-2T vector, which encodes a cleavable N-terminal GST-tag. The plasmid carrying human RhoA_{F25N} gene was kindly provided by Dr. Katrin Rittinger (MRC, National Institute for Medical Research, London). In this thesis RhoA will refer to this mutant. Expression after IPTG induction was carried out ON, at 20°C in Rosetta 2 cells. A GST-trap column was used for the first purification step. Tag cleavage was performed with thrombin on the column. As a final purification step, size exclusion chromatography was performed (see Chapter 2). Table 5.1 shows details of the purification procedure.

The protein eluted as monomer at an elution volume of 66 ml from the size exclusion column, with a 260/280 ratio close to 1 (Figure 5.6). The ratio is high due to the GDP bound to the protein that co-purified from the cells. The purity was determined by SDS-PAGE. The yield was typically 14mg/L of culture.

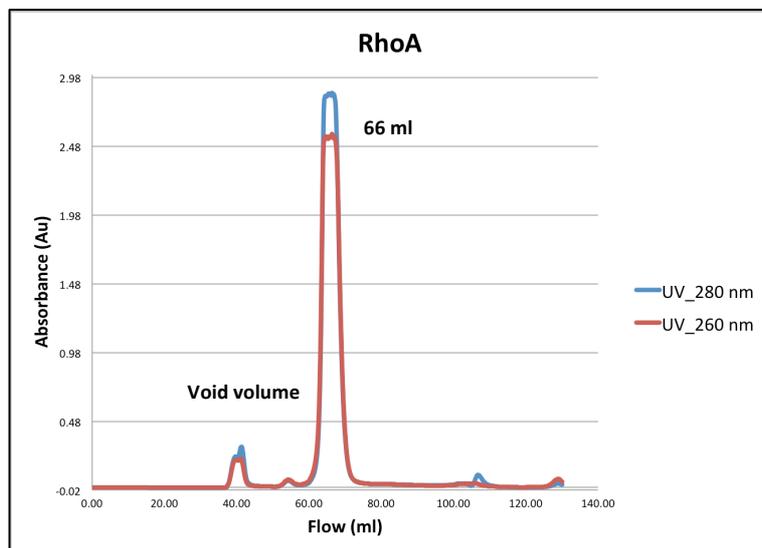


Figure 5.6 Size exclusion chromatography profile of RhoA. From purification of 1.5 L of culture, 20 mg of protein was eluted. The absorbance at 260 nm is shown in *red* and 280 nm in *blue*. The chromatogram was run with a HiLoad 16/60 Superdex 75 pg column.

Table 5.1 Protein purification steps

Purification steps		Solutions
1. Cell Lysis	Bacteria were resuspended in Lysis Buffer or LyB (10ml LB for 750 ml of bacteria) with addition of a protease inhibitor cocktail tablet (EDTA free) and DNase I (trace). The cells were lysed by sonication on ice for 2 * 50 s with 1 minute gap, 60% power; then centrifuged for 20 min at 20,000 g, 4 °C.	Lysis buffer (LyB) = 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM MgCl ₂ , 5 mM DTT
2. Affinity Chromatography	A GST-trap column was used (GST column-5 ml, GH), connected to an AKTA prime system. After equilibrating the column with the BB, the Crude Extract (CE) or soluble phase was applied. The binding step was performed at 0.5 ml/min until the UV line started to decrease. Then the column was extensively washed with BB until the UV lines become flat.	Binding Buffer (BB) = 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM MgCl ₂ , 1 mM DTT
3. GST-tag cleavage	The column was extensively washed with the TCB until the UV lines become flat again. Thrombin sample dissolve in the TB was applied to the column. After ON incubation with at 4°C, gently shaking (100U/20 mg of protein), the column was connected to Benzamidine-trap column and the cleaved protein was eluted using the CB. This step was performed at 1 ml/min.	Thrombin Cleavage Buffer (TCB) = 50 mM Tris pH 8, 150 mM NaCl, 5 mM MgCl ₂ , 1 mM DTT
4. Size exclusion chromatography (SEC)	A Hi load 16/60 Superdex 75 pg gel filtration column was run a flow rate 1 ml/min with SEC buffer.	SEC buffer = 50 mM Bis/Tris pH 6, 150 mM NaCl, 5 mM MgCl ₂ , 1 mM DTT

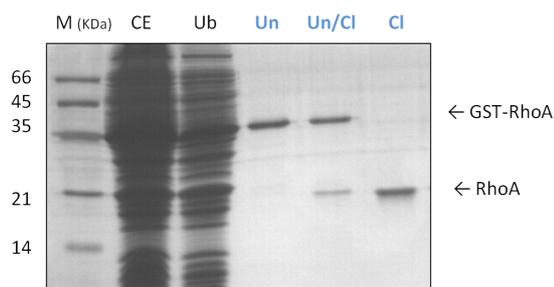


Figure 5.7 12% SDS-PAGE of RhoA purification – initial tests. Samples loaded show details of RhoA purification while the optimisation of the procedure was in progress. Due to its high solubility, the protein was purified from the soluble phase (see CE=crude extract) and an additional column was required to recuperate the Protein remaining in the unbound fraction (Ub) from the affinity column. The further lanes show the uncleaved (Un) and cleaved species (Cl). The column Un/Cl is an example of the test run for checking the thrombin amount required. 2 ug of protein were loaded in each lane. (M=Marker).

RhoA was then concentrated or mixed with RhoGAP (see sections 5.3.2.3-5.3.2.4) in order to prepare a stock solution for crystallisation of the single protein RhoA or for the crystallisation of the complex. RhoA was concentrated to 12 mg/ml and setup for crystallisation experiments immediately. The protein was conserved at -20°C for maximum of one week. After this time, crystals were not obtained.

5.3.1.2 RhoA-GDP-AlF_x-TSA and RhoA-GDP crystallisation

Several crystallisation experiments were performed. The RhoA-GDP-AlF_x sample was prepared by adding 10 mM NH₄F and 2 mM AlCl₃ from concentrated stock solution (0.5 M) to the concentrated RhoA stock solution, without adjusting the final protein concentration. For establishing crystallisation conditions, several commercial screens were tested at the EMBL High Throughput Crystallisation Laboratory (HTX Laboratory), EMBL, Grenoble (Figure 5.8; Table5.2; see Chapter 2).

Buffer	Salt	Precipitant
0.1 M Tris pH8.5	0.2 M NaCl	25% (w/v)PEG 3350
0.1 M Tris pH8.5	0.2 M Li ₂ SO ₄	25% (w/v) PEG 3350
0.1 M Hepes pH7.5	0.2 M Li ₂ SO ₄	25% (w/v) PEG 3350
0.1 M Hepes pH7.5	0.2 M MgCl ₂	25% (w/v) PEG 3350
	0.2 M KF	25% (w/v) PEG 3350

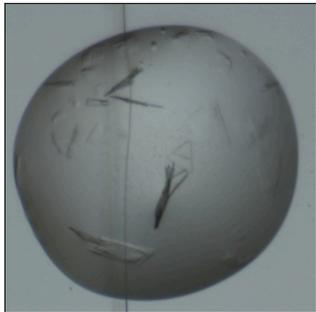


Figure 5.8-Table 5.2 Crystals obtained from initial robot conditions. The table shows the robot conditions in which crystals were observed. The picture shows crystals that diffracted to 2.5 Å in the best direction.

In order to determine the best conditions obtained, the diffraction properties of the plate-like crystals obtained in each condition were tested for their diffraction properties at beamline ID23-1 at the ESRF. All the crystals were thin and diffracting at reasonable resolution only in one direction. The crystals grown from 0.1 M Tris HCl pH 8.5, 0.2 M NaCl and 25%PEG 3350, displayed the best diffraction quality (2.5 Å in the best

direction). Further optimisation was then performed in order to obtain larger and thicker crystals. Several tests were also performed to promote the accommodation of an aluminium fluoride molecule inside the active site. Optimisation of the salt and PEG concentration was necessary to scale up the crystals obtained from the HTX Laboratory from nano-drops to micro-drops. The replacement of NH_4F with NaF increased the crystal quality. Crystals were also obtained at lower pH values (5-7), necessary for maintaining aluminium in solution. Further aluminium and sodium fluoride in the sample and in the reservoir solution were added to improve the probability of aluminium fluoride binding. Crystals were grown at 4°C and RT. Eventually, RhoA-GDP- AlF_x -TSA crystals were obtained at RT by the sitting-drop method from solutions containing 10 mg/ml RhoA, 50 mM Bis/tris pH 6.5, 150 mM NaCl, 5 mM MgCl_2 , 1mM DTT, 2 mM AlCl_3 , 10mM NaF equilibrated against 24% (w/v) PEG 3350 and 0.3 M NaCl. Plate crystals appeared overnight and were mounted directly from the mother liquor using a mesh loop and cryo-cooling as described in E. Pellegrini *et al.*, 2010 (see Chapter 2).

5.3.1.3 Data collection, structure solution and refinement

The structure of RhoA in complex with GDP was solved using molecular replacement with X-ray diffraction data collected at 100 K from a single crystal. All diffraction data were collected on beamline ID29 at the ESRF, Grenoble, France. The diffraction images were then integrated with XDS (Kabsch, 2010). The previously published structure of RhoA-GDP (PDB code 1FTN; Wei *et al.*, 1997) without bound ligand and water molecules was used as a search model for MolRep (Vagin & Teplyakov, 2010). Refinement and validation have been performed as described in Chapter 2.

5.3.2 RhoA-GDP- AlF_x -RhoGAP_{R85A}-TSA and RhoA-GDP- MgF_x -RhoGAP_{R85A}-TSA

5.3.2.1 Site-directed mutagenesis of RhoGAP to generate the R85A mutant

The Arginine finger residue of RhoGAP, arginine 85, was mutated to an alanine. The pGEX-2T expression vectors carrying human p29RhoGAP were provided by Dr. Katrin

Rittinger (MRC, National Institute for Medical Research, London). p29RhoGAP comprised the carboxyl terminal fragment of p50 RhoGAP (Lancaster *et al.*, 1994). The mutagenesis experiment was performed by Dr. Y Jin. Details are described in her PhD thesis (Jin, 2012). In this thesis, RhoGAP_{R85A} will refer to this mutant.

5.3.2.2 RhoA and RhoGAP_{R85A} expression and purification

The two recombinant proteins were cloned into the same vector. Thus, the expression and purification were carried out as described for RhoA. Expression after IPTG induction was carried out ON, at 20°C in Rosetta 2 cells. A GST-trap column was used for the first purification step. Tag cleavage was performed with thrombin on the column. As a final purification step, size exclusion chromatography was performed (see Chapter 2), with SEC buffer was prepared with Bis/Tris pH 7. This did not affect the size exclusion chromatography profile of RhoA. In the case of RhoGAP_{R85A}, the protein eluted as monomer at an elution volume of 56 ml, with a 260/280 ratio of 0.6 (Figure 5.9). The yield of the entire process was typically 1.6 mg/L culture.

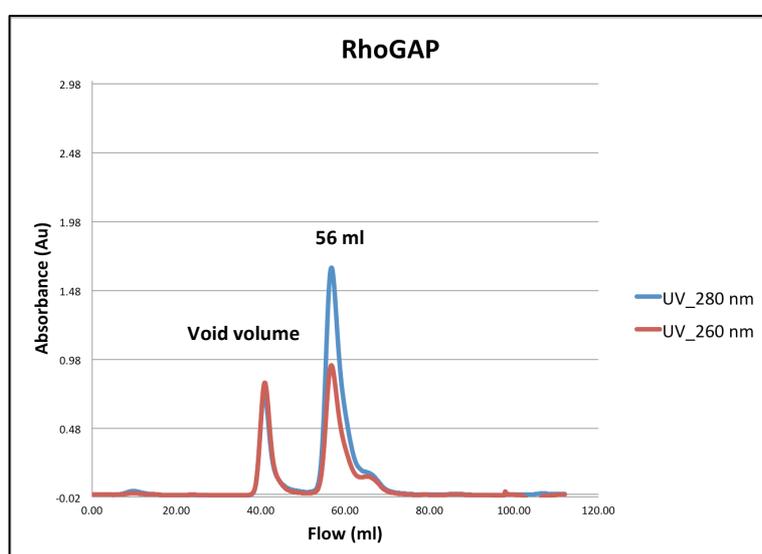


Figure 5.9 Size exclusion chromatography profile of RhoGAP. From 6 L culture, 10 mg of protein was eluted in a unique peak. The absorbance at 260 nm is shown in *red* and 280 nm in *blue*. The chromatogram was run with a HiLoad 16/60 Superdex 75 pg column.

5.3.2.3 Generation and crystallisation of the RhoA-GDP-AlF_x-RhoGAP_{R85A}-TSA complex

A unique stock of the metallofluoride protein complex was prepared every time immediately after the size exclusion chromatography experiments. The two proteins were mixed in equimolar amounts and 20 mM NaF and 2 mM AlCl₃ from concentrated stock solutions (0.5 M) were added to the mix. Usually, the initial complex concentration was around 11 μM. The complex was incubated overnight at 4°C and then concentrated to 230 μM for crystallisation experiments. The complex concentration was monitored by Bradford assay. After the first successful crystallisation experiments it was established that 230 μM corresponds to a concentration of 8 mg/ml.

At first, published conditions were tested (Rittinger *et al.*, 1997). Sea urchins and thin plates in unique clusters were obtained. Since separation of the plates from the sea urchins was challenging it was decided to find new crystallisation conditions by using the service provided by the HTX Laboratory. Three samples were prepared for this experiment: the metallofluoride complex and the single proteins mixed with aluminium fluoride and concentrated as in the heterodimer sample. The last two samples act as control experiments, which allowed the definition of whether crystals were of the complex or of individual components. Several efforts were made to reproduce all the hits obtained. Eventually, showers of microcrystals and plate clusters were obtained from sitting drop experiments over a reservoir containing 100 mM Bis/Tris pH 5.8 or pH 6 and PEG 3350 23-26% (w/v).

Optimisation was required to produce crystals useful for data collection. The combination of the microbatch method and seeding techniques proved successful. Microbatch experiments were set up using sitting drop plates (D'Arcy *et al.*, 2004; Del Campo & Lambowitz, 2009). The protein sample was mixed at different ratios and at different concentrations with the reservoir solution described for the sitting drop experiments. In this way, the optimal microbatch conditions were established: 2 μl of sample plus 2 μl reservoir, covered by 25 μl of paraffin oil. The protein concentration used was the same as for the sitting drop. In this way thicker crystals were obtained, but they still formed in clusters or were very thin. Three generations of seeding were required to obtain good crystals. Microseeding was carried out through the preparation of a stock seeding solution prepared by crushing a few crystals (often a cluster) and

diluting them in 50 μl of reservoir solution. Drops were prepared in two different ways: Seed crystals were transferred to the microbatch drop with an otter whisker or by substituting partial reservoir volumes with the seed stock (typically 0.5-0.25 μl) (Bergfors, 2003; Patrick D. Shaw Stewart 2011). The progression of these experiments was monitored in two ways: by eye using a microscope with polarised lens and by diffraction quality. Diffraction experiments were carried out at the microfocus beamline ID23-2, ID14-4 or at ID29, according to the crystal size. Each time data collection was performed in order to check for any resolution improvement. In the end, single and well diffracting crystals were obtained (Figure 5.10). The final crystals obtained were in different conditions from those published for the WT complex (RhoA-GDP-AIF_x-RhoGAP-TSA, Rittinger *et al.*, 1997).

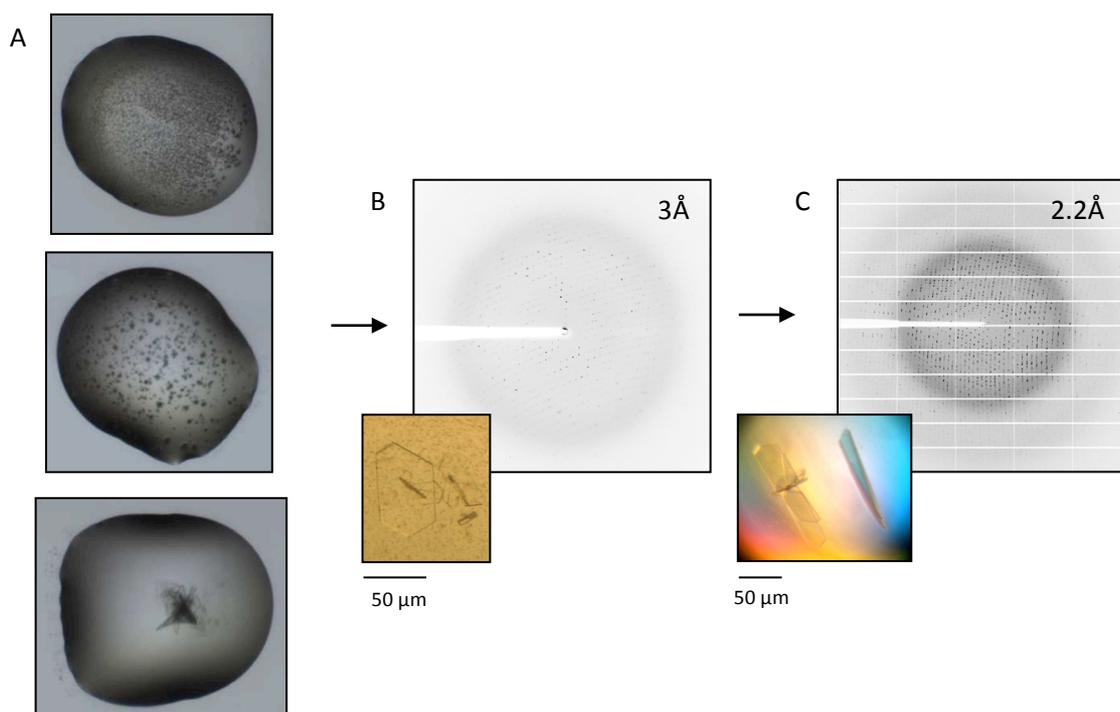


Figure 5.10 RhoA-GDP-AIF_x-RhoGAP_{R85A}-TSA crystal optimisation. Images of the crystals and the diffraction patterns show the main steps in the optimisation progress. A) Sea urchins and clusters obtained from the robot screening in 100 mM Bis/Tris pH 5.5-6.5, 20-25% (w/v) PEG 3350; similar initial hits were obtained by repeating in sitting drop plate the same conditions. B) Improvements obtained after one round of microseeding experiments in microbatch. Diffraction data to 3 Å resolution were collected at ID23-2, on a MAR 225 CCD detector. At this step the crystals were thin and up to 30 μm in the largest dimension. C) Single crystals that diffracted to higher resolution were obtained after further seeding experiments. The data were collected at ID29, with a Pilatus 6M detector.

5.3.2.4 Generation and crystallisation of RhoA-GDP-MgF_x-RhoGAP_{R85A}-TSA complex

The generation and the crystallisation of the magnesium fluoride complex was carried out in similar manner to that described for the aluminium fluoride complex. The complex was prepared by adding 20 mM MgCl₂ and concentrating to 30 mg/ml as indicated by a Bradford assay. Sitting drop experiments were carried out in the conditions described for the aluminium fluoride complex. As before, no good crystals were obtained. Thus, microbatch and microseeding experiments were also used in this case. Conditions were screened and eventually the complex was crystallised at 15mg/ml with 2 µl of the sample were added to 4 µl of the reservoir. Paraffin oil (20 µl) was used to cover the drop. Again, at least three microseeding generations were necessary for growing useful crystals.

5.3.2.5 Data collection, structure solution and refinement

The crystals obtained from microbatch experiments were prepared for flash cooling by immersion in the mother liquor made up with 25% PEG 400. Diffraction data from RhoA-GDP-AlF_x-RhoGAP_{R85A}-TSA crystals at 2.2 Å resolution were collected at 100K, at beamline ID29. Diffraction data from RhoA-GDP-MgF_x-RhoGAP_{R85A}-TSA crystals at 2.1 Å resolution were collected at 100K, at beamline ID14-4. Reflections were integrated with iMOSFLM and XDS (Battye *et al.*, 2011; Kabsch, 2010). The structure of the complex was solved by molecular replacement using the previous structure (PDB code 1TX4 or 1OW3; Rittinger *et al.*, 1997; Graham *et al.*, 2002) as a search model with the bound ligands and water molecules removed. Ligands were included after a few refinement cycles. Refinement and validation were performed as described in Chapter 2.

5.3.2.6 Sulphur-SAD data collection

In order to determine the identity of ions, a long wavelength experiment was performed at 14.2 beamline at the BESSY Synchrotron, in Berlin. The experiment was run at a remote wavelength from the adsorption edge of Sulphur or Phosphorus (see Appendix

A). At this wavelength, the observation of the anomalous scattering of light elements becomes possible, such as chloride and potassium, because the adsorption edge is much closer with regard to the usual data collection wavelength of 0.97Å.

Since the anomalous scattering signal of light elements is low at the wavelengths achievable at MX synchrotron beamlines, the success of the experiment depends on the combination of several factors. Conditions for cryocooling crystals should be optimised in order to avoid excess liquid around the crystal and on the loop. This is important to reduce the X-ray absorption effects, which will increase the background scattering (Kitago *et al.*, 2005; Teng, 1990). The correct or 'best' wavelength for the experiment should be considered carefully. For protein structure determination, it has been suggested to be in the range of 1.5 - 3 Å (Chayen *et al.*, 2000; Weiss *et al.*, 2001). Finally, the redundancy of the data required should be considered. High redundancy allows more accurate determination of the anomalous differences, which in the end will allow a better location of anomalous scatters (Dauter *et al.*, 2002; Harrop *et al.*, 1996; Weiss *et al.*, 2001).

The long wavelength experiment was performed on the same crystals used for data collection described previously. The wavelength was set to 2Å (E = 6.2 keV) and two collections of 360 degrees were performed in order to have high redundancy. The anomalous map was calculated after the first and the second collection. For the second map the two data collection datasets were merged together and the possible benefit from using more data was judged. The diffraction images were integrated with XDS (Kabsch, 2010). After scaling, F_s and phases were calculated using SFALL from the ccp4 suite (Brunger, 1992). The program required as input the *mtz* files with separated F_{hkl} and F_{-h-k-l} , and the *pdb* file of the native protein. A final map was generated using a fast Fourier transform, through FFT in the ccp4 suite (Read, 1988). Inspection of the peaks in the resulting map was performed by eye using the molecular modelling program Coot (Emsley & Cowtan, 2004; Emsley *et al.*, 2010).

5.4 Results and Discussion

5.4.1 RhoA-GDP-AlF₃-TSA crystallisation

RhoA-GDP-AlF_x-TSA has primarily been studied in our laboratory for investigating the role of the charge balance hypothesis in small G proteins. As described in the general introduction, the removal of a positive charge inside the active site of a protein, which is relevant in transition state stabilisation, should be compensated by a corresponding decrease in charge of the transition state analogue. Since Arg85_{RhoGAP} is a positively charged residue involved in the stabilisation of the transition state in RhoA-GDP-AlF₄⁻-RhoGAP-TSA and RhoA-GDP-MgF₃⁻-RhoGAP-TSA, the charge balance principle was investigated by studying the RhoA-GDP-AlF_x-TSA, where the Arg85_{RhoGAP} is absent. This complex is also of interest, because it is not a dead mutant. RhoA has an intrinsic GTPase activity and the RhoA elements that perform this reaction and stabilise the transition state when RhoGAP is absent have not been defined (see Chapter 1).

Structural studies of the RhoA-GDP-AlF_x-TSA complex proved difficult. ¹⁹F-NMR data showed the formation of an aluminium trifluoride complex in solution. NMR spectra, recorded by Dr Jin Yi, showed three peaks, corresponding to an aluminium trifluoride species with no net charge (Jin, 2012) (Figure 5.11). This confirmed that the expected RhoA-GDP-AlF₃ complex is formed in solution. As demonstrated for PGK, it is proposed that an aluminium atom coordinated in octahedral geometry by three fluorides and one water molecule is the bound species (Cliff *et al.*, 2010). However, only a crystal structure could define the bound aluminium fluoride species.

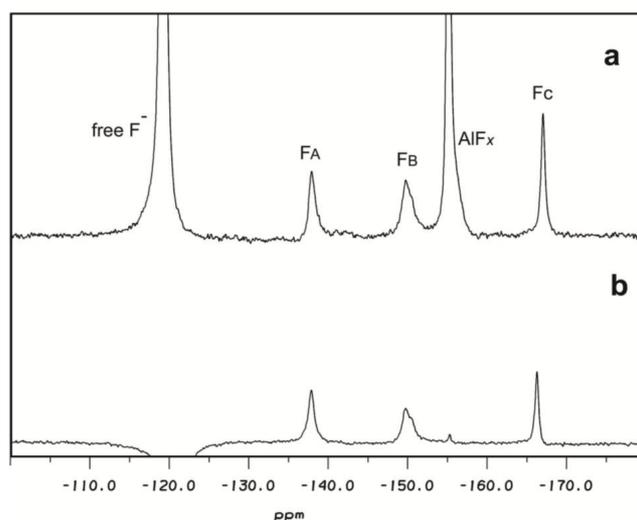


Figure 5.11 ^{19}F -NMR spectra of RhoA-GDP- AlF_3 -TSA complex. Spectra of a) RhoA-GDP- AlF_3 complex in 50 mM Tris at pH 7.5, containing 0.5 mM RhoA (with 1 eq. GDP bound), 5 mM MgCl_2 , 1 mM DTT, 10 mM NH_4F , 1 mM AlCl_3 using pulse program *zg*. b) The same sample using presaturation pulse program, *zgpr* to saturate the free F^- signal at -119 ppm to remove the problem of non-protein bound free F and AlFx resonances in fast exchange. $F_A = -137.9$ ppm, $F_B = -149.9$ ppm, and $F_C = -166.9$ ppm (Jin, 2012).

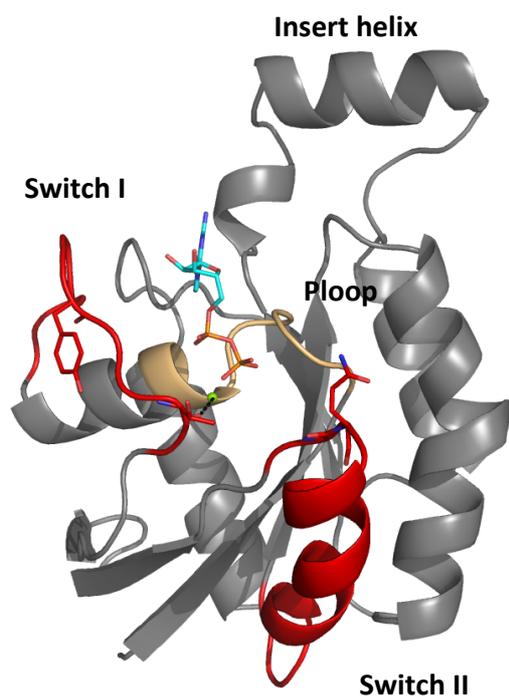
Referring to the RhoA_{G14V}-GTP γ S structure (PDB code 1A2B; Ihara *et al.*, 1998), the aluminium fluoride was expected to be in the position of the γP , coordinated by the Tyr34 and Gln63 residues belonging to the Switch I and Switch II regions respectively. The Switch regions were then expected to be in the Rho-GTP bound state. However, the crystallisation of RhoA-GDP- AlF_3 -TSA resulted in a RhoA-GDP structure that crystallised in the orthorhombic space group $P2_12_12_1$, with one molecule in the asymmetric unit (Table 5.3). Thus, the Switch regions crystallised in Rho-GDP conformation (Figure 5.12). None of the peaks in the difference map around the active site could be assigned to an aluminium trifluoride moiety. Inspection of the electron density maps showed no peaks corresponding to bound aluminium fluoride in other locations in the protein. Further efforts to reproduce the NMR results in the solid state were unsuccessful.

Table 5.3 Data collection and refinement statistics

	RhoA-GDP
Space group	<i>P2₁2₁2₁</i>
Wavelength (Å)	0.976
Unit cell dimensions (Å) <i>a,b,c</i>	31.80, 66.27, 83.68
Resolution range (Å)¹	35.38 (1.37-1.3)
Number of unique reflections	44107
Multiplicity	4.6 (4.5)
Completeness (%)	99.5 (99.3)
R_{merge}	0.037 (0.265)
<1/σ(I)>	20.3 (5.1)
Wilson B factor (Å²)	15.5
Water molecules	207
R_{factor} (%)	15.56
Free R_{factor} (%)	18.43
RMS deviations:	
Bonds (Å)	0.013
Angles (°)	1.636

¹Values for higher resolution shell are in brackets

The absence of bound aluminium fluoride was probably due to dynamic properties of the RhoA-GDP-AlF₃-TSA sample. This makes a scenario likely in which the Switch regions involved in aluminium fluoride coordination and the aluminium fluoride itself are continually in motion in solution. In the solid state, where one of the two states should be fixed, the Rho-GDP conformation was stabilised by crystal packing. Each RhoA molecule interacts through hydrophobic interactions with the symmetry related molecules in the crystal. The benzene group of Tyr34 from the Switch I of one molecule is indeed packed between the benzene group of Phe134 and the Met157 side chain of another molecule (Figure 5.13).



	Overall	P loop	SwitchI	SwitchII
1A2B	1.31	0.25	1.91	1.65
1FTN	0.24	0.07	0.13	0.23

Figure 5.12-Table 5.4 The structure of RhoA-GDP at 1.3Å resolution. In this ribbon representation of RhoA-GDP, Switch regions and further relevant regions are highlighted as in figure 5.2: Switch I (residues 27-36) and Switch II (residues 61-78) are in *red*, the catalytic magnesium is shown as a *green* sphere, GDP is in *cyan* sticks and the P-loop is in *brown*. In the table C α rmsd values are shown between RhoA_{G14V}-GTP γ S (PDB code 1A2B; Ihara *et al.*, 1998) and RhoA-GDP (PDB code 1FTN; Wei *et al.*, 1997). Rmsd xyz displacement was calculated with the program *Superpose*, from CCP4 suite (Krissinel & Henrick, 2004).

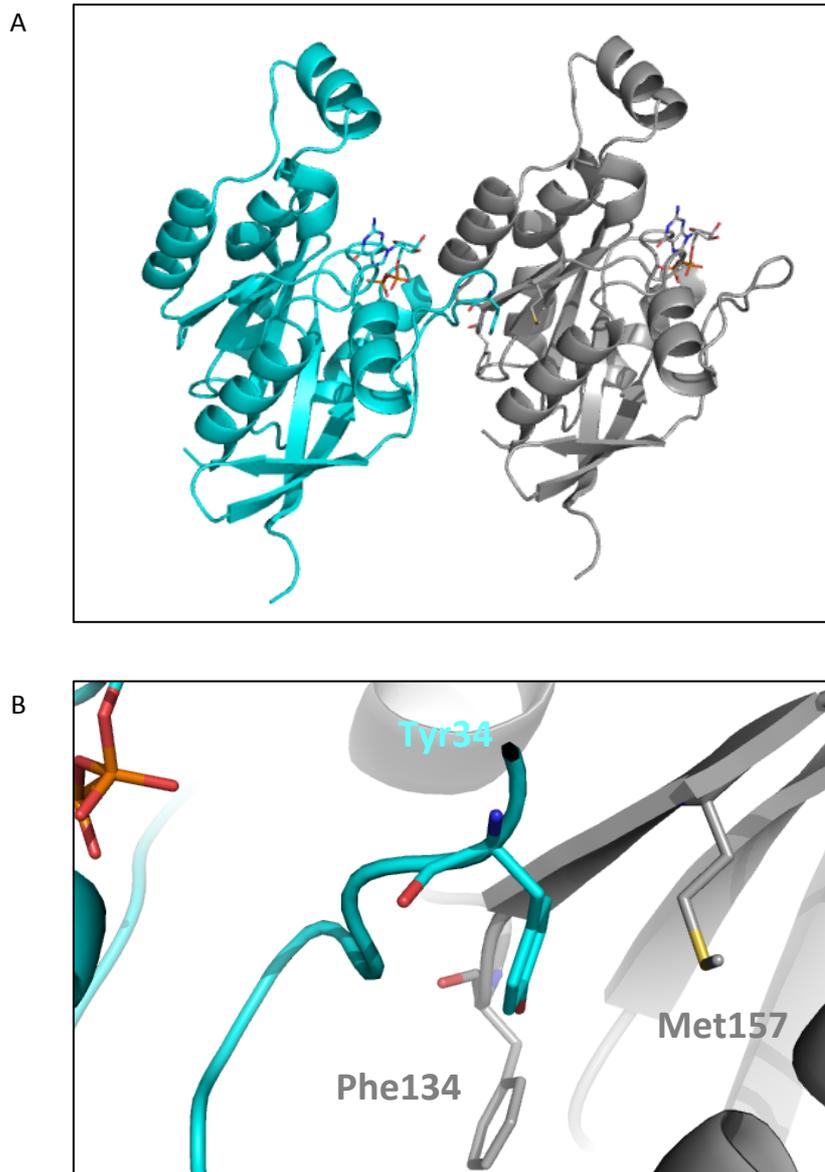


Figure 5.13 Crystal packing contacts between RhoA molecules. The relationship between two molecules in the crystal is shown (A). Details of the interaction are shown in the panel (B). Residues involved in symmetry related molecules interactions that stabilise the switch in the ADP bound state, are shown in sticks.

Nevertheless, we obtained a further structure of RhoA-GDP at higher resolution (1.3 Å) than previously obtained 1FTN, 2.1 Å (Wei *et al.*, 1997) . Surprisingly, electron density was visible for key catalytic residues that have occupancy zero in the published structure. Differences in the Switch II region are particularly interesting. While in 1FTN Gln63 is still inside the active site, RhoA-GDP shows that Gln63 no longer interacts

with the active site when the nucleotide is hydrolyzed (Figure 5.14). Thus the RhoA-GDP structure solved in this thesis, provides a complete and detailed snapshot of the RhoA conformation in the GDP-bound state. In the next paragraphs the name RhoA-GDP will refer to the structure solved in this thesis.

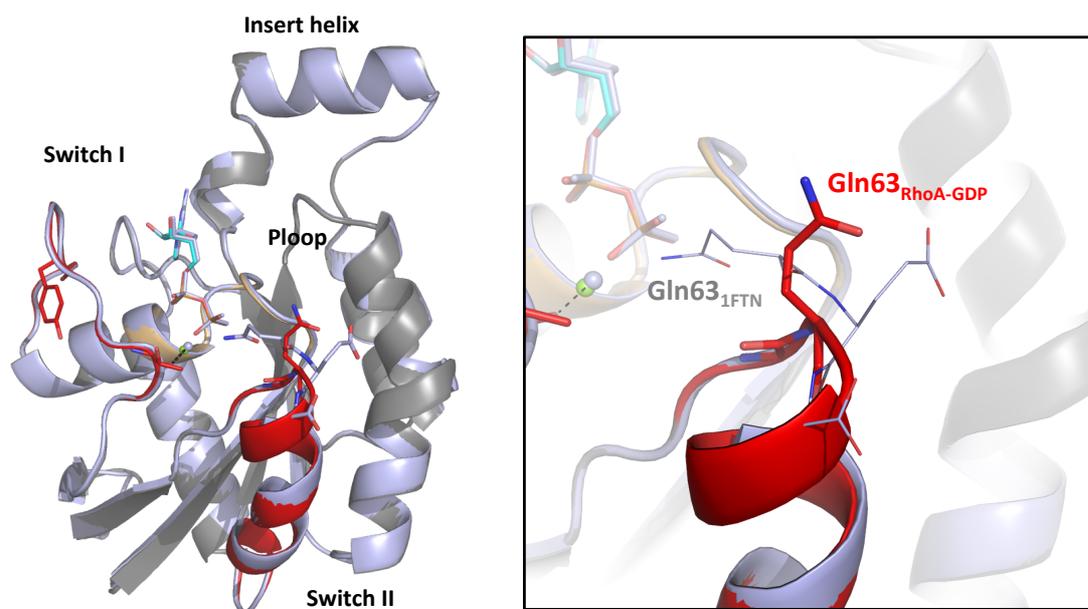


Figure 5.14 Structure of Switch II region in RhoA-GDP and RhoA-GDP-1FTN. In the overlay the structure of the Switch II region in Rho-GDP is in *red*, while Rho-GDP-1FTN is in *light blue*. The panel magnifies the difference in conformation of Gln63 in both structures. Gln63 is shown in sticks. Gln63 has occupancy zero in 1FTN structure. The overlay is on the Switch I region.

5.4.2 The structures of RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA

Since the crystallisation of RhoA-GDP-AlF₃ failed, it was decided to study the charge balance theory in a small G protein, by crystallising RhoA-GDP-AlF_x/MgF_x-TSA in complex with RhoGAP_{R85A}, where Arg85 has been mutated to an alanine. As in the case of RhoA-GDP, this complex is still active and it was also of interest to investigate which are the transition state features of GTP-hydrolysis performed by this complex.

The RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA and the RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA complexes crystallised in the monoclinic space group *P2*₁ with 2 molecules per

asymmetric unit. The structures were solved to 2.2 and 2.1 Å resolution respectively (Table 5.5).

The two proteins assumed the same relative orientation as in the WT complex RhoA-GDP-AlF₄⁻-RhoGAP-TSA structure (PDB code 1TX4; Rittinger *et al.*, 1997) or in the RhoA-GDP-MgF₃⁻-RhoGAP-TSA structure (PDB code 1OW3; Graham *et al.*, 2002) with an rmsd between C α atoms of 0.5 and 0.7 Å respectively. The difference Fourier maps showed clear peaks for the expected ligands (GDP and aluminium fluoride) and the absence of electron density for the Arg85_{RhoGap} side chain (Figure 5.15). Surprisingly, density for the backbone of residues 31-33 in the RhoA effector loop was clearly visible in both structures. These residues are disordered in RhoA-GDP-AlF₄⁻-RhoGAP-TSA structure. Both aluminium and magnesium are coordinated by the same number of fluorides atoms as in the WT structures: aluminium is coordinated by four fluorides in octahedral geometry, while magnesium is coordinated by three fluorides in trigonal bipyramidal geometry, despite the removal of a charged residue from the active site.

Table 5.5 Data collection and refinement statistics

	RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA	RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA
Space group	<i>P2</i> ₁	<i>P2</i> ₁
Wavelength (Å)	0.976	0.939
Unit cell dimensions		
<i>a, b, c</i> (Å)	72.55 , 66.06, 76.76	70.53, 68.24, 74.37
<i>α, β, γ</i> (°)	90.00, 96.00, 90.00	90.00, 93.24, 90.00
Resolution range (Å)¹	48.72 (2.32-2.2)	50.24 (2.21-2.1)
Number of unique reflections	36,197	41,361
Multiplicity	3.0 (3.0)	2.4 (239)
Completeness (%)	98.5(98.6)	91(90.9)
R_{merge}	0.082 (0.529)	0.153 (0.507)
<1/σ(I)>	8.5 (2.1)	4.7 (2.1)
Wilson B factor (Å²)	33.1	27.45
Water molecules	159	273
R_{factor} (%)	21.76	21.6
Free R_{factor} (%)	26.63	27.55
RMS deviations:		
Bonds (Å)	0.01	0.017
Angles (°)	1.395	1.692

¹Values for higher resolution shell are in brackets

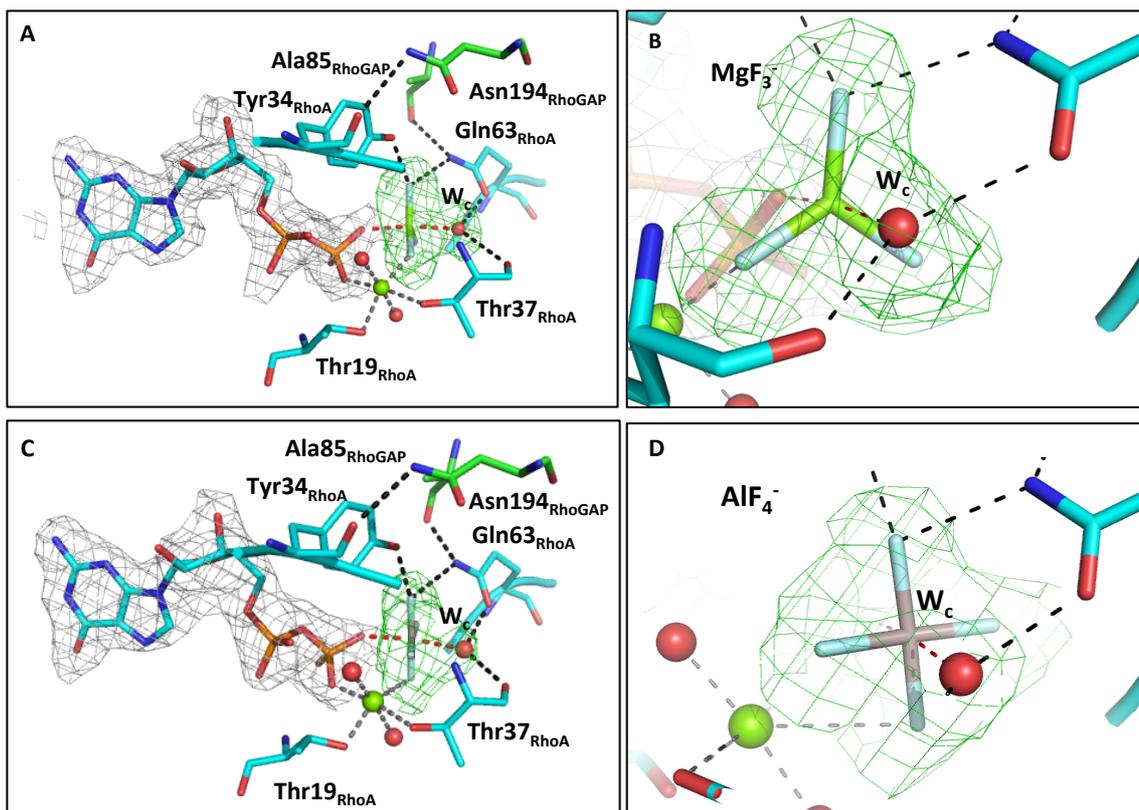


Figure 5.15 Difference electron density in the active site of the structures of RhoA-GDP- AlF_4^- -RhoGAP_{R85A}-TSA and RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA. The difference Fourier map for each fluoride species (*in green*) at a counter level of 3σ is shown (Fourier map in *grey* contoured at 1σ). At this resolution the catalytic water molecule (W_c) is included in the same continuous electron density. A-B) RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA. C-D) RhoA-GDP- AlF_4^- -RhoGAP_{R85A}-TSA. Details about residues, dashed lines and colours are explained later in the chapter.

At first sight this is in contrast with the charge balance hypothesis, in which the loss of a positive charge should cause a concomitant reduction of negative charge in the bound ligand. This result was confirmed by ^{19}F -NMR, in which the AlF_4^- is an average peak, the integration of which shows four atoms are present, and the MgF_3^- species show three peaks (Figure 5.16).

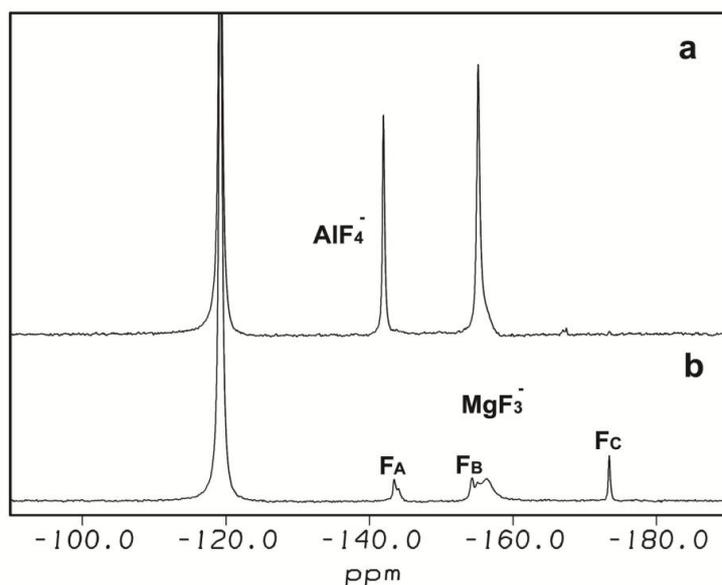


Figure 5.16 ^{19}F -NMR spectra of the RhoA-GDP- AlF_4^- -RhoGAP_{R85A}-TSA and RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA complexes. Spectra of a) RhoA-GDP- AlF_4^- -RhoGAP_{R85A}-TSA complex in 50 mM Tris at pH 7.5: 0.5 mM RhoA (with 1 eq. GDP bound), 5 mM MgCl_2 , 1 mM DTT, 10 mM NH_4F , 1 mM AlCl_3 . The rotationally averaged peak has a chemical shift at -142.0 ppm and its integration shows there are 4 fluorines in the complex. b) RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA complex which shows three peaks at $F_A = -143.4$ ppm, $F_B = -154.3$ ppm, and $F_C = -173.4$ ppm (Jin, 2012).

The next section will discuss the importance of these structures in the understanding of catalysis and will try to investigate how the charge balance hypothesis can be applied to this particular system. Since RhoA-GDP- AlF_4^- -RhoGAP_{R85A}-TSA and RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA show similar features, RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA will be used as model in the following paragraphs. This because MgF_3^- is a better transition state analogue (see Chapter 1) and the structure was determined at higher resolution.

5.4.3 RhoA-GDP- MgF_3^- -RhoGAP_{R85A}-TSA and Rho-GDP structures provide novel insight into the GTP-hydrolysis reaction

Through the comparison of the RhoA-GDP- MgF_3^- -RhoGAP_{R85A} structure with the RhoA_{G14V}-GTP γ S (PDB code 1A2B; Ihara *et al.*, 1998) and RhoA-GDP- MgF_3^- -RhoGAP (PDB code 1OW3; Graham *et al.*, 2002) structures, the GTP-hydrolysis transition state is clearly defined by two consecutive steps: a stabilisation step and a

catalytic step. In the first step, the binding of RhoGAP to RhoA induces the conformational changes necessary for the reaction start. In the second, the Arginine finger stimulates RhoA activity. The hydrolysis reaction can be carried out using either step one or step two, this asks the question: what are the necessary components for catalysis and how is charge balanced? Further dramatic changes occur at the Switch regions when RhoGAP releases RhoA, which moves out from the active site residues involved in catalysis.

5.4.3.1 The stabilisation step

The stabilisation step will be described by comparing the RhoA_{G14V}-GTP γ S, RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-MgF₃⁻-RhoGAP-TSA structures.

In order to carry out the reaction, RhoA requires the binding of RhoGAP to stabilise the Switch I (residues 28-38) and Switch II (residues 61-78) regions.

The Switch regions are the protein segments mainly involved in this step. As described, they contain several key residues for the coordination of the transition state and are involved in the interaction between the proteins. The binding of RhoGAP causes conformational changes in the Switch regions that induce fundamental changes in the coordination of the γ P or transition state analogue. In RhoA_{G14V}-GTP γ S, the triphosphate moiety is coordinated by the side chains of Tyr₃₄_{RhoA}, Lys₁₈_{RhoA}, Gly₆₂_{RhoA}, Thr₃₇_{RhoA} and by the Mg²⁺ ion. The catalytic water molecule is located at a position 10° away from the hypothetical line for nucleophilic attack, at 3.6 Å from the phosphorous atom (Ihara *et al.*, 1998) (Figure 5.17).

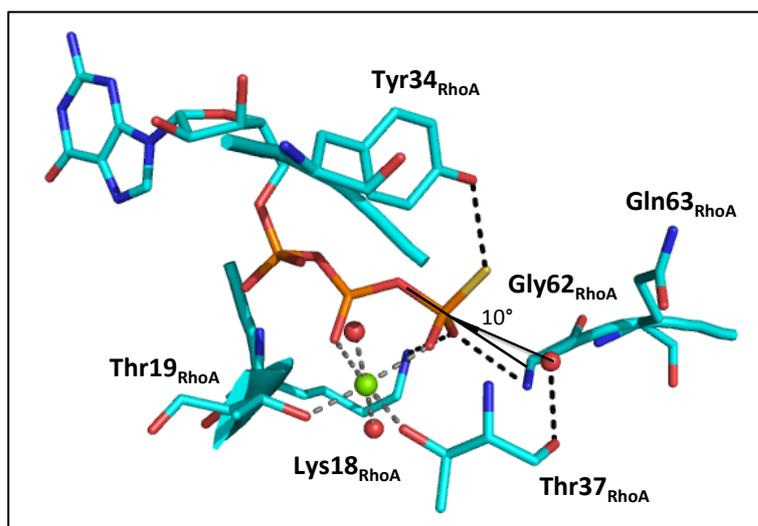


Figure 5.17 GTP γ S bound to RhoA in the RhoA_{G14V}-GTP γ S structure. The GTP γ S and the predicted catalytic water molecule coordination are shown. Potential H-bonds are shown as *black* dashed lines. *Black* lines represent the proposed line for the nucleophilic attack and the distance between the phosphorous atom and the predicted catalytic water molecule.

It has been proposed that the formation of the complex occurs through the interaction of the GAP protein with the Switch regions. In particular, the Switch II region seems to act as a pivot point around which rigid-body motions occur on the progression from the ground state to the transition state (Rittinger *et al.*, 1997). This causes conformational changes in the Switch I region, where the Tyr34_{RhoA} is in an open conformation. This is clearly shown by the overlay of the two structures (Figure 5.18). However, the analysis of the superimposition of RhoA_{G14V}-GTP γ S protein with RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA, leads to different conclusions. RhoGAP stabilises the Switch I and II regions through several hydrogen bonds. The Switch I region is stabilised by the interaction between Asn194_{RhoGAP} and Tyr34_{RhoA}. The binding causes a slight rotation of the phenol ring and maintains the closed conformation rather than the open one. Indeed, the side chain of Asn194_{RhoGAP} interacts with the backbone of Tyr34, stabilising the OH---O bond with the tri-phosphate (Figure 5.18). The importance of this interaction has already been described in previous work. In the Cdc42-GDP-AlF₃-Cdc42_{R305A}-TSA structure this binding has already been described (Nassar *et al.*, 1998).

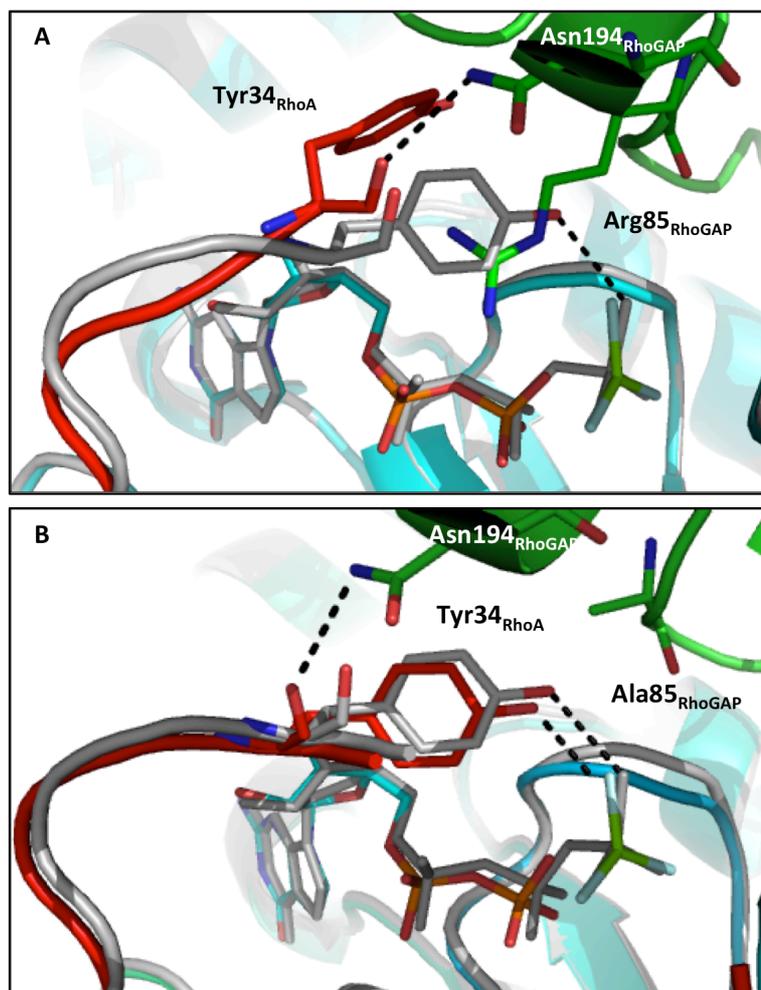


Figure 5.18 Stabilisation and conformational changes of Tyr34_{RhoA} in the Switch I region. The Switch I region in RhoA-GDP-MgF₃⁻-RhoGAP-TSA and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA is shown in *red*. A) Overlay of RhoA_{G14V}-GTPγS (in *grey*) and RhoA-GDP-MgF₃⁻-RhoGAP-TSA (RhoGAP is in *green* and RhoA in *cyan*). The Tyr34_{RhoA} progresses from the closed conformation to the open one. B) Overlay of RhoA_{G14V}-GTPγS (in *grey*) and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA (RhoGAP is in *green* and RhoA in *cyan*). Tyr34_{RhoA} is stabilised in the closed conformation. The H-bonds are represented with a *black* dashed line. In both RhoA-RhoGAP structures the amine group of Asn194_{RhoGAP} side chain interacts with the carbonyl of Tyr34_{RhoA} backbone. Residues 35-38 from the Switch I region have been deleted for clarity.

In the case of the Switch II region, GAP binding causes a shift of a segment comprised of residues 61-69 through the formation of hydrogen bonds between the two proteins. As previously described, Lys122_{RhoGAP}, and Asn202_{RhoGAP} interact with Asp65_{RhoA} while Val197_{RhoGAP} and Asn220_{RhoGAP} interact with Tyr66_{RhoA}. As a result of this shift, the side chain of Gln63_{RhoA} moves in a way that aligns the catalytic water molecule in line with the magnesium atom (analogous to the phosphorus it is a surrogate for) and it

coordinates the transition state. Ala85_{RhoA} backbone is hydrogen bonded with Gln63_{RhoA} side chain. Furthermore both Asp65_{RhoA} and Tyr66_{RhoA} stabilise this position by interacting with the backbone carbonyl oxygen atom of Gln63_{RhoA} (Figure 5.19).

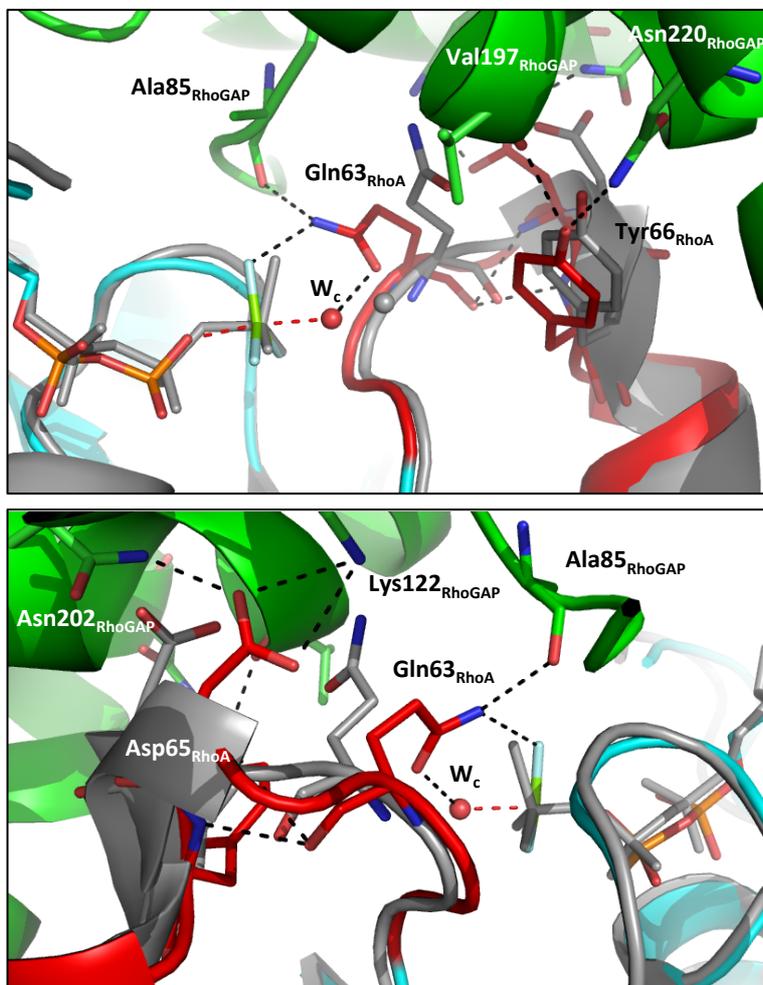


Figure 5.19 Stabilisation and conformational changes in the Switch II region. A cartoon representation of the Switch II region (in red) conformational changes induces in RhoA (in cyan) by RhoGAP (in green), in the structure of RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA. RhoA is overlaid with RhoA_{G14V}-GTPγS (in grey) for comparison. Black dashed lines represent hydrogen bonds. Red dashed lines represent the interactions between βP-O, MgF₃⁻ and the catalytic water molecule. Lys122_{RhoGAP}, Asn202_{RhoGAP}, Val197_{RhoGAP}, Asn220_{RhoGAP}, Ala85_{RhoGAP}, Asp65_{RhoA}, Tyr66_{RhoA} and Gln63_{RhoA} are represented in sticks. The same view is shown from two different orientations, 180° apart. Residues 35-38 from Switch I region have been deleted for clarity. W_c is the catalytic water molecule.

This scenario appears to be in contrast with previous theories. As shown in figure 5.20, it has been proposed that in the RhoA-RhoGAP system there is an active role of the Arginine finger in positioning Gln63_{RhoA} and the catalytic water for the nucleophilic attack (Rittinger *et al.*, 1997). The RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA structure shows that this event is independent of the presence of the Arginine finger and that after binding the catalytic water molecule is in line with the magnesium atom. We observe an H-bond between the Gln63_{RhoA} side chain and the Ala85_{RhoGAP} backbone, but even if there is a connection between the two residues, this stabilisation interaction is independent from the nature of Arg85.

It is not possible to describe precisely the starting position of this residue, due to a bulky Val14 side-chain in RhoA_{G14V}-GTP γ S. RhoA_{G14V} has been described as a dominant mutant and the Val maintains the Gln63_{RhoA} too far from here to be able to start the reaction. Moreover, the bulky sulphur atom sterically shields the phosphorus atom from the close approach of the nucleophile water molecule and interfere with the stabilisation of the transition state by Gln63_{RhoA} (Ihara *et al.*, 1998). This observation does not conflict with the evidence for the absence of an active role of the Arg85_{RhoGAP} side chain in the positioning of the Gln63_{RhoA}/catalytic water molecule.

Thus, the stabilisation step is the first event required for activation of intrinsic RhoA GTPase activity. The GAP protein performs an allosteric function, leading to the reaction starting through hydrogen bond formation in the Switch I and the Switch II regions. After binding, Gln63_{RhoA} and the catalytic water molecule are correctly positioned for nucleophilic attack.

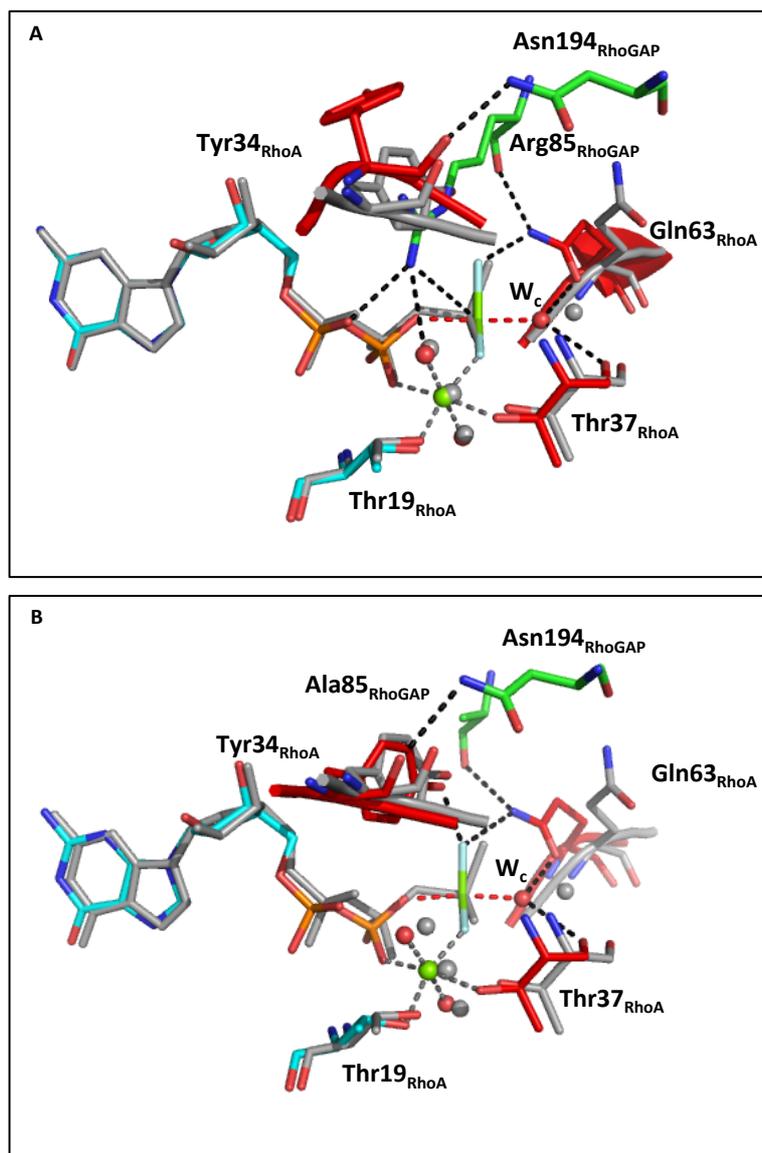


Figure 5.20 Allosteric effects of RhoGAP binding. The contribution of the binding effect of RhoGAP are shown by comparing the previous structures, and the one proposed with the RhoA-GDP-MgF₃⁻-RhoGAP-TSA structure. A) The stick representation of the overlay between the active site of RhoA_{G14V}-GTPγS (in *grey*) structure and RhoA-GDP-MgF₃⁻-RhoGAP-TSA (RhoGAP is in *green* and RhoA in *cyan*) structures shows the effect on the Switch regions (in *red*) and the catalytically water molecule caused by the entrance of the Arginine finger. B) The stick representation of the overlay between the active site of RhoA_{G14V}-GTPγS (in *grey*) structure and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA (RhoGAP is in *green* and RhoA in *cyan*) structures show the events independent of the Arginine finger. Colours and dashed lines are similar to previous figures. *Grey* dashed lines represent the magnesium coordination. W_c is the catalytic water molecule.

5.4.3.2 The catalytic step

This section describes the catalytic steps through the comparison of RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-MgF₃⁻-RhoGAP-TSA. Once the stabilisation event has occurred, the Arginine finger displaces Tyr34_{RhoA} and stabilises it in an open conformation (Figure 5.21). As previously described, the Arginine finger interacts with the nucleotide phosphates (α P and γ P/MgF₃⁻), stabilising the leaving group by decreasing the negative charge around the γ P. Both Lys18_{RhoA} and Gly62_{RhoA} maintain the interactions with the γ P (or MgF₃⁻), as described in the RhoA_{G14V}-GTP γ S structure. No further residue movements occur, except for Switch I region displacement caused by the conformational change of Tyr34_{RhoA}.

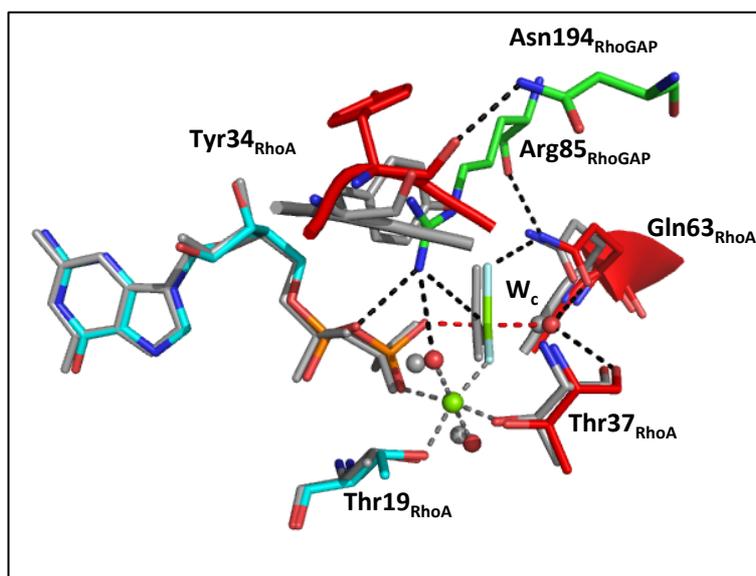


Figure 5.21 The Arginine finger flips out the Tyr34_{RhoA}. A stick representation of the overlay between RhoA active site (in *cyan*) from RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA (in *grey*) and RhoA-GDP-MgF₃⁻-RhoGAP-TSA (RhoA in *cyan* and RhoGAP in *green*) structures. Residues from Switch regions are in *red*.

5.4.3.3 The releasing step

Once the reaction has occurred, RhoGAP releases RhoA in a GDP-bound state. In order to analyse the final conformational changes, the structure of RhoA-GDP solved in this thesis to higher resolution (1.3 Å), is used. The overlay between the RhoA-GDP-MgF₃⁻-

RhoGAP and the RhoA-GDP shows a large conformational change in the Switch regions (Figure 5.22). In particular, after catalysis, both the Tyr34_{RhoA} and Gln63_{RhoA} residues no longer form part of the active site. This conformation is stabilised by a rotation of the Thr37_{RhoA} in the Switch I region. Indeed, Thr37_{RhoA} still coordinates the magnesium atom, but through the main chain instead of the side chain.

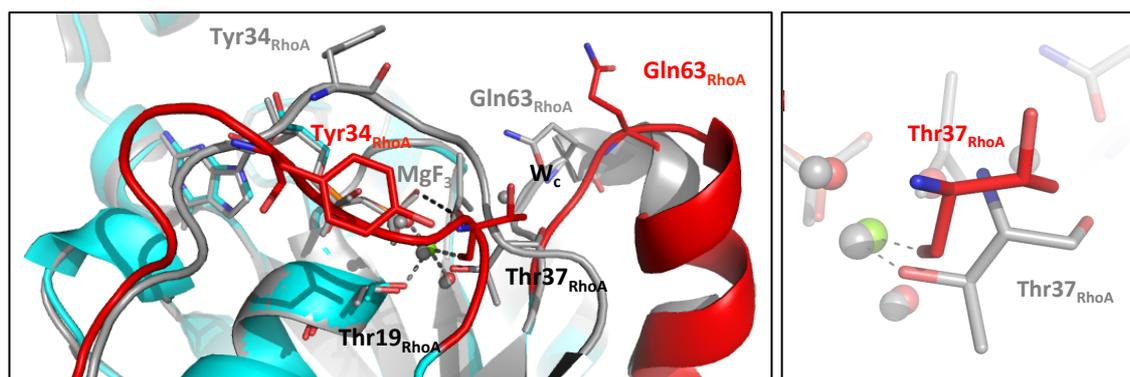


Figure 5.22 Dramatic changes of the Switch regions in RhoA-GDP. A cartoon representation of the Switch regions (in red) with RhoA-GDP (in cyan) and RhoA from RhoA-GDP-MgF₃⁻-RhoGAP-TSA (in grey). Tyr34_{RhoA}, Thr37_{RhoA} and Gln63_{RhoA} are shown in sticks. The left panel magnifies the rotation of Thr37_{RhoA}. Colours and dashed lines are similar to previous figures.

5.4.4 Residual activity in RhoA and RhoA-RhoGAP_{R85A}

The previous section has provided a complete and detailed scenario of the conformational changes that occur from the ground state to the transition state that enhance the hydrolysis of GTP in RhoA. Interestingly, as well as the Arginine finger and Gln63_{RhoA}, a third relevant residue has been described: Tyr34_{RhoA}, which is the same tyrosine that in other Ras proteins such as Ran and Rap is known to be involved in catalysis. In RhoA-GTP and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA Tyr34_{RhoA} is in a closed conformation, coordinating one of the oxygen atoms of the GTP-γP. Since RhoA has an intrinsic GTPase activity and RhoGAP_{R85A} is still active, it was decided to investigate how these systems could hydrolyse the nucleotide and how Tyr34_{RhoA} is involved in the reaction.

5.4.4.1 The intrinsic activity of RhoA

Despite the formation in solution of the RhoA-GDP-AlF₃-TSA complex, no structural data are available. Thus, only a hypothesis can be proposed on how this basal hydrolysis works. The RhoA_{G14V}-GTP γ S structure shows that the sulphur atom in the γ P interacts with the hydroxyl group of Tyr34_{RhoA} (in the closed conformation) and the amine group of Lys18_{RhoA}, while both Gln63_{RhoA} and the catalytic water molecule are too far from the γ P to allow nucleophilic attack. Similar O₃- γ P coordination has been found in two other structures: the RhoA_{Q63L}-GMPPNP structure (PDB code 1KMQ; Longenecker *et al.*, 2003), which is another permanently GTP-bound state RhoA protein, and in the RhoA of the 3KZ1 structure, where RhoA is in complex with a GEF domain (Chen *et al.*, 2010). Thus, the RhoA-GTP active site is likely the one described by these three structures. In this scenario Gln63_{RhoA} could move occasionally in a way to place the water molecule in line with the γ P, at distances that allows associative or dissociative phosphoryl transfer. In this catalysis Tyr34_{RhoA} is likely stabilising the transition state, by working as a pseudo-arginine: it makes an H-bond interaction with the γ P group and positive charge is localized in a way that would increase the electronegative interaction with it. This could be enough for stabilising the negative charge around the leaving group in a reaction that happens occasionally.

5.4.4.1 The RhoGAP_{R85A} residual activity

RhoGAP_{R85A} is still able to accelerate RhoA GTP-hydrolysis 160-fold (Graham *et al.*, 1999). As the structures of RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA show, the catalysis is independent of the Arginine finger and no other residues are provided by the GAP. The charge balance theory suggests that the mutant system should compensate the loss of the positive charge with a corresponding unit reduction in negative charge through the loss of one fluoride from the magnesium or aluminium fluoride moiety in the mutant complex (see Chapter 1). However, both fluoride species maintain their charge and the Arginine finger is replaced by a tyrosine. As said above, tyrosine could work as pseudo-arginine. However, this is not enough to stabilise the negative charge around the leaving group and thus allow efficient catalysis.

Two main hypotheses are proposed for explaining how the RhoGAP_{R85A} accelerates the reaction. The first proposes the presence of a counter ion, which decreases the nucleotide negative charge, by interacting with the phosphate nucleotide region. The second proposes a tightening effect achieved in the RhoA active site upon RhoGAP_{R85A} binding.

5.4.4.1.1 Potential counter ions around α -phosphate

Looking at the sequence of structures RhoA_{G14V}-GTP γ S, RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-MgF₃⁻-RhoGAP-TSA, it is evident that the water molecules around the nucleotide increase in number and change in position and residue coordination (Figure 5.23), with the exception of the ones coordinating the magnesium ion. The same observation could be made on the sequence of structures of RhoA_{G14V}-GTP γ S, RhoA-GDP-AlF₄⁻-RhoGAP_{R85A}-TSA and RhoA-GDP-AlF₄⁻-RhoGAP-TSA. However, for the aluminium fluoride structures the difference in resolution is too large to make them comparable.

The displacement of the water molecules is relevant because they change the charge environment around the nucleotide in the progression from the ground state to the transition state. Two water molecules characterise GTP γ S solvation in RhoA_{G14V}: one is involved in the base coordination and the other is associated with the 2'-hydroxyl group of the ribose.

Upon RhoGAP binding, further water molecules interact with the nucleotide phosphate region. The major difference between the WT and the R85A mutant structure is the interaction between the α - and β -phosphates with the Arg85_{RhoGAP} and surrounding water molecules. The Arg \rightarrow Tyr mutation changes the solvation subtly around the oxygen of α -phosphate on the side of the mutated area. This introduces the possibility that the water molecule coordinating one of the oxygen atoms of the α -phosphate is a counter ion instead of a water molecule. The presence of a counter positive charged ion in RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA structure would re-establish charge balance. A water molecule in the same position is also present in the corresponding aluminium fluoride structure. However, due to the limited resolution of the mutant structure, it is

difficult to identify the nature of the atoms or their protonation state according to the electron density.

An attempt was made to identify any potential ions by collecting a further dataset of RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA crystal at an energy of 6 keV (2 Å wavelength). 2 Å is a long wavelength, which allows observation of the anomalous scattering of light elements, such as sulphur or phosphorous. Moreover, as the adsorption edge of lighter elements such as chloride and potassium is much closer with regard to the usual data collection wavelength of 0.97Å, it is possible to define whether the assigned water molecules are ions. The long wavelength data collection experiment was performed at the BESSY II Synchrotron in Berlin, at a beamline dedicated to these experiments (14.2). However, the anomalous map obtained was too noisy to identify sulphur or phosphorus atoms in the anomalous difference Fourier maps. This may have been due to the crystal conditions, where crystals were cryo-cooled in a large amount of paraffin oil belonging to the microbatch experiment. The oil may have disturbed the experiment, by increasing X-ray adsorption and thus preventing the correct measurement of the anomalous differences, which is a critical point in a long wavelength collection experiment.

Thus, no conclusions could be made about the presence of a counter ion in the RhoA active site. However, the change in solvation is a clear event, which is affecting in some way the reaction.

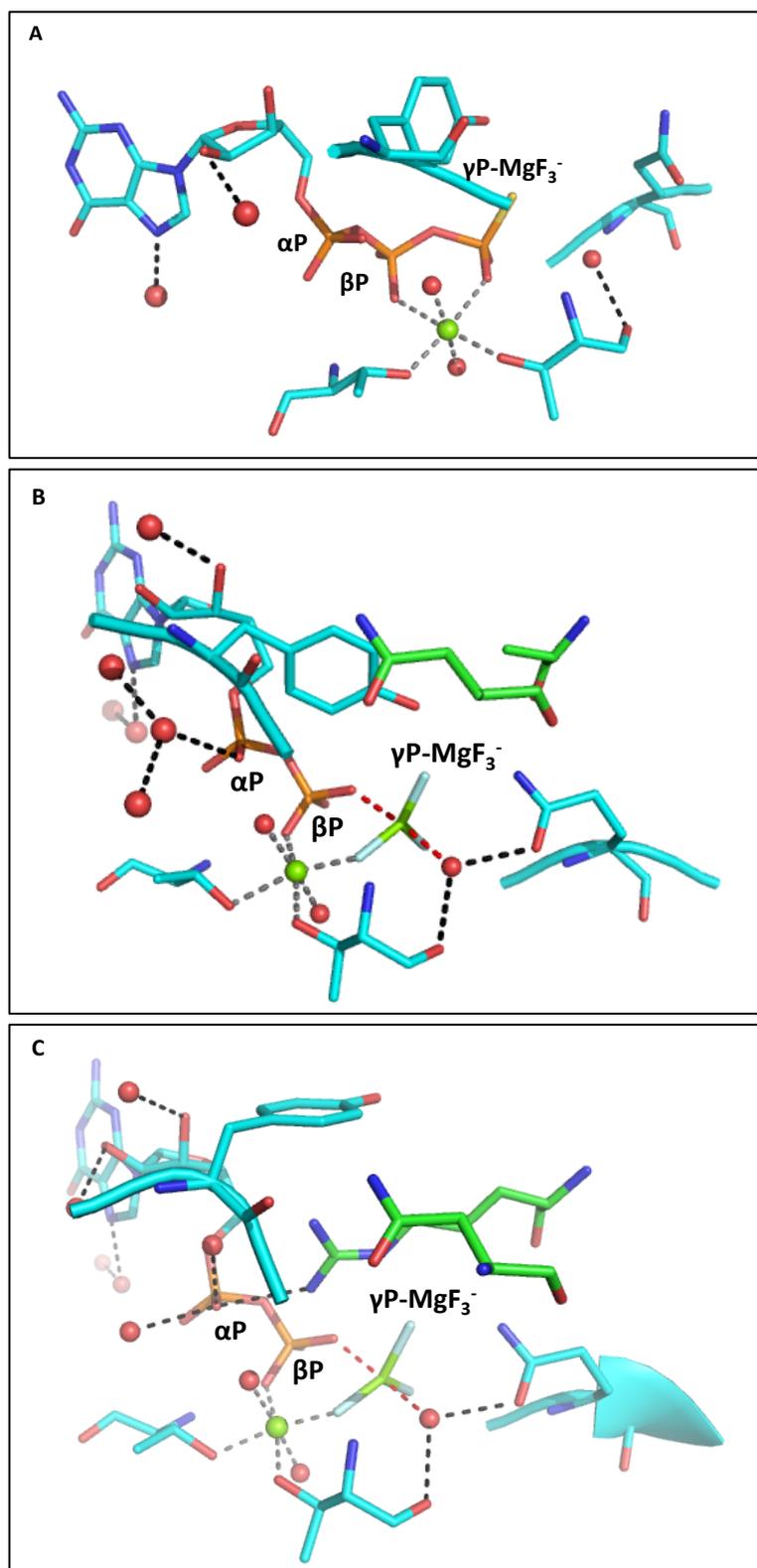


Figure 5.23 GTP solvation. The three panels represent the nucleotide solvation in RhoA-GTPγS (A), RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA (B) and RhoA-GDP-MgF₃⁻-RhoGAP-TSA (C). Colours and dashed lines are similar to previous figures.

5.4.4.1.2 The tightening effect around γ -phosphate during catalysis

As described above, the binding of RhoGAP causes the correct positioning of Gln63_{RhoA} in order to align the catalytic water molecule for nucleophilic attack. New hydrogen bonds are created around the γ P-MgF₃⁻. Upon binding F_C (equivalent to the S atom in the RhoA_{G14V}-GTP γ S structure) is coordinated by the catalytic water molecule and the amino group of Gln63_{RhoA}. The catalytic water molecule is in line at 2.12 Å distance from the Mg_{TSA} and the new interaction restricts its freedom. All the hydrogen bonds present in RhoA_{G14V}-GTP γ S structure are still present. Surprisingly, hydrogen bonds in RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA are shorter and much more similar to the ones in RhoA-GDP-MgF₃⁻-RhoGAP-TSA (Table 5.6). This tightening effect is induced upon RhoGAP binding; it induces stability in the RhoA active site and probably decreases the energy barrier as described by enzyme definition. This effect could be enough for accelerating the GTP-hydrolysis of 160-fold, without further leaving group stabilisation.

Table 5.6 Tightening effects. The table shows the hydrogen bonds coordinating the oxygen or fluorine atoms of the γ P or Mg_{TSA} (magnesium belong to the TSA). *Red* distances represent the distances between the catalytic water molecule and γ P/ MgF_3^- . SC (side chain) or BB (back bone) indicate where the chemical group is place in the residue. W_M are the waters coordinating the magnesium. W_C is the catalytic water molecule.

O_1/F_A	Lys18 _{RhoA}	Gly62 _{RhoA}		
RhoA _{G14V} -GTPyS	NH-2.9 Å (SC)	NH-3.31 Å (BB)		
RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA	NH-2.66 Å (SC)	NH-2.68 Å (BB)		
RhoA-GDP-MgF ₃ ⁻ -RhoGAP-TSA	NH-2.63 Å (SC)	NH-2.69 Å (BB)		

O_2/F_B	Mg	W_{M1}	W_{M2}	T37 _{RhoA}
RhoA _{G14V} -GTPyS	2.92 Å	3.1 Å	2.9 Å	O-3.1 Å (SC) NH-3.15 Å (BB)
RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA	1.83 Å	3.1 Å	2.76 Å	O-2.95 Å (SC) NH-2.9 Å (BB)
RhoA-GDP-MgF ₃ ⁻ -RhoGAP-TSA	1.94 Å	2.95 Å	2.92 Å	O-2.84 Å (SC) NH-2.69 Å (BB)

S/F_C	Y34 _{RhoA}	Wc	Q63 _{RhoA}	R85 _{RhoGAP}
RhoA _{G14V} -GTPyS	OH-2.92 Å (SC)	-	-	-
RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA	OH-2.48 Å (SC)	2.82 Å	NH-2.81 Å (SC)	-
RhoA-GDP-MgF ₃ ⁻ -RhoGAP-TSA	-	2.85 Å	NH-2.73 Å (SC)	NH-2.6 Å (SC)

Wc	Q63	T37	γ P/ Mg_{TSA}
RhoA-GTPyS	-	O-2.8 Å (SC)	3.55 Å
RhoA-GDP-MgF ₃ ⁻ -RhoGAP _{R85A} -TSA	O-2.61 Å (SC)	O-3.1 Å (SC)	2.13 Å
RhoA-GDP-MgF ₃ ⁻ -RhoGAP-TSA	O-2.73 Å (SC)	O-2.89 Å (SC)	2.11 Å

5.5 Conclusions

The structures solved here have brought novel insights into the RhoGAP mediated RhoA GTP-hydrolysis. In particular, through the overlay of four different structures, representing four different steps in the reaction, how RhoGAP contributes to the reaction has been analysed. Figure 5.24 shows a sequence of four snapshots that summarize the main changes occurring in the entire reaction around the $\gamma\text{P}/\text{MgF}_3^-$. Similar snapshots can be created using the corresponding aluminium fluoride structure. First, RhoGAP binds RhoA through the recognition of the Switch regions. The binding stabilises the Switch I region in the RhoA-GTP conformation, while inducing a shift in the Switch II region. Thus, it appears that Switch I region, rather than Switch II, acts as a pivot point around which RhoGAP induces conformational changes in RhoA on progression from the ground state to the transition state. As result of this, the catalytic water molecule is moved close enough to the gamma phosphate for nucleophilic attack to occur. This event is independent of the Arginine finger and it is caused by an allosteric effect from GAP binding. Only after this step does the RhoGAP enter in an active role, by displacing Tyr34_{RhoA} with the Arg85_{RhoGAP}. Finally, upon RhoGAP release, both Switch regions no longer interact with the active site. The RhoA-GDP conformation is stabilised by the rotation of Thr37_{RhoA}. Thus, the Gap protein acts actively only in the catalytic step, stimulating the intrinsic RhoA activity, while it performs an allosteric function during the stabilisation step.

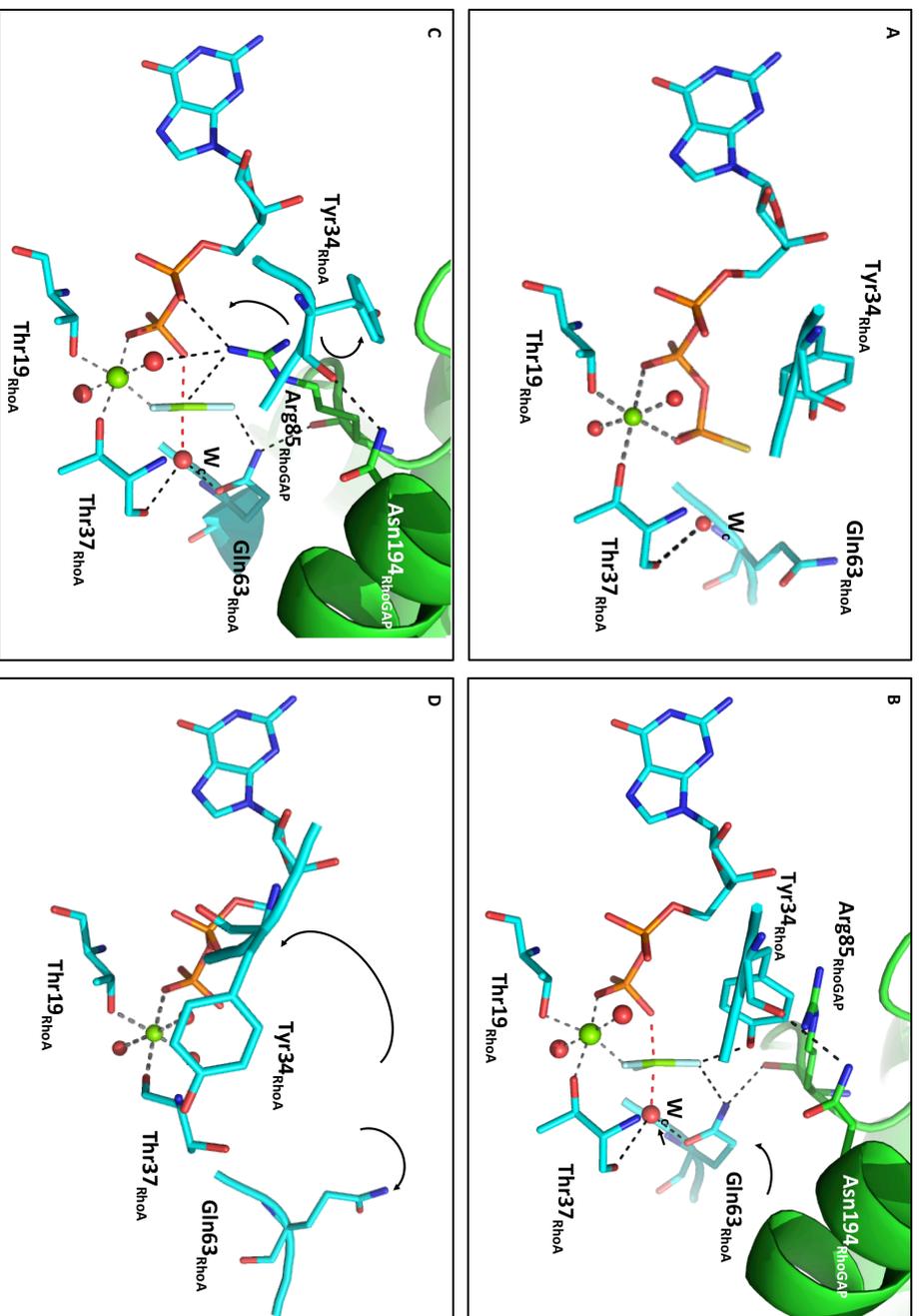


Figure 5.24 Snapshots of the RhoGAP mediated RhoA GTP-hydrolysis. A cartoon representation of the three steps reaction around the γP , built through the overlay of RhoA_{G14V}-GTP γS , RhoA-GDP-MgF₃-RhoGAP_{R85A}-TSA, RhoA-RhoGAP-GDP-MgF₃-RhoGAP-TSA, RhoA-GDP and RhoGAP (PDB code 1RGP; Barrett *et al.*, 1997). RhoA protein is represented in *cyan*, RhoGAP in *green*, Mg molecule in *green* and water molecules in *red*. A) RhoA is in the GTP bound state. B) RhoGAP binds RhoA. C) The Arg85_{RhoGAP} finger flips out the Tyr34_{RhoA}, stabilising it in open conformation; the reaction can be carried out. D) RhoGAP releases RhoA.

In this scenario, a new role has been described for Tyr34_{RhoA}. As has already been observed by Scrima A. et al., Tyr34 in RhoA is highly conserved across the main Ras proteins: Ras (Tyr32), RhoA (Tyr34), Cdc42 (Tyr32), Rap (Tyr32) and Ran (Tyr39). Upon GAP binding, the tyrosine assumes either an open or closed conformation, based on which further residues will be involved in the GTP-hydrolysis reaction. In the case of RhoA and Cdc42 it has been demonstrated that when the Arginine finger is missing, RhoGAP_{R85A} is still able to increase the hydrolysis rate of Rho-GTP by 160-fold, (Graham *et al.*, 1999), while the similar mutant Cdc42GAP_{R305A} conserves 20% of its activity (Nassar *et al.*, 1998). In these cases tyrosine remains in the closed conformation, assuming a possible role in catalysis. Indeed, taking into account the pseudo-arginine role that the tightening effect upon RhoGAP binding, the reaction could still work. Whether a counter ion is present is still an open question. Further long wavelength data collection experiments with better cryocooled crystals will be performed for answering this question.

However, in the case of the small GTP proteins from the Rho family, the Arginine finger is required in order to obtain maximum reaction efficiency, leaving doubtful a role for the tyrosine in catalysis under physiological conditions. Considering the cycle described in figure 5.24, it appears that Tyr34_{RhoA} is more involved in regulation of the switching cycle rather than transition state stabilisation. Indeed, Tyr34_{RhoA} together with Arg85_{RhoGAP} and Thr37_{RhoA} perform the RhoA RhoGAP mediated GTP-hydrolysis. Tyr34_{RhoA} side chain is the hallmark, which defines where the Arginine finger should enter. When Tyr34_{RhoA} coordinates the γ P, the Switch I region is in closed conformation. The entry of the Arginine finger moves the Switch I region in the direction of the GDP bound conformation (semi-open conformation). Finally the rotation of Thr37_{RhoA} prevents Tyr34_{RhoA} from coming back in the active site (open conformation). Thus while the catalysis is working, the switch cycle is progressing from the GTP bound state to the GDP bound state (Figure 5.25).

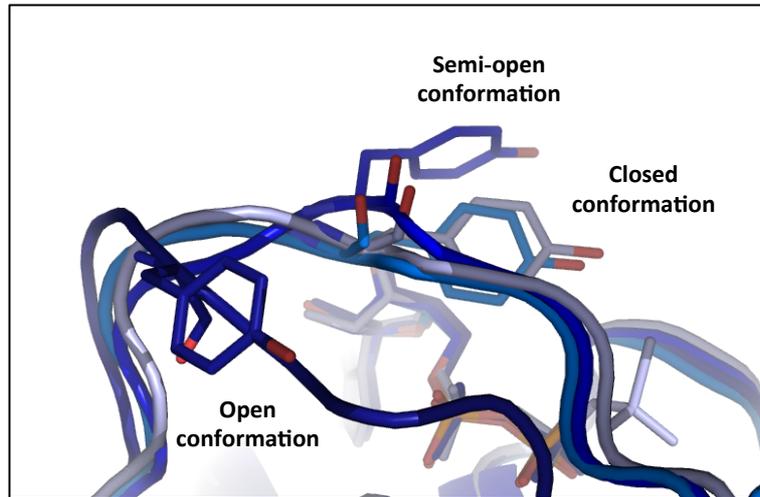


Figure 5.25 Switch I conformational changes from the GTP-bound state to the GDP-bound state. The conformational changes assumed by the Switch I loop are shown. In RhoA_{G14V}-GTP γ S and RhoA-GDP-MgF₃⁻-RhoGAP_{R85A}-TSA Switch I is in closed conformation; in the RhoA-GDP-MgF₃⁻-RhoGAP-TSA structure is in semi-open conformation, while it is in open conformation in the RhoA-GDP structure.

The relevance of Tyr34_{RhoA} and Thr37_{RhoA} in this cycle is highlighted by the fact that both residues are targets of bacterial toxins and virus enzymes. In case of infection by the viruses *H. somnus* and *V. parahaemolyticus*, enzymes mediate AMPylation at Tyr34_{RhoA} and Thr37_{RhoA}, respectively (Worby *et al.*, 2009; Yarbrough *et al.*, 2009). AMPylation leads to the inactivation of the GTPase activity by preventing the interaction with downstream effectors, thereby inhibiting actin assembly in infected cells. Thr37_{RhoA} in the Rho family proteins is the target of *Clostridium difficile* toxins A and B, which monoglucosylate it and maintained the small G protein in off-state (Just *et al.*, 1996).

In conclusion, through the structures presented here, evidence has been provided that define how RhoGAP stimulates RhoA GTP-hydrolysis. Firstly a structural proof has been provided for the fundamental allosteric role of RhoGAP, which induces the proper orientation and polarisation of the catalytic water molecule. Secondly, the possible elements involved in RhoA intrinsic GTP-hydrolysis and the residual activity of RhoGAP were found. Thirdly it was defined the fundamental role in the switching cycle of the conserve Tyrosine in RhoA-RhoGAP system, rather than a possible role in catalysis. This series of events is probably also occurring in the similar proteins Ras and Cdc42.

Chapter 6 The effect of the MEK6 docking interaction on p38 α

6.1 Introduction

6.1.1 MAPKs belong to the MAPKs cascade, a druggable pathway

Mitogen activated protein kinases (MAPKs) are ubiquitous signalling molecules, which belong to the so called MAPK cascades. MAPK signalling confers switch-like responses to a wide variety of extracellular stimuli that lead to cell fate decisions such as differentiation, proliferation, apoptosis and senescence (Chen *et al.*, 2001; Deng *et al.*, 2004; Johnson & Lapadat, 2002).

MAPK signalling is characterised by a cascade of multiple kinases, through which the signal is transduced in the form of phosphorylation events from the upstream kinase to the downstream one. MAPK kinase kinases (MAPKKKs) activate MAPK kinases (MAPKKs), which in turn phosphorylate the MAPK (Figure 6.1). This last phosphorylation event is a dual phosphorylation, which occurs on a conserved threonine and a conserved tyrosine residue in the phosphorylation lip (P-Lip) or activation loop (Anderson *et al.*, 1990). The same sites are the targets of dual specificity tyrosine and serine/threonine phosphatases that will switch off the activated kinase via dephosphorylation (Dickinson & Keyse, 2006; Saxena & Mustelin, 2000). The activated MAPK will activate downstream proteins, such as transcription factors or others protein kinases (Tanoue *et al.*, 2001).

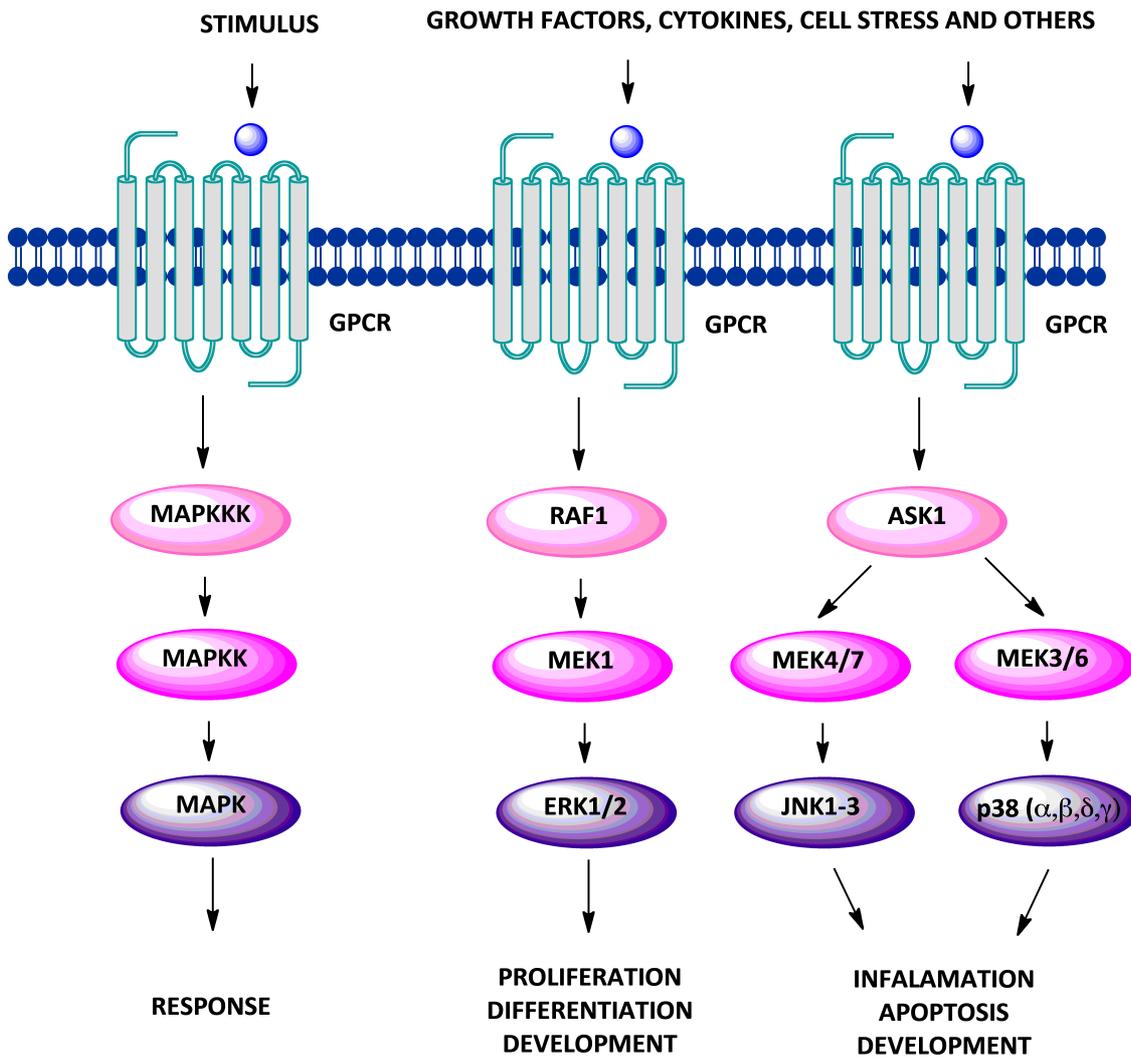


Figure 6.2 MAPK cascades in mammals. The cartoon shows how the MAPK cascade is activated upon ligand binding to a G-protein-coupled receptor (GPCR). Activation of mitogen activated protein kinase (MAPK) cascades proceeds by the sequential activation of a MAPK kinase kinase (MAPKKK), a MAPK kinase (MAPKK) and a MAPK. The cascade for ERK, JNK and p38 is then shown. ERK is activated mainly by growth factors and phorbol esters, while JNK/SAPK and p38 are activated by extracellular stresses, such as UV irradiation and osmotic stress, and by inflammatory cytokines (Garai *et al.*, 2012; Ip & Davis, 1998; Kyriakis & Avruch, 1996; Laughlin *et al.*, 2012; Nishida & Gotoh, 1993) (modified from Pierce & Lefkowitz, 2001).

There are three major MAPK subgroups described in vertebrates: ERK, p38 and JNK/SAPK (Tanoue *et al.*, 2001). The three families have 40 to 50% sequence identity (Caffrey *et al.*, 1999) and each one contains several isoforms (Figures 6.1-6.2).

Each MAPK is activated by a specific upstream MAPKK, through the common mechanism described above (Figure 6.1). The conserved threonine and tyrosine are

within a Thr-*X*-Tyr motif, where *X* is proline, glutamate, and glycine for ERK, JNK, and p38, respectively (Cobb & Goldsmith, 1995; Davis, 1995). The phosphorylation activates MAPK by 100- to 1000-fold, and is necessary to achieve maximum enzymatic activity (Johnson *et al.*, 1996; Payne *et al.*, 1991; Robbins & Cobb, 1992; Robbins *et al.*, 1993).

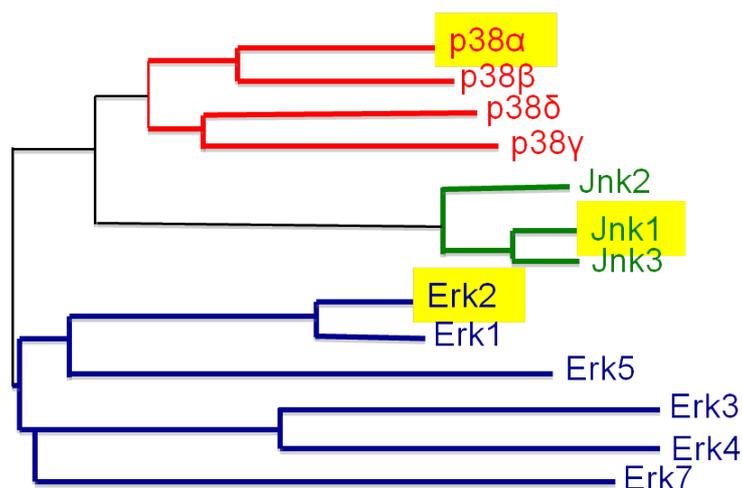


Figure 6.2 The Human MAPK family tree. The family tree of human MAPKs was built through local alignment on the EXPASY web site using ClustalW (www.expasy.org; Thompson *et al.*, 1994). Isoforms highlighted in *yellow* are structurally characterised.

MAPK cascades are involved in several human pathologies and many drugs are already in development. ERK2 pathway components are drug targets for cancer (Roberts & Der, 2007; Sebolt-Leopold & Herrera, 2004). JNKs are drug targets for apoptosis related diseases such as Alzheimer's disease, Parkinson's disease, type II diabetes, hearing loss, and also for autoimmune diseases (Manning & Davis, 2003; Resnick & Fennell, 2004). The p38 MAP kinase pathway is a therapeutic target for inflammatory diseases such as psoriasis, rheumatoid arthritis and chronic obstructive pulmonary disease (Kaminska, 2005; Kumar *et al.*, 1995). Currently, p38 α , one of the four isoforms of p38, is under study for the treatment of rheumatoid arthritis, pain relief, dementia and stroke (Adams *et al.*, 2001; Poornam *et al.*, 2009; Schindler *et al.*, 2007).

In the case of p38 α , most of these drugs bind competitively with the nucleotide in the

ATP binding site (Adams *et al.*, 2001; Kaminska, 2005). Only a few bind to a region adjacent to the ATP binding site, which is termed the DFG-out site. The term “out”, indicates the conformational changes in this motif induced by drug binding (Regan *et al.*, 2003; Wroblewski & Doweyko, 2005).

However, all these drugs appear to have broad specificity: ATP is used by all kinases, and DFG is highly conserved across all the protein kinase family.

6.1.2 Docking interactions are relevant in MAPK-specific drug design

The understanding of how MAPKs interact specifically with upstream and downstream signalling proteins has highlighted a possible specific MAPK target.

Through mutagenic analysis (Bardwell *et al.*, 2001; Barsyte-Lovejoy *et al.*, 2002; Gum & Young, 1999; Tanoue *et al.*, 2000; Tanoue *et al.*, 2001; Zuniga *et al.*, 1999) and crystallographic studies (Chang *et al.*, 2002; Heo *et al.*, 2004; Liu *et al.*, 2006; Remenyi *et al.*, 2005; Zhou *et al.*, 2006) it has been defined that MAPK interacting proteins, recruit MAPKs by binding at loci outside the active site, through an almost linear peptide docking motif called the D-motif (Enslin *et al.*, 2000; Kallunki *et al.*, 1996; Zuniga *et al.*, 1999). The longest D-motifs characterised have the consensus sequences $X-\phi_H-X_2-(\text{Arg/Lys})_{1-2}-(X)_{2-6}-\phi_A-X-\phi_B$ (where ϕ_A , ϕ_B and ϕ_H are hydrophobic residues (Leu, Ile, or Val) (Gavin & Nebreda, 1999; Smith *et al.*, 1999), while shorter motifs can lack the first four residues (Sharrocks *et al.*, 2000) (Table 6.1).

Table 6.1 Structure-based sequence alignment of the D-motifs. The basic residues and ϕ residues are shown in *blue* and *orange* letters, respectively (modified from Chang *et al.*, 2002).

Docking protein	Docking site	MAPK
MEK1	KKKPTP-IQL-NPNP	ERK
MEK2	RKPVLPA-LTI-NP	ERK
LIN-1	GMPNP-LNL-TATS	ERK
SAP-1	RSKKPKG-LGL-APT	ERK
ELK-1	KGRKPRD-LEL-PLSPS	p38
MEK6	SKGKKRNPGLKI-PKEA	p38
MEK3b	GSKRKKD-LRI-SCNS	p38
MEF2A	RKPD-LRV-VIPPSS	p38
MEF2C	RKPD-LRV-LIPPGS	JNK
MEK7	ARRRID-LNL-DISP	JNK
NFAT4	LERPSRDH-LYL-PLEP	JNK
C-JUN	KILKQSMT-LNL-AD	JNK
JIP-1	RPKRPTT-LNL-FPQVP	JNK, p38
ATF-2	VHKHKHE-MTL-KFGPA	ERK, p38
PTP-SL	GLQEERRGSN-VSL-TLDM	ERK, p38
HePTP	RLQEERRGSN-VAL-MLDV	ERK, p38
STEP	GLQEERRGSNVS-LTL-DM	
	$R/K_{1-2}-(X)_{2-6}-\phi_A X \phi_B$	

The binding site of this motif on MAPKs is named the D-motif binding site. This site is formed by an acidic patch known as the **CD domain** (for Common Docking) (Tanoue *et al.*, 2000) and a **hydrophobic docking groove** (Chang *et al.*, 2002; Gum & Young, 1999). The CD domain accommodates the ϕ_H and basic residues, while the hydrophobic docking groove is supposed to bind ϕ_A and ϕ_B (Akella *et al.*, 2010; Heo *et al.*, 2004; Lee *et al.*, 2006).

Structures of ERK2, JNK1 and p38 α in complex with docking peptides, show a similar mechanism of binding among proteins that interact with the same MAPK, but substantial differences between the three families (Garai *et al.*, 2012; Heo *et al.*, 2004; Laughlin *et al.*, 2012; Zhang *et al.*, 2011; Zhou *et al.*, 2006). Indeed, D-motifs are

selective and pathway specific despite relatively weak (micro molar) binding constants (Bardwell *et al.*, 2009; Bardwell, 2006; Smith *et al.*, 2000).

Therefore, D-motifs (also known as D-domains, DEJL motifs, or kinase interaction motifs) are probably the main specificity determinants that distinguish different MAP kinases (Bardwell, 2006; Garai *et al.*, 2012), which make the D-motif binding site a possible target for designing substrate based inhibitors specific for each MAPK. This idea is well established for protein kinases, but new for MAPKs (Akella *et al.*, 2008).

6.1.3 The docking interaction by MAPKK promotes MAPK activation

MAPKK binds MAPK by the N-terminal tail, which the D-motif belongs. This interaction is followed by the formation of transient enzyme-substrate complex at the active site (Tanoue *et al.*, 2001), where the double-phosphorylation occurs.

Therefore, the structural and biochemical studies, which investigate the docking interaction between MAPKK docking peptide and MAPKs, are not only useful for drug design studies, but also provide information about the preliminary structural rearrangements that occur during MAPK activation. So far, four structures have been solved showing MAPKK-MAPK interactions: ERK2 in complex with a MEK2 and MEK1 peptide and p38 α in complex with MKK3b (Chang *et al.*, 2002; Garai *et al.*, 2012; Zhou *et al.*, 2006). Recently, a structure of p38 α in complex with a MEK6 peptide was also solved (Garai *et al.*, 2012). In this last structure p38 α is mutated in the D-motif binding site, which makes the model proposed questionable.

The subject of this chapter is the structural characterisation of the docking interaction between p38 α and MEK6. In the p38 α construct used no mutated residues are in the D-motif binding site. In order to understand the results of this chapter, knowledge of how kinases, and in particular p38, becomes activated is required. Therefore, in the next sections a general overview about conserved kinase motifs within their role in the active kinase state, will be presented. Successively, the structures of the inactive and active conformations of p38 will be described with the goal of highlighting the molecular processes necessary for activating a kinase in general and p38 specifically. Eventually the p38 α D-motif binding site will be described and the known effects induced by the MAPKK docking interaction on p38 α discussed.

6.1.3.1 Structural features of activate state of protein kinases

Despite the diversity of input signals that kinases receive, the result is always the reorganisation of catalytic active site, which then becomes suitable for carrying out phosphorylation reaction. The catalytic site is located between the two lobes of protein kinases, termed the N- and C-terminal lobes. It is characterised by a set of highly conserved structural features, which define the so called “active state” of a protein kinase (Jura *et al.*, 2011).

Based on the structure of the active state of c-AMP-dependent protein kinase A (PKA; PDB code 1ATP), the relevant features of the internal active site architecture are the following (Kornev & Taylor, 2010):

- ATP is properly accommodated in the active site; this is due to the phosphorylation of the P-Lip (or activation loop) at specific sites, which induces a hinge motion of the two lobes and changes the relative orientation from a more open in the inactive state to a closer one in the active state.
- The P-Lip is involved in both nucleotide and substrate binding. The N-terminal region of the P-Lip contains the DFG motif, with the aspartate pointing towards the nucleotide phosphate group and playing a critical role in coordinating one of the two magnesium ions. The C-terminal region acts as platform for docking the substrate peptide.
- The secondmagnesium ion is coordinated by the catalytic base, which is the aspartate belonging to the HRD motif.
- ATP α P- β P are coordinated by a conserved lysine residue, which belongs to the strand β_3 in the N-lobe; the same lysine forms a salt bridge with a well conserved glutamate in helix α C.
- ATP β P- γ P interact with Glycine-rich loop, within the consensus Gly-X-Gly-X-X-Gly sequence. This region is also called the phosphate anchor, due to its role in recruiting the nucleotide (Hemmer *et al.*, 1997).

6.1.3.2 Structural conformational changes from the inactive to the active form of p38

The structure of the apo, un-phosphorylated form of p38 α is available both for the

murine (2.1 Å, PDB code 1P38) and human species (2.3 Å, PDB code 1WFC), which are 99% identical (Wang *et al.*, 1997; Wilson *et al.*, 1996a). For the description of the probable p38 α active conformation, the structure of bis-phosphorylated p38 γ (PDB code 1CM8; Bellon *et al.*, 1999) will be analysed.

Crystal structures have shown that p38 α has a typical protein kinase conformation: it is a small protein composed of the typical N- and C-terminal lobes. The N-terminal lobe is composed largely of a β -sheet, whereas the C-terminal lobe is largely α -helical (Figure 6.3). The catalytic site is at the junction between the two domains (Knighton *et al.*, 1991).

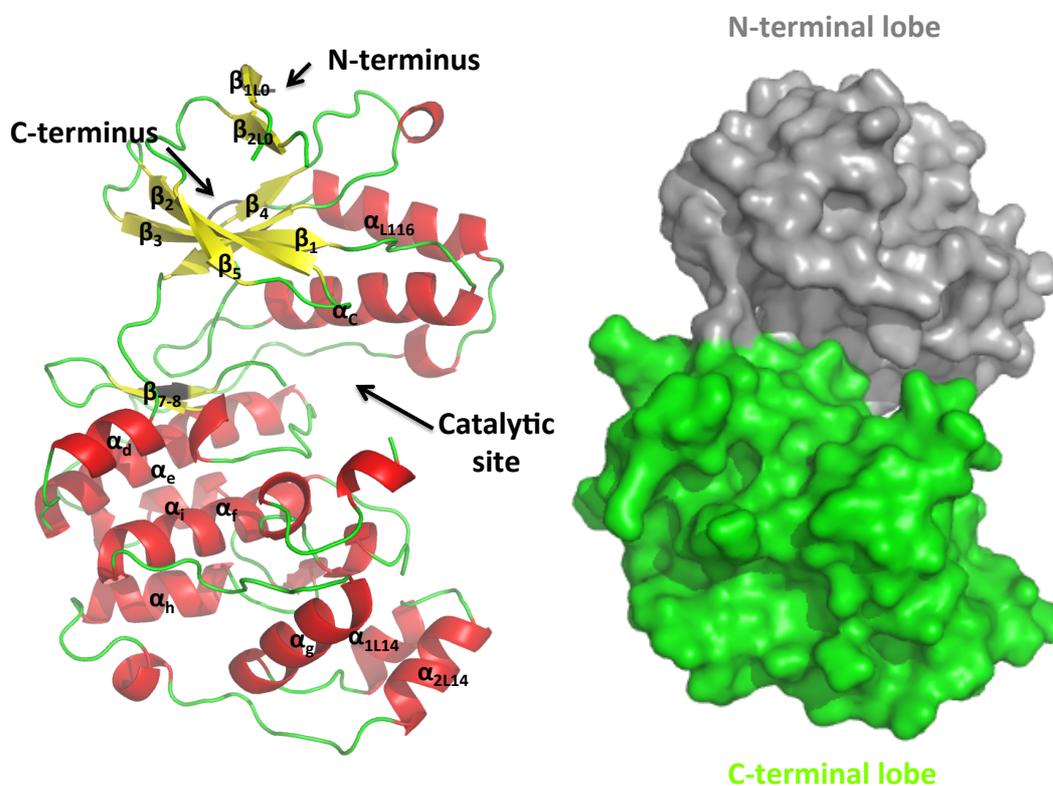


Figure 6.3 Secondary structure of p38 α . The secondary structure of human p38 α is represented as a ribbon representation (PDB code 1WFC; Wilson *et al.*, 1996a). The nomenclature is the common one for kinases. The surface representations show the two proteins lobes. The N-terminal lobe (in *grey*) contains both N and C-terminus of the protein, denoted by the *black* arrows.

Based on comparisons with PKA, the **N-terminal domain** (residues 1-109 and 320-

352) creates a binding pocket for the purine base of ATP (Walker *et al.*, 1982). Further, it contains the so called Glycine-rich loop (residues 31–36) (Figure 6.4). The Glycine-rich loop is poorly defined in both structures of un-phosphorylated p38 α . In the murine kinase it is partially disordered, while in the human structure it is highly mobile, with average B-factor of 61 Å² (total molecule B-factor is 24.5 Å²).

The **C-terminal domain** (residues 110-318) contains the HDR motif with the base catalyst, the magnesium binding sites and the phosphorylation lip (P-lip) within the DFG motif (residues 168-185) (Wilson *et al.*, 1996b) (Figure 6.4). The P-Lip contains the phosphorylation sites. MAPKKs activate p38 α by phosphorylating residues Tyr182 and Thr180 (Mansour *et al.*, 1994). These residues are well defined in the electron density in both structures. Thr180 is on the surface, while Tyr182 interacts through a water molecule with the side chain hydroxyl group of Thr221. The rest of the P-Lip is partially disordered in both structures: residues 171-178 are poorly defined in the electron density for the murine, while 171-174 are not visible in the human structure.

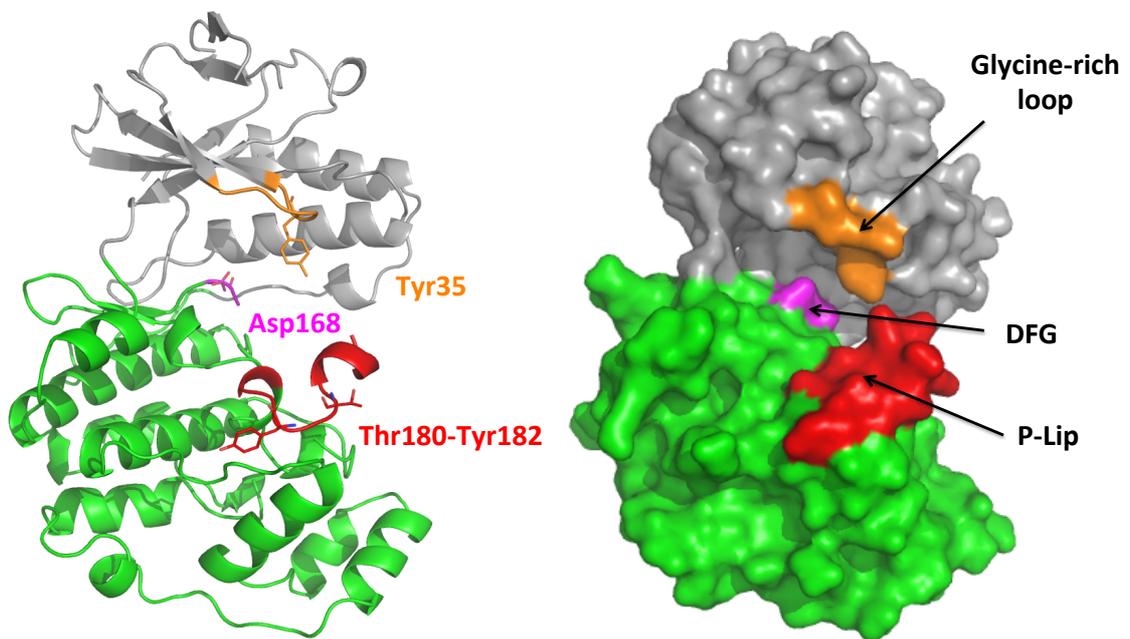


Figure 6.4 Functional motifs in p38 α . Ribbon and surface representation of functional motifs in the human p38 α structure (PDB code 1WFC; Wilson *et al.*, 1996a). The Glycine-rich loop is in *orange*, with Tyr35 in sticks. This tyrosine has occupancy zero in the structure. DFG motif is in *magenta* with Asp168, in sticks. The glycine of DGF motif is not visible in the electron density of human p38 α structure. The P-Lip is in *red*, with the phosphorylation sites in sticks. Part of it is disordered.

In the structure of the p38 active state, an ATP analogue is bound in the active site. As expected, Asp171_{p38 γ} (Asp168_{p38 α}) belonging to the DFG motif interacts with a magnesium ion that coordinates the nucleotide analogue. The base catalyst Asp153_{p38 γ} (Asp150_{p38 α}) belonging to the HDR motif points toward the active site. Lys56_{p38 γ} (Lys53_{p38 α}) from strand β_3 and Glu74_{p38 γ} (Glu71_{p38 α}) from helix αC form a salt bridge which is involved in ATP coordination (Figure 6.5; Bellon *et al.*, 1999).

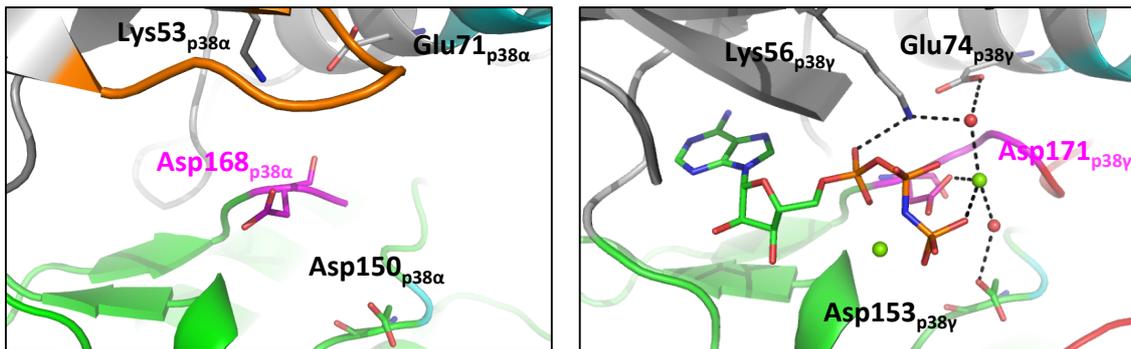


Figure 6.5 p38 active site. The panels magnify the active site of both inactive (left) and active (right) state of p38, where residues involved in catalysis are shown in sticks. Colours are similar to previous figures. H-bonds are represented with *black* dashed lines.

The comparison of the structures of the inactive p38 α and active p38 γ shows that phosphorylation causes a reconfiguration of the P-Lip and of the protein active site. Phosphorylation is supposed to cause the alignment of two fundamental intramolecular networks between the N-lobe and C-lobe, which are defined as the “regulatory spine” and the “catalytic spine” (Kornev *et al.*, 2006; Kornev *et al.*, 2008). Both networks contain hydrophobic residues, which become assembled upon kinase phosphorylation. The comparison of inactive and active state of CDK and Src kinases suggested that the DFG motif has a different orientation in the two states (Jura *et al.*, 2011). In particular, the aspartate and the phenylalanine side chains exchange positions due to a crankshaft-like motion of the peptide backbone. In the active state conformation, DFG is in the “in” conformation with the aspartate pointing toward the active site and the phenylalanine is assembled in the core of the regulatory spine. In the inactive conformation the DFG is in the “out” conformation, with the phenylalanine pointing toward the active site. The

inactive phenylalanine conformation prevents both the ATP binding and the assembling of the catalytic spine. As discussed, the DFG-out conformation is also induced by certain drug molecules.

Upon kinase phosphorylation, a change in the helix $\alpha 3$ orientation is also expected, which promotes the formation of the lysine-glutamate salt bridge and ATP binding. Structures of inactive and active kinase show that the helix $\alpha 3$ goes from an “outward” conformation associated to DGF-out state, to “inward” conformation associated to the DGF-in state.

In summary, phosphorylation at the P-Lip cause relevant structural rearrangements, which lead to the DGF-in conformation and consequently to proper orientation of helix $\alpha 3$. Indeed in the p38 active state, the DFG is in the “in” conformation and helix $\alpha 3$ is in inward state, which defines p38 being in the active state (Figure 6.6).

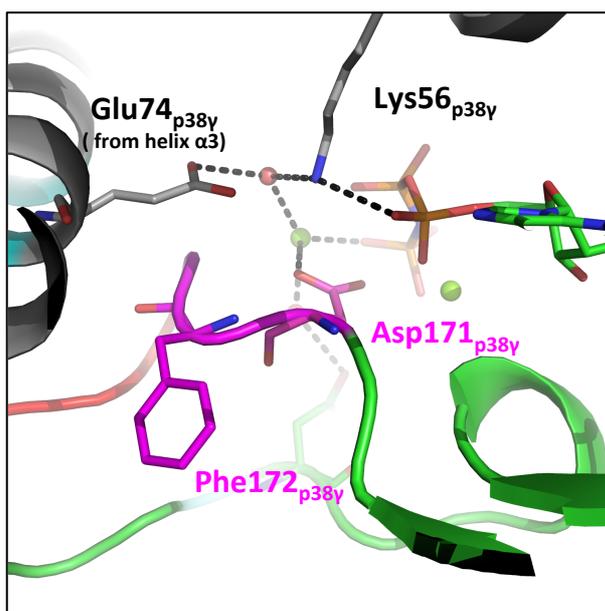


Figure 6.6 DFG-in conformation in the active state of p38 γ . The active site of the bis-phosphorylated p38 γ with bound AMP-PNP is shown (PDB code 1CM8; Bellon *et al.*, 1999), with a zoom on the DGF motif. Colours are similar to previous figures. H-bonds are represented with *black* dashed lines.

The comparison between the inactive p38 α and active p38 γ shows that this rearrangement is guided by the phosphorylation of the TxY motif and consequently, by

In particular, the formation of the P-Thr180 binding site (TPO) brings together three arginine residues, which belong to three different protein regions relevant in catalysis: Arg73_{p38 γ} (Arg70_{p38 α}) from helix α_3 , Arg152_{p38 γ} (Arg149_{p38 α}) from the catalytic motif HDR and Arg176_{p38 γ} (Arg174_{p38 α}) from the P-Lip (Figure 6.7).

Surprisingly, the Phenylalanine in both human and murine structure of inactive p38 is outward of the active site, suggesting for a DFG-in conformation. However, since the Phenylalanine is the last visible residue before the disordered region, its orientation is doubtful.

6.1.3.3 D-motif binding site in the inactive p38 α

As described above, the D-motif binding site on MAP kinases is formed by an acidic patch in the C-terminal extension of the kinase core known as the CD domain (Tanoue *et al.*, 2000) and a hydrophobic docking groove (Chang *et al.*, 2002; Gum & Young, 1999). The **CD domain** is an electrostatic surface depression (Zhou *et al.*, 2006). It comprises acidic residues, which are Asp313, Asp315 and Asp316 in p38 α . Moreover, it is part of a long loop (Loop16 or L₁₆), which connects the C-terminal lobe with the N-terminal lobe. Residues belonging to this loop can also participate in the interaction. The **hydrophobic docking groove** is located near helices α_d and α_e and in the β_7 - β_8 reverse turn (Akella *et al.*, 2010; Heo *et al.*, 2004; Lee *et al.*, 2006), which contains a sub-domain named ED, characterised by mutagenic analysis (Tanoue *et al.*, 2001) (Figure 6.8).

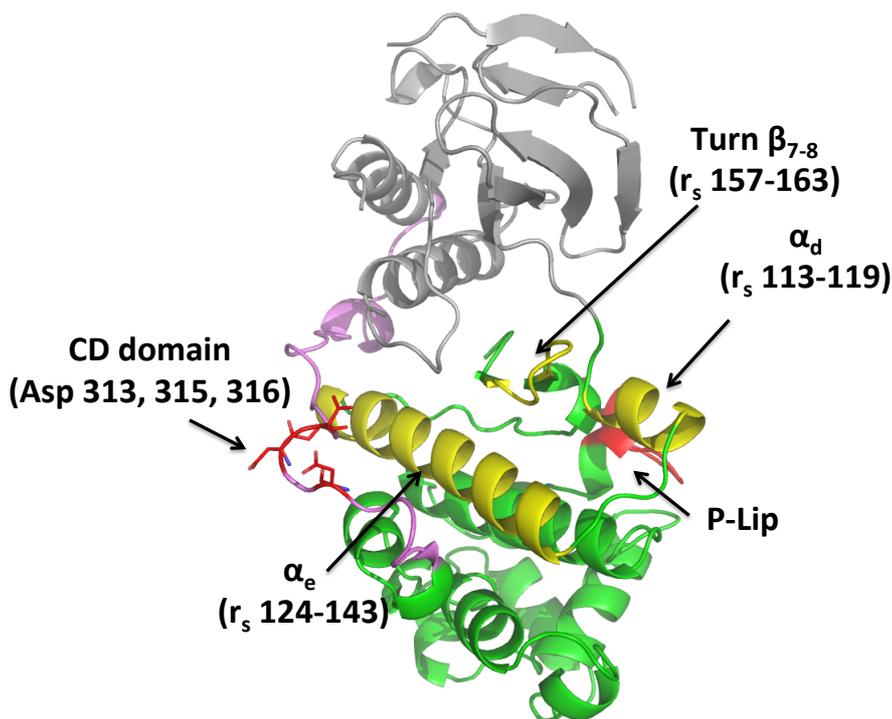


Figure 6.8 D-motif binding domain in p38 α . The secondary structure elements involved in the D-motif binding domain are shown for human p38 α (PDB code 1WFC; Wilson *et al.*, 1996b). The ED domain belongs to the β_7 - β_8 reverse turn. The CD domain is part of Loop16, which is shown in *pink*.

6.1.3.4 How does the MAPKK docking interaction promote p38 activation?

Structures of MAPKs in complex with peptides containing the MAPKK D-motif have shown that docking interactions cause long-range conformational changes, particularly in the activation loop. This is a preliminary event which prepares the MAPK for activation by phosphorylation.

Murine p38 α has been co-crystallised with docking peptides from MKK3b (pepMKK3b-p38 α), a MAPK kinase, and with MEF2A (pepMEF2A-p38 α), one of the transcription factors activated by p38 α (Chang *et al.*, 2002). Peptide sequences are reported in table 6.1. Since the structural changes induced by both peptides are equal, both structures were used for building a model of p38 α activation (Akella *et al.*, 2010). Both peptides bind in an elongated manner, without showing any binding to the CD domain. The selective interactions are in the hydrophobic pocket built around residue

Ile116. The peptide-protein interaction is stabilised by sidechain-backbone hydrogen bonds between sidechains of Gln120_{p38α}, Gln160_{p38α} and His126_{p38α} and the peptide backbone. Gln160_{p38α} had been previously characterised by mutagenic analysis and it is part of the so called ED domain. Despite these experiments, no interaction with aspartate 161 has been observed. For this reason, the ED domain will not be referred to in this chapter. A similar peptide-protein interaction is present in the recently published structure of p38α in complex with MEK6 peptide (pepMEK6-p38α_{C126S}, Garai *et al.*, 2012), where only the peptide residues interacting with the hydrophobic groove are visible in the electron density.

The binding of the peptide through the hydrophobic region induces disorder in the P-Lip, which should facilitate its binding to the MAPKK active site. Further conformational changes also occur in the ATP binding loop. Indeed, it has been proposed that the docking interaction promotes the nucleotide binding, through conformational changes at the inactive p38 active site (Figure 6.9).

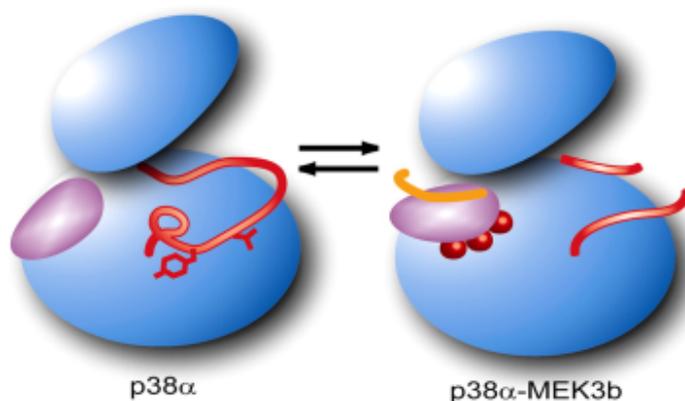


Figure 6.9 Conformational changes induced by docking interaction in inactive p38α. The cartoon shows conformational changes occurring upon docking interaction. The peptide is shown in *orange*. The interaction with the hydrophobic groove (helix α_d in *purple*) induces disorder in the P-Lip. New water molecules appear after peptide binding (*red* sphere). This last observation has been described in the structure of double-phosphorylated p38α in complex with MKK3b peptide, but not in pepMKK3b-p38α structure (modified from Akella *et al.*, 2010).

Similar hydrophobic and electrostatic interactions between MAPKK peptide and MAPK D-motif binding site have been found in the structures of ERK2 complexes with

docking peptides (Zhou *et al.*, 2006). However, the structural rearrangements induced are different in ERK from p38.

The model of ERK2 activation has been built through two peptide-MAPK structures. ERK2 has been co-crystallised with a peptide (pepMEK2) derived from the activating kinase MEK2 and also with a docking peptide derived from hematopoietic protein tyrosine phosphatase (pepHePTP), which dephosphorylates and retains the enzyme in the cytoplasm (Pulido *et al.*, 1998; Saxena & Mustelin, 2000; Zuniga *et al.*, 1999). Peptide sequences are shown in table 6.1. The two complexes are very similar (rmsd 0.4 Å) therefore, equivalent observations have been made on both (Zhou *et al.*, 2006). There are several features specific for ERK2-activation model, which appear to be absent in the p38 α -activation model. The D-motif peptide adopts a helical conformation at the N-terminus, which promotes the binding to the CD domain. This binding causes long-range conformational changes in Loop16, which finally induce conformational changes in the P-lip. The un-phosphorylated ERK2 structure shows that only one of the two phosphorylation sites is exposed to the solvent. Upon the docking interaction, both phosphorylation sites are exposed to the solvent (Figure 6.10).

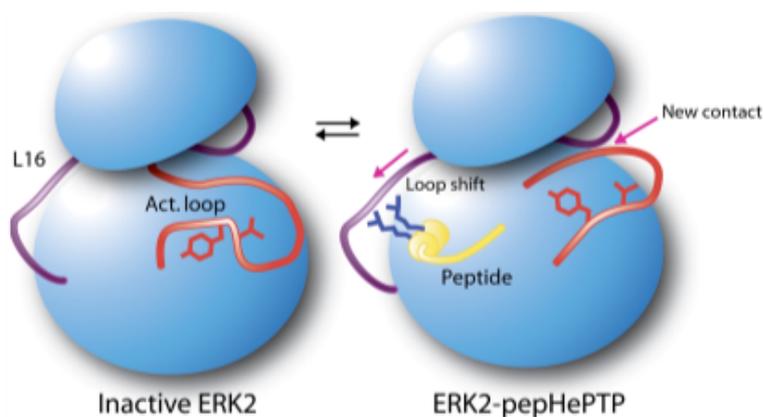


Figure 6.10 Conformational changes induced by docking interaction in inactive ERK2. The cartoon shows which conformational changes are required for exposing the ERK2 phosphorylation sites to the solvent. The two arginines of pepHePTP (RLQ**ERR**GSNVALMLDV) in *yellow*, interact with the CD domain in Loop16 (in *purple*). Following the formation of a new contact between the Loop16 and the P-Lip, both phosphorylation sites are exposed. N- and C- terminal lobes are represented by small and large *blue* eggs. P-Lip is in *red* (modified from Zhou *et al.*, 2006).

Thus, despite the fact that p38 and ERK are both MAPKs and a common phosphorylation- activation mechanism has been described, the structures solved so far reveal that MAPK activation can be achieved in different ways. D-motif peptides bind differently and different conformational changes are observed in the P-Lip. While in ERK2 the docking interaction is mainly through the CD domain, and it causes a change in P-Lip conformation, in p38 α the loop becomes completely disordered upon binding of MAPKK at the hydrophobic groove.

However, there are still some open questions about the p38 α -activation model proposed. No interactions with the CD domain have been detected in the crystallographic structures of p38 α in complex with a docking peptide. However, mutagenesis analysis has shown that peptide binding is decreased or absent when aspartate residues of the CD domain are mutated to alanines (Tanoue *et al.*, 2000). Modelling of pepMKK3b-p38 α and pepMEF2A-p38 α , has shown that basic residues from the peptide, not visible in the electron density, could interact with the CD domain (Chang *et al.*, 2002). Furthermore, ^{15}N - ^1H TROSY spectra on pepMKK3b-p38 α complex (Akella *et al.*, 2010) revealed that Asp313, 315, 316 from p38 α and surrounding residues move in response to peptide binding. Thus, the role of peptide binding to p38 α CD domain remains unclear.

The second open question relates to the disorder induced in the p38 α P-Lip, which could be a crystal artefact. The ^{15}N - ^1H TROSY data on pepMKK3b-p38 α has shown that long-range conformational changes occur when the peptide is bound. However, due to a problem of assignment, no observation was made for the P-lip, leaving open the question of the solution state of the loop in the docking complex. The assignment problems are likely due to the absence of defined structure in the P-Lip. Thus, a better and more stable reference model of the apo-p38 α is needed for investigating the veracity of the p38 α -activation model proposed.

6.2 Aim of this project

The aim of this project was to investigate the interaction between MEK6 and p38 α through the structure of p38 α in complex with a MEK6 D-motif peptide. As explained in the introduction, the structural characterisation of this complex will provide useful information for drug design and possibly, novel insight into p38 α activation.

For these crystallisation experiments a mutant of p38 α was used, in which the apo-structure showed a completely ordered P-Lip and, furthermore, a more stable Glycine-rich loop. The comparison between the apo-structure and the pepMEK6-p38 α structure shows the molecular basis of initial MEK6 binding and the conformational changes induced in p38 α . Despite the already published structure of pepMEK6-p38 α , in the structure solved here the C-terminus of the peptide is partially visible in the electron density, which allows the definition of possible secondary elements and possible interactions with the CD domain. Moreover, the published structure contains a mutation in the hydrophobic groove (C162S), whereas the structure determined here does not have this mutation.

Thus, the structures solved in this thesis (p38 α K53R and pepMEK6- p38 α _{K53R}) provide answers to the open questions presented in section 6.1.3.3. Therefore, a model of p38 α -activation will be shown at the end of this chapter.

6.3 Experimental procedures

6.3.1 Plasmids

Several constructs were used in the study. Some for crystallisation purposes, others for kinetic studies (see Chapter 7). Two plasmids pET-BS(+) and pET15b were employed for protein expression. The first encodes a non-cleavable N-terminal His₆-tag, the second is cleavable. Cloning was performed partially in Sheffield (in the NMR group of Prof. Jon Waltho) and partially in our laboratory.

For MEK6, three constructs were used:

- **pET-BS(+)-MEK6**: full length form of human MEK6; the construct had been retro-mutated from MEK6DD, in order to have an inactive form. This form requires phosphorylation by an upstream kinase to be activated. The mutagenesis experiment was performed in Sheffield;
- **pET-BS(+)-MEK6DD**: the full length, constitutively active form of human MEK6; this plasmid was donated by Vertex pharmaceuticals, Cambridge, MA, USA. The original construct was transferred into pET15b in order to have a cleavable His₆-tag (see below for details);
- **pET15b-ΔMEK6DD**: MEK6DD truncated at residue 46, a constitutively active form of human MEK6 ; the truncation eliminates the flexible region, supposed to be responsible for p38α recruitment;

For p38α, three constructs were used:

- **pET-BS(+)-p38α**: the full length, **inactive (non-phosphorylated)** form of human p38α.
- **p38αK53R**: full length, inactive and mutated form of human p38α. The protein is supposed to be unable to bind ATP. The mutagenesis experiment was performed in our laboratory (see below for details);
- **p38αT180A**: full length, inactive and mutated form of human p38α. The protein can be phosphorylated only at one site (Tyr182). The mutagenesis

experiments and protein purification were performed in Sheffield (Xiaoxia *et al.*, 2011).

6.3.1.1 MEK6DD cloning in pET15b

The MEK6DD DNA fragment was transferred from pET-BS(+) to pET15b in order to obtain an N-terminal cleavable His₆-tag. Both plasmids were cleaved overnight at 37°C by the restriction enzyme *NdeI* (New England Biolabs, Ipswich, MA, USA). The recognition site of the restriction enzyme is CA[^]TATG and it was chosen for its ability to cut at the first ATG codon, after the His₆-tag DNA sequence. The cleavage was verified by electrophoresis and both the MEK6DD fragment and pET15b plasmid were purified using the gel extraction kit from Qiagen (Hombrechtikon, Switzerland). Since only one restriction enzyme was used, the cutting plasmid was incubated with alkaline phosphatase (Sigma, St. Louis, MO, USA) in order to dephosphorylate the DNA plasmid to prevent recircularisation. Finally, the dephosphorylated plasmid was mixed with the MEK6DD fragment in different ratios (1:3, 1:6, 1:9, 1:12, named T-samples) and incubated for one hour at room temperature with T4 DNA Ligase (Rapid DNA ligation kit, Fermentas). As a control, equivalent samples were prepared without the MEK6DD fragment (named C-samples). All eight ligation products were transformed into *E. Cloni* DH5 α cells (Lucigen corporation, Middleton, WI, USA). Evaluation of the experiment was performed by comparing for each ratio the number of colonies between the T-sample plate and its respective C-sample plate. Eventually, a few colonies were selected and the correct incorporation of the gene verified by sequencing.

6.3.1.2 Mutagenesis of p38 α in p38 α K53R

Mutagenesis of p38 α was performed on His₆-tagged human p38 α in pET-BS(+). p38 α K53R was constructed by PCR-based mutagenesis, using *Pfu* polymerase (Fermentas, Vilnius, Lithuania) in magnesium supplemented buffer. The following primers were used: 5'tacgtgtggcagtgaggaagctctccagacc-3', with antisense 5'-ggctctggagagcttctctcaactgccacaagta-3'. The melting temperature of both oligonucleotides was 72°C. The following PCR program was run:

1. first denaturation: 3' 95°C
2. cycle (repeated 18 times): 1' 95°C, 1' 65°C 1', 8' 68°C
3. 10' 68°C

PCR product was treated with *DpnI* restriction enzyme (Stratagene, La Jolla, CA, USA), which eliminates the original templates by recognising methylated DNA. Treated PCR product was transformed into XL1-Blue Competent cells (Stratagene, La Jolla, CA, USA). Positive clones were picked and mutagenesis was verified by sequencing.

6.3.2 Protein expression and purification

All proteins were expressed and purified in a similar way. Due to the human origins, Rosetta2 cells were used for expression. Expression was carried out ON, at 20°C.

As all the constructs have an N-terminal His₆-tag, a nickel column was used for the first purification step. Thrombin was employed to cleave the His₆-tag for the cleavable constructs. Size exclusion chromatography was then performed. Details of purification are reported in Table 6.2.

ΔMEK6DD protein eluted as dimers at an elution volume of 78 ml from the size exclusion column of 130 ml bed volume, with a 260/280 ratio of 0.64 (Figure 6.11). The yield of the entire process was around 10 mg/L culture.

MEK6 and MEK6DD were purified using a smaller column (24 ml bed volume). The full length proteins were used in kinetic experiments (see Chapter 7), which required lower protein amounts than the crystallisation experiments. For this volume, elution was typically around 15 ml, with a 260/280 ratio in the range of to 0.65. The final yield was 6 mg/L culture for MEK6 and 10 mg/L culture for MEK6DD.

In all MEK6 purifications, the correspondence between the highly purified protein and the peak described was confirmed by SDS PAGE The protein was then concentrated and stored at -20°C.

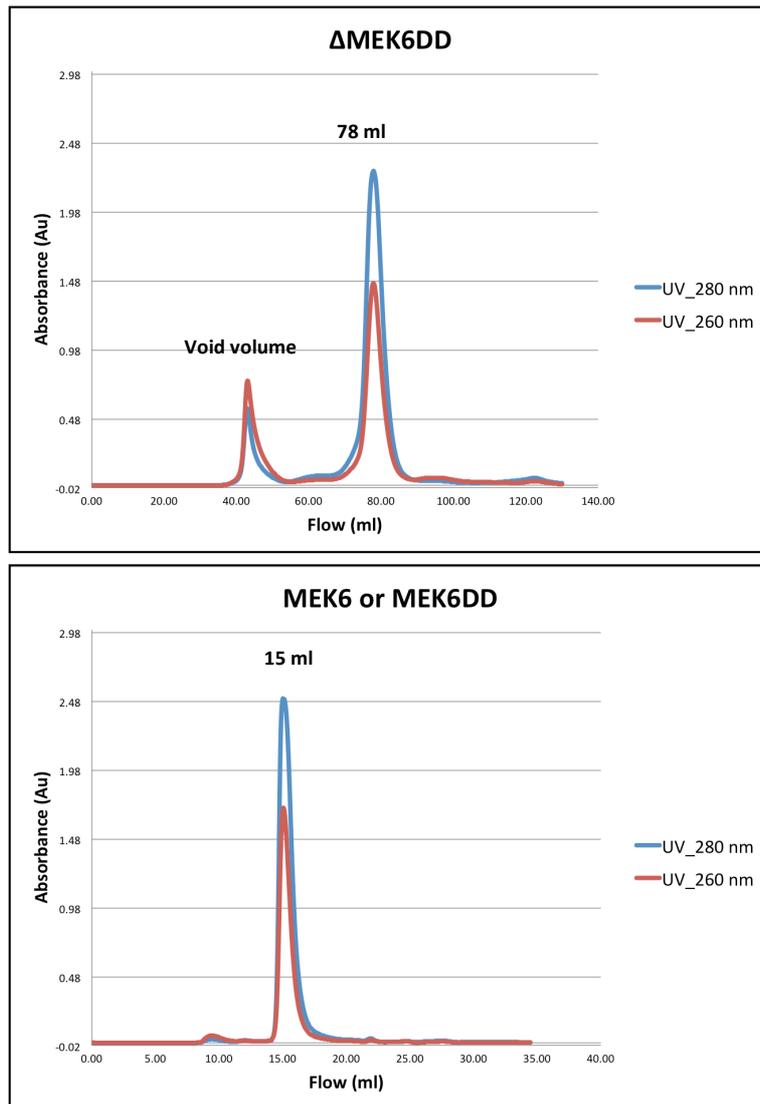


Figure 6.11 Size exclusion chromatography profile of Δ MEK6DD, MEK6 and MEK6DD. The absorbance at 280 nm is shown in *blue*, while the absorbance at 260 nm is shown in *red*. The Δ MEK6DD chromatogram was run with a HiLoad 16/60 Superdex 200 pg, while the MEK6 or MEK6DD chromatogram was run with Superdex 200 10/300 GL. Both chromatograms are the results of purification from 3L of culture.

For p38 α constructs, a further anion exchange chromatography step was necessary for separating the phosphorylated form from the un-phosphorylated form. Phosphorylation is probably due to endogenous bacterial kinases. A NaCl gradient was used to separate the two forms. The success of the experiment was verified on native PAGE gels, where the two species are clearly separated (Figure 6.12). The protocol for native PAGE gel will be described in the next chapter. Un-phosphorylated p38 α was then collected and

concentrated before storage at -20°C . The final yield for both p38 α and p38 α K53R was 20 mg/L culture.

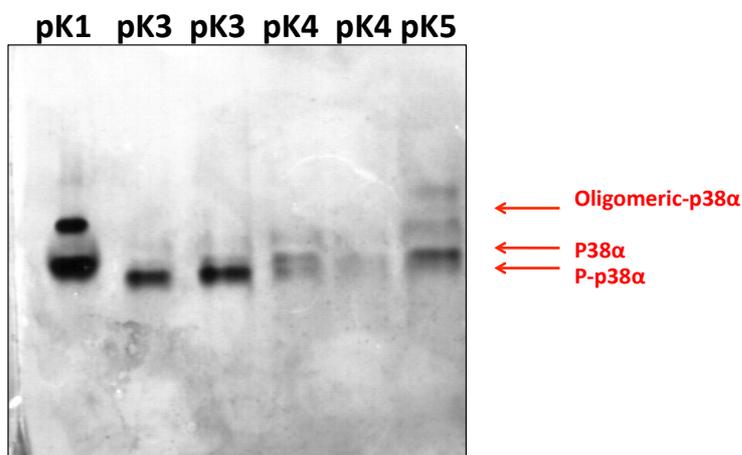
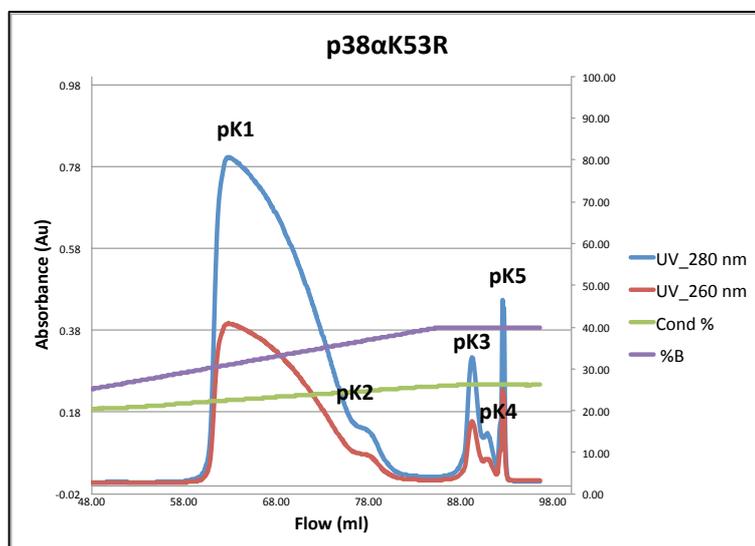


Figure 6.12 Anion exchange chromatography of p38 α K53R. The elution profile from 15 to 30% of buffer B is shown. 10 mg of protein were used for the experiment. In the native PAGE gel below, one or two fractions are run for each peak (pK). The distinction between un-phosphorylated p38 α and mono-phosphorylated p38 α will be elucidated in the next chapter. The absorbance at 280 nm is shown in *blue*, the absorbance at 260 nm is shown in *red*, %B is in *purple*. Conductance, expressed in %, is in *green*.

Table 6.2 Protein purification steps

Purification steps		Solutions for MEK6 constructs	Solutions for p38 α constructs
1. Cell Lysis	Bacteria were resuspended in the Lysis Buffer or LyB (10ml LyB for 750 ml of bacteria) with addition of a protease inhibitor cocktail tablet (EDTA free) and DNase (trace). The cells were lysed by sonication on ice for 3 * 50 s with 1 min gap, 60% power; then centrifuged for 20 min at 20,000 g, 4 °C.	Lysis buffer (LyB)= 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM MgCl ₂ , 2 mM DTT, 10 mM Imidazole (40 mM only for Δ MKK6)	Lysis buffer (LyB)= 50 mM Tris pH 7.4, 500 mM NaCl, 10 mM MgCl ₂ , 1 mM DTT, 10 mM Imidazole, 5% Glycerol
2. Affinity Chromatography	A Ni-NTA prepacked column was used (His-trap column-5 ml, GH), connected to an AKTA prime system. After equilibrating the column with the Binding Buffer, the binding step was performed at 0.5 ml/min until UV absorbance started to decrease. Then the column was washed extensively with BB until the UV absorbance was constant. The protein was eluted with a linear gradient at 1ml/min.	Binding Buffer (BB)= 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM MgCl ₂ , 0.5 mM DTT, 10 mM Imidazole (40 mM only for Δ MKK6) Elution Buffer (EB)= 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM MgCl ₂ , 0.5 mM DTT, 500 mM Imidazole	Binding Buffer (BB)= 50 mM Tris pH 7.4, 500 mM NaCl, 10 mM MgCl ₂ , 10 mM Imidazole, 5% Glycerol Elution Buffer (EB)= 50 mM Tris pH 7.4, 500 mM NaCl, 10 mM MgCl ₂ , 500 mM Imidazole, 5% Glycerol
3. His₆-tag cleavage	The protein buffer was exchange with a p10 column in the Thrombin cleavage buffer. After ON incubation with Thrombin at 4°C, gently shaking (100U/20 mg of protein), the affinity chromatography step was repeated in order to separate the cleaved protein from the un-cleave. A Hitrap benzamidine column was connected to His trap column to purify the cleaved protein from Thrombin. This step was performed at 1 ml/min.	Thrombin cleavage buffer (TCB)= 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM MgCl ₂ , 0.5 mM DTT, 2mM CaCl ₂	Thrombin cleavage buffer (TCB)= 50 mM Tris pH 7.4, 500 mM NaCl, 10 mM MgCl ₂ , 2mM CaCl ₂ , 5% Glycerol
4. Size exclusion chromatography (SEC)	A HiLoad 16/60 Superdex 200 pg or a Superdex 200 10/300 GL was run at a flow rate of 1 mL/min with SEC buffer. MEK6 constructs require this step at 4°C.	SEC buffer= 50 mM HEPES pH 7.5, 250 mM NaCl, 10 mM MgCl ₂ , 5 mM DTT, 5% Glycerol	SEC buffer = 50 mM Tris pH 7.4, 50 mM NaCl, 10 mM MgCl ₂ , 1mM DTT, 5% Glycerol
5. Mono Q	A MonoQ 5/50 GL column was run for p38 α constructs. Loading was performed at 0.3 ml/min in 50 mM salt. A slow gradient from 150 mM to 500 mM salt was selected after several trials (1ml/min).		Buffer A = 50 mM Tris pH 7.4, 50 mM NaCl, 10 mM MgCl ₂ , 1mM DTT, 5% Glycerol Buffer B = 50 mM Tris pH 7.4, 1 M NaCl, 10 mM MgCl ₂ , 1mM DTT, 5% Glycerol

6.3.3 Crystallisation experiments

After a MonoQ purification step, un-phosphorylated p38 α K53R was concentrated to 10 mg/ml and stored at -20°C. For crystallisation trials, p38 α K53R was prepared by adding NH₄F (10 mM) and AlCl₃ (2mM) as additives. For p38 α K53R in complex with the MEK6 peptide (pepMEK6-p38 α _{K53R}), p38 α K53R was incubated ON at 4°C with a 3-fold molar excess of peptide. The peptide SKGKKRNPGLKIPKA was purchased from Eurogentec (Angers, France).

p38 α K53R and pepMEK6-p38 α _{K53R} crystallised in different conditions. Crystallisation conditions were established by testing several commercial screens at the EMBL High Throughput Crystallisation Laboratory (Grenoble, France). Crystals of p38 α K53R were obtained at 20°C by the sitting drop method from solutions containing 10 mg/ml p38 α K53R in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 10mM MgCl₂, 1mM DTT, 5% (v/v) glycerol, 10 mM NH₄F and 2 mM AlCl₃ equilibrated against 3.5 M Na Formate pH 7.0 (Figure 6.13). Crystals were transferred to a cryoprotection buffer prepared by equilibrating protein buffer solution against reservoir solution supplemented with 20% (v/v) glycerol in a sitting drop plate and harvested using a micromesh loop (MiteGen, Ithica, NY, USA), plunged into liquid nitrogen and stored at 70K.

Crystals of pepMEK6-p38 α _{K53R} were obtained as above but equilibrated against a buffer containing 25% (w/v) PEG 3350 and 0.1 mM Bis/Tris pH6.5 (Figure 6.13). Crystals were transferred to a cryoprotection buffer, prepared as described above (reservoir supplemented with 20% (v/v) PEG400) and harvested using a micromesh loop (MiteGen, Ithica, NY, USA), plunged into liquid nitrogen and stored at 70 K.

6.3.4 Data collection, structure solution and refinement

Diffraction data were collected at beamlines ID23-1 and ID23-2 at the ESRF (Grenoble, France) on an ADSC Q315 CCD detector (ID23-1) or a MAR225 CCD detector (ID23-2) to between 2.2 and 2.8 Å resolution. Crystals of p38 α K53R contained multiple leaves of crystals and the combination of a micro-focussed or micro-beam in combination with automated mesh scans was essential. Crystals of pepMEK6-p38 α _{K53R} formed long needles (100 x 20 x 20 μ m³) in clusters that were difficult to visualise (Figure 6.13). Crystals were visualised and the best diffraction volumes were defined by diffraction

cartography as implemented in the beamline GUI MxCuBE (Bowler *et al.*, 2010; Gabadinho *et al.*, 2010).

Data were processed with XDS (Kabsch, 2010). The structures were solved by molecular replacement with MolRep (Vagin & Teplyakov, 2010). For p38 α K53R, the PDB accession code 1WFC (Kamps & Sefton, 1986) was used as search model. For pepMEK6-p38 α _{K53R}, p38 α K53R structure was used as search model. Refinement was carried out alternately using Phenix.refine (Afonine *et al.*, 2012) and by manual rebuilding with the program COOT (Emsley & Cowtan, 2004), as described in Chapter 2. Single peptide residue occupancy was refined using Phenix.refine (Afonine *et al.*, 2012).

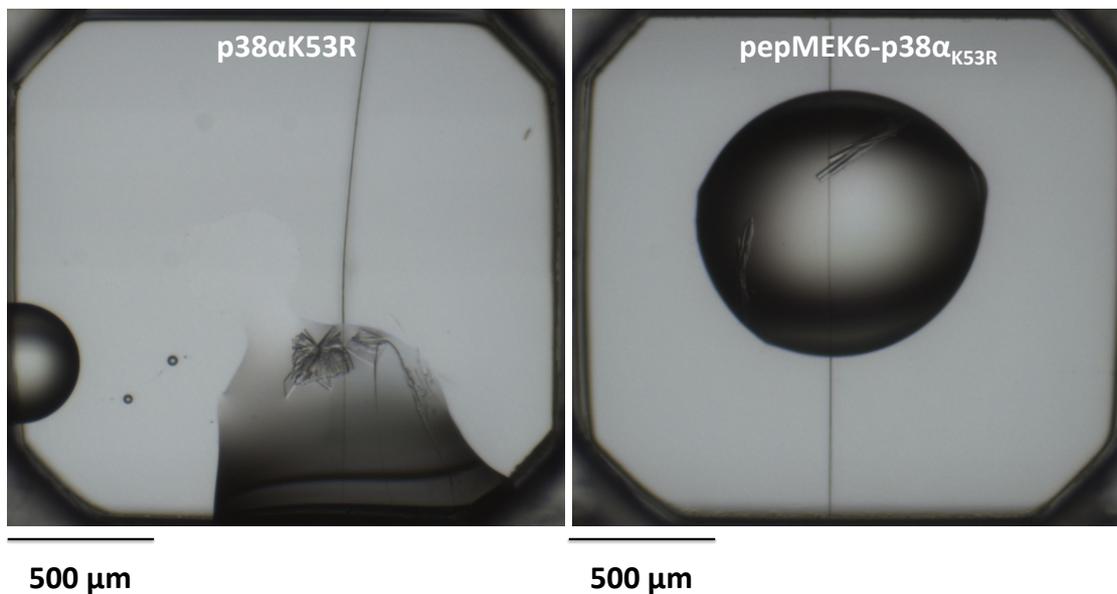


Figure 6.13 Crystals of p38 α K53R and pepMEK6-p38 α _{K53R}.

6.4 Results and Discussion

6.4.1 The structure of p38 α K53R

6.4.1.1 p38 α K53R is an inactive mutant

Structures of both murine and human recombinant p38 α have been solved to high resolution. Both structures show disorder in a few regions, such as the Glycine-rich loop and the phosphorylation lip region. The structure of p38 α K53R was solved at 2.8 Å resolution. Surprisingly, the structure of the mutant is much more ordered in the P-Lip and in the Glycine-rich loop than WT p38 α . This mutant was used for further crystallographic (in this Chapter) and kinetic experiments (see Chapter 7).

Lys53 in subdomain II of p38 α is fully conserved among the protein kinase family members (Hanks *et al.*, 1988). This lysine has been described to be essential in the phosphoryl transfer reaction (Gibbs & Zoller, 1991; Kamps & Sefton, 1986). In the structure of double-phosphorylated p38 γ , Lys56, which corresponds to Lys53 in p38 α , is coordinating the α P of AMP-PNP (Bellon *et al.*, 1999). Mutation of this conserved lysine in several protein kinases has resulted in loss of kinase activity, or kinases with low residual activity (Gibbs & Zoller, 1991; Kamps & Sefton, 1986; Robbins *et al.*, 1993; Snyder *et al.*, 1985). Mutation of this lysine in ERK2 (K52A and K52R), drastically lowered the kinase activity by reducing k_{cat} , conserving 5% of activity of the wild type protein. Structural analysis of ERK2K52R with ATP has revealed that the substitution of a lysine with an arginine altered the way that the ATP β -phosphate group is bound due to an interaction with the guanidinium group. Furthermore, the coordination sphere of Mg^{2+} is altered. Thus, the mutation creates a non-productive binding mode for ATP. This suggests a relevant role of Lys52 in orienting ATP binding in ERK2 (PDB code 1GOL; Robinson *et al.*, 1996). Interestingly, the ERK2K52R structure revealed an ordered Glycine-rich loop with regard the WT, due to a hydrogen bond between the Tyr35 backbone and the guanidinium group of Arg52. Similar mutagenesis experiments have been performed on the human form of p38 α . Biochemical data show an elimination of kinase activity (Kumar *et al.*, 1995).

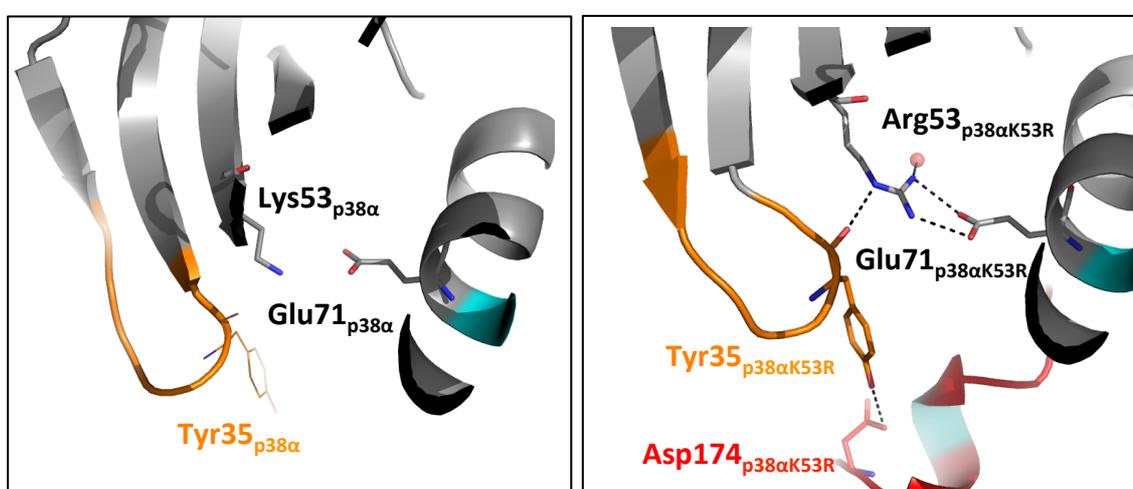
p38 α K53R crystallised in the orthorhombic space group $P2_12_12_1$ with one molecule in the asymmetric unit (Table 6.3). General features, such as lobe orientation, are similar to the human and murine p38 α structures.

Table 6.3 Data collection and refinement statistics

p38αK53R	
Space group	$P2_12_12_1$
Wavelength (Å)	1.07
Unit cell dimension (Å) a,b,c	45.42, 87.02, 122.92
Resolution range (Å)¹	20 (2.95-2.8)
Number of unique reflections	10748 (1584)
Multiplicity	3.9 (3.9)
Completeness (%)	87.9 (90.8)
R_{merge}	0.174 (0.582)
<I/σ(I)>	8.3 (3.1)
Wilson B factor (Å²)	28.05
Water molecules	64
R_{factor} (%)	16.73
Free R_{factor} (%)	22.6
RMS deviations:	
Bonds (Å)	0.008
Angles (°)	1.225

¹Values for higher resolution shell are in brackets

The structure provides evidence for the inability of the protein to bind nucleotide. Arg53 is extended in the active site cleft in a similar way to Lys53 in the WT structure. However, the replacement of the lysine with an arginine causes the formation of a new hydrogen bond network among the Arg53 side chain, the Glu71 side chain and the Tyr35 backbone (Figure 6.14). This interaction, absent in the wild type structure, is close to the hypothetical nucleotide phosphate binding site and it is comprised of residues involved in the catalysis and the Glycine-rich loop. Unexpectedly, the DFG motif is in “in” conformation.



6.14 Effects of mutation K53R on the Glycine-rich loop. Details of p38 α phosphate anchor or Glycine-rich loop are shown in human p38 α structure (PDB code 1WFC; Wilson *et al.*, 1996a) and in p38 α K53R. Tyr35 has occupancy zero in the WT structure (represented in lines). New hydrogen bonds introduced by the mutation are shown in *black* dashed lines. The Glycine-rich loop is in *orange*, the P-Lip in *red*.

The mutation has two likely effects. Firstly, it stabilises the Glycine-rich loop or phosphate anchor in one conformation, which decreases the flexibility of this region, required for recruiting the nucleotide. Tyr35_{p38 α K53R}, which has low occupancy in the human and murine structure of p38 α structure, is now interacting with Arg53_{p38 α K53R} through the backbone as described in the ERK2K52R structure. The conformation assumed by Tyr35_{p38 α K53R} and the orientation assumed by Arg53_{p38 α K53R} are identical to the one in the ERK2 structure, even if the two structures are not similar in the hydrogen bond network caused by the mutation. Secondly, the arginine impedes ATP binding due

to steric hindrance. This is shown in figure 6.15, where an arginine has been modelled instead of Lys56 in the structure of active p38 γ . The dot surface representation of the VdW sphere shows clearly that a conformational change is required for accommodating both the charge and size of the guanidinium group. It is probable that the larger amino acid sidechain cannot be accommodated with productive ATP binding.

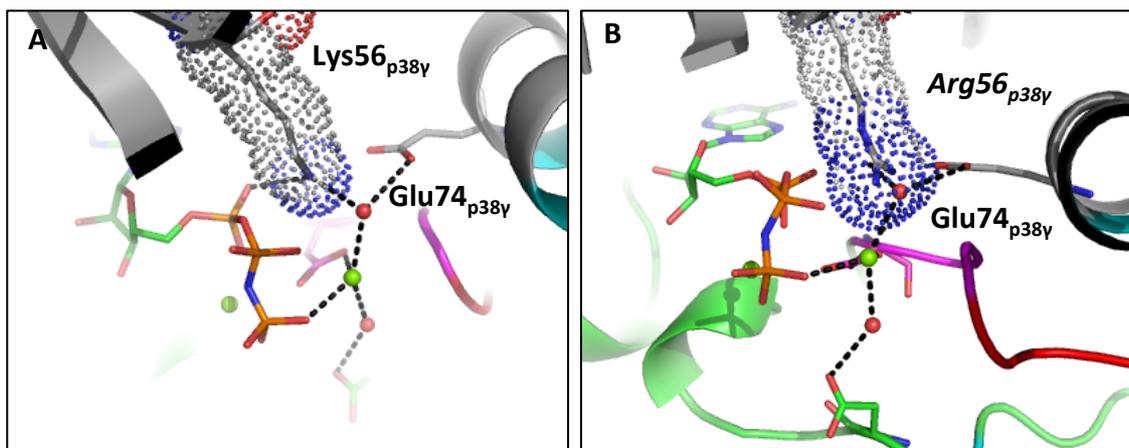


Figure 6.15 Possible effects of mutation K53R in the active site of active p38. The active site of bis-phosphorylated p38 γ is shown (PDB code 1CM8; Bellon *et al.*, 1999). VdW sphere of Lys56 (A) and Arg56 (B) are shown in dots. Glu74 and one of the water molecules coordinating magnesium are too close to the guanidinium group of Arg56. The mutation has been performed with *Pymol*.

6.4.1.2 The functional motifs of p38 α are well ordered in p38 α K53R structure

The mutation K53R leads to a better ordered structure, in comparison with the human and murine structure of p38 α . As in ERK2K52R, all the Glycine-rich region is well ordered in the electron density with acceptable B-factors for the resolution achieved (43.7 Å² vs 25.6 Å² of the entire structure). Importantly, the P-Lip (Gly170-Thr185) is almost completely ordered. The structure of un-phosphorylated p38 α P-lip has never been observed before. Clear density is visible for all the residues. The hydroxyl group of the Tyr35 is oriented toward the phosphorylation lip and it appears to coordinate the Asp176 side chain. Thr180 and Tyr182 maintain the same conformation and interactions described in p38 α structure. Furthermore, more residues are visible in the N- and C-terminal regions. An additional interaction not present in WT p38 α was

observed in the N-terminal lobe, where a hydrogen bond between the side chain of Gln3 and the Ile346 backbone, stabilises the interaction between the protein termini. In this thesis the name N-terminal hinge will refer to this interaction (Figures 6.16-6.18).

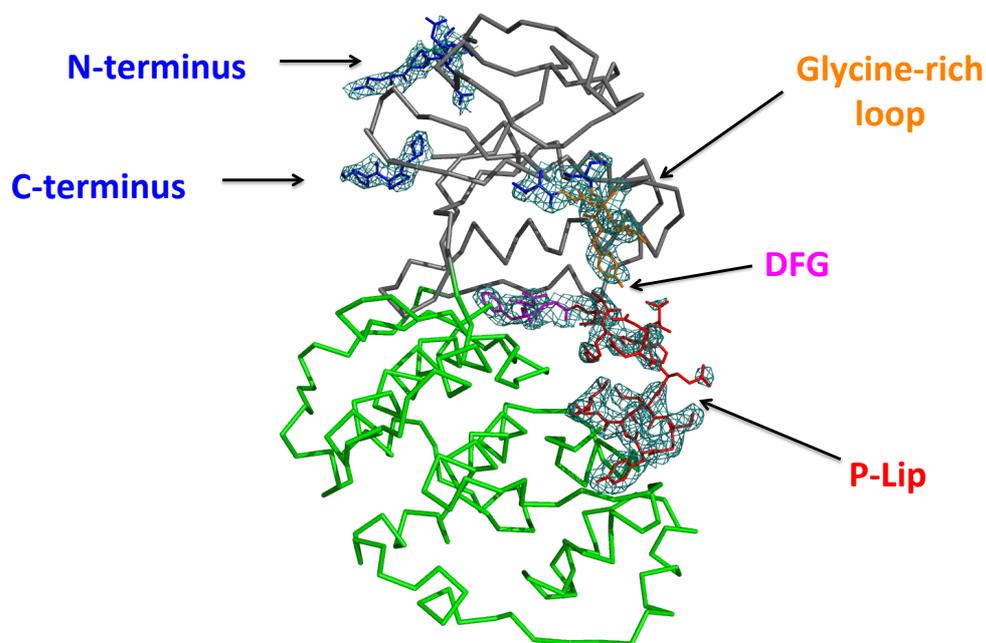


Figure 6.16 Electron density map of functional motifs in p38 α K53R. Electron density countered at 1σ is shown for N- and C-terminus (in *blue*), Glycine-rich loop (in *orange*), DFG motif (in *magenta*) and P-Lip (in *red*). The motifs are shown as sticks.

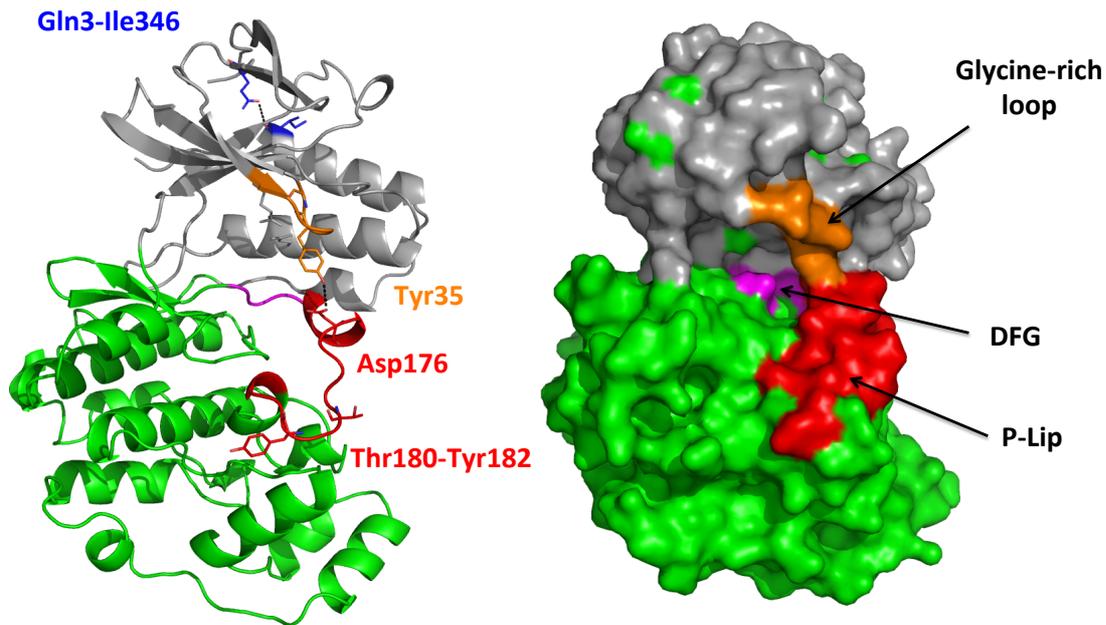


Figure 6.17 Functional motifs in the structure of p38 α K53R. The structure of p38 α K53R is shown. Functional motifs are shown in different colours, as described in figure 6.4. Gln3, Ile346, Tyr35, Thr180 and Tyr182 are shown as sticks. Hydrogen bonds are represented by *black* dashed lines.

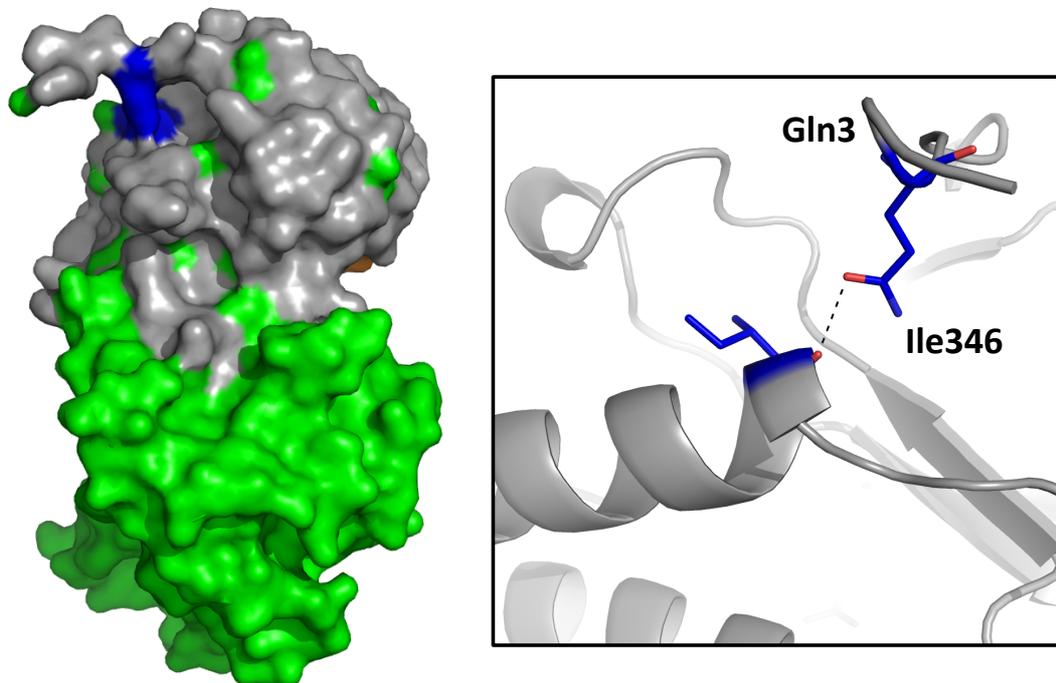


Figure 6.18 The N-terminal hinge. The N-terminal hinge is represented in *blue* on the surface representation. The panel is a magnification of the H-bond (3.2Å) connecting Gln3 to Ile346.

6.4.1.3 Is auto-inhibition in the p38 α K53R a crystal artefact?

Further stabilisation of the structure is achieved by crystal contacts. Interestingly, residues from the N-terminus occupy the docking groove of a symmetry related molecule. Similar interactions have been described for several Ser/Thr kinases, such as c-AMP-dependent protein kinase (cAPK), casein kinase-1, and cyclin-dependent kinase 2 (CDK2) (De Bondt *et al.*, 1993; Knighton *et al.*, 1991; Xu *et al.*, 1999). A C-terminal groove interaction is also present in WT human p38 α . As further structural details are visible in the p38K53R structure, such as a complete N-terminus, the interaction can be further described. The structure shows that the interaction is not only hydrophobic, as described previously, but Gln120, which belongs to the D-motif binding groove, is hydrogen bonded to the N-terminus. However, this interaction goes to a residue remaining from the cleavage of the His₆-tag (Figure 6.19). Thus, the tail-groove interaction observed is partially an artefact of the construct used.

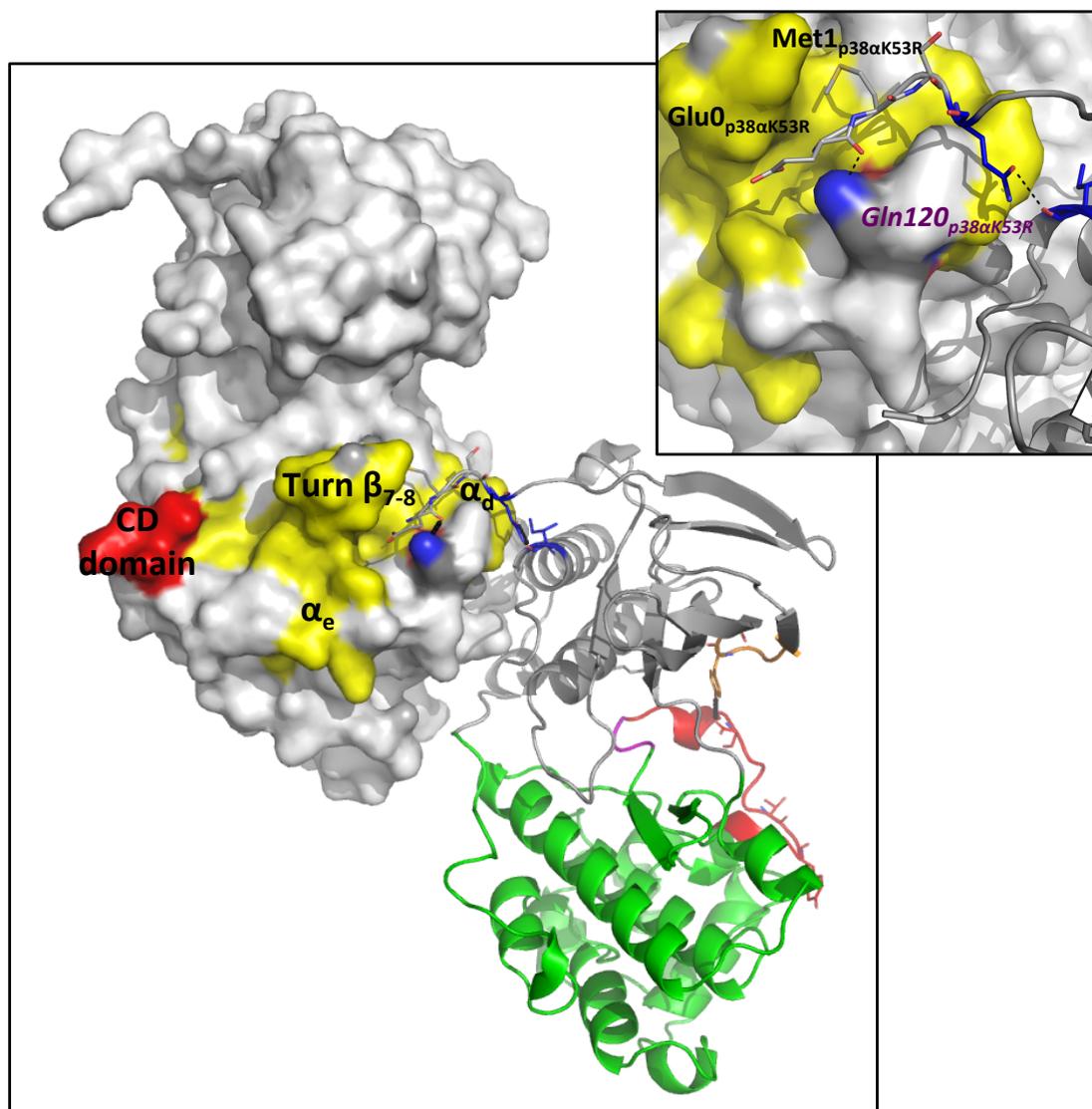


Figure 6.21 Docking interactions in symmetry related molecules of p38K53R. The docking interaction between two symmetry related molecules is shown in the D-motif binding site. This region is coloured in *red* and *yellow* as described in figure 6.6. The panel is a magnification of the interaction. Residues belong to different molecules are shown with labels in different colours (*purple* instead of *black*). Hydrogen bonds are in *black* dashed lines.

6.4.2 The structure of p38αK53R bound to the docking peptide of MEK6

6.4.2.1 The MEK6 peptide binds in the D-motif binding site of p38αK53R

p38αK53R has been successfully crystallised in complex with a 16 residue synthetic peptide containing the MEK6 docking sequence. The structure was solved at 2.4 Å resolution, a similar resolution to other p38-peptide complexes reported in the literature

(1.9-2.3 Å). The complex crystallised in space group $P3_121$, with one molecule per asymmetric unit (Table 6.4).

Table 6.4 Data collection and refinement statistics

pepMek6-p38αK53R	
Space group	$P3_121$
Wavelength (Å)	0.87
Unit cell dimensions	
a,b,c (Å)	82.38, 82.38, 123.33
α, β, γ (°)	90.00, 90.00, 120.00
Resolution range (Å)¹	20 (2.52-2.4)
Number of unique reflections	17587 (1638)
Multiplicity	3.9 (3.2)
Completeness (%)	90.3 (59)
R_{merge}	0.093 (0.524)
<I/σ(I)>	13.3 (2.1)
Wilson B factor (Å²)	29.18
Water molecules	253
R_{factor} (%)	16.18
Free R_{factor} (%)	22.7
RMS deviations:	
Bonds (Å)	0.008
Angles (°)	1.25

¹Values for higher resolution shell are in brackets

The choice of the sequence used was based on the alignment proposed in table 6.1 (Chang *et al.*, 2002). The peptide contains the D-motif region, preceded by several lysine residues: SKGKKRNPGLKIPKEA. Previous experiments defined that the binding constant for a similar peptide (SKGKKRNPGLKIPK) is 7 μ M (Garai *et al.*, 2012). *S* is residue number four in the MEK6 sequence, while it is residue -9 based on the D-motif-peptide nomenclature (Figure 6.20). In this sequence the D-motif-basic residue is an arginine, while a lysine is in between the two hydrophobic residues.

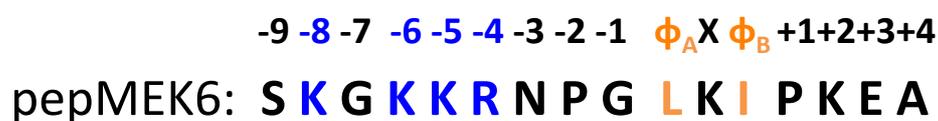
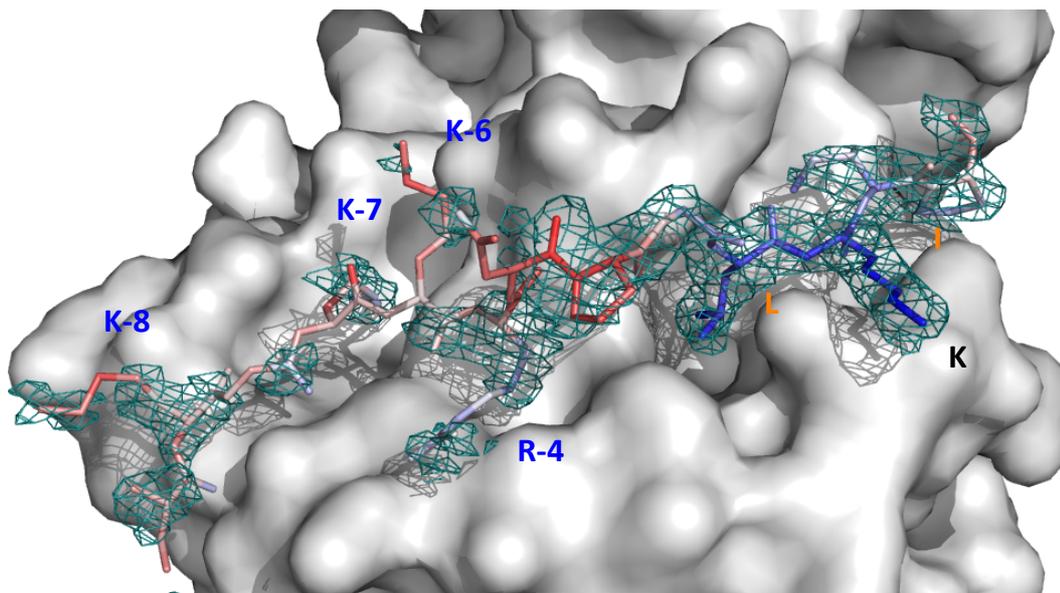


Figure 6.20 Peptide sequence nomenclature. Residues number zero are ϕ_A and ϕ_B , (in *orange*). Basic residues are in *blue*. The nomenclature reported is the one used in Chang *et al.*, 2002.

Residues PGLKIP (inter-Pro) are well defined in the electron density, (Figure 6.21; Table 6.5). The N-terminus of the peptide is partially visible in the electron density, in particular around the three lysine residues. The C-terminal residues K, E and A are not visible in the electron density.

It was possible to build the main chain from Ser-9 to Pro+1 and to determine whether possible secondary structure elements were present. In total, 12 of the 16 peptide residues were built. Residues KEA (+2 - +4) and residue -6 are not visible. Side chains have been successfully built for the intra-Pro residues, while in the N-terminus only side chains of S-9, K-8 and the guanidinium group of R-4 are visible in the electron density contoured at 1 σ (Figure 6.21).



	B-factors (\AA^2)		B-factors (\AA^2)	
S -9	49.3	P -2	59.6	
K -8	51.6	G -1	40.9	
G -7	45.2	L ϕ_A	28.5	
K -6	48.3	K	26	
K -5	52.2	I ϕ_B	41	
R -4	46.3	P +1	45.5	
N -3	57.4			

Figure 6.21-Table 6.5 B-factor analysis of MEK6 peptide. The MEK6 peptide is coloured based on the atomic B-factors (*scaled blue to red via white*, minimum = 20 \AA^2 , maximum = 70 \AA^2). B-factors correlate with the electron density distribution, shown in *green* mesh at 0.8 σ . Low B-factors and well-defined electron density correspond to the hydrophobic region, while high B-factors and less defined electron density, belong to the basic-residues zone.

Indeed the peptide was found to bind in an extended conformation in the hydrophobic groove between α_d and α_e and the reverse turn between β_7 and β_8 in the C-terminal lobe of the kinase (Figure 6.22), as described in pepMKK3b-p38 α , pepMEF2-p38 α and pepMEK6-p38 α_{C162S} (PDB codes 1LEZ, 1LEW, 2Y8O; Chang *et al.*, 2002; Garai *et al.*, 2012).

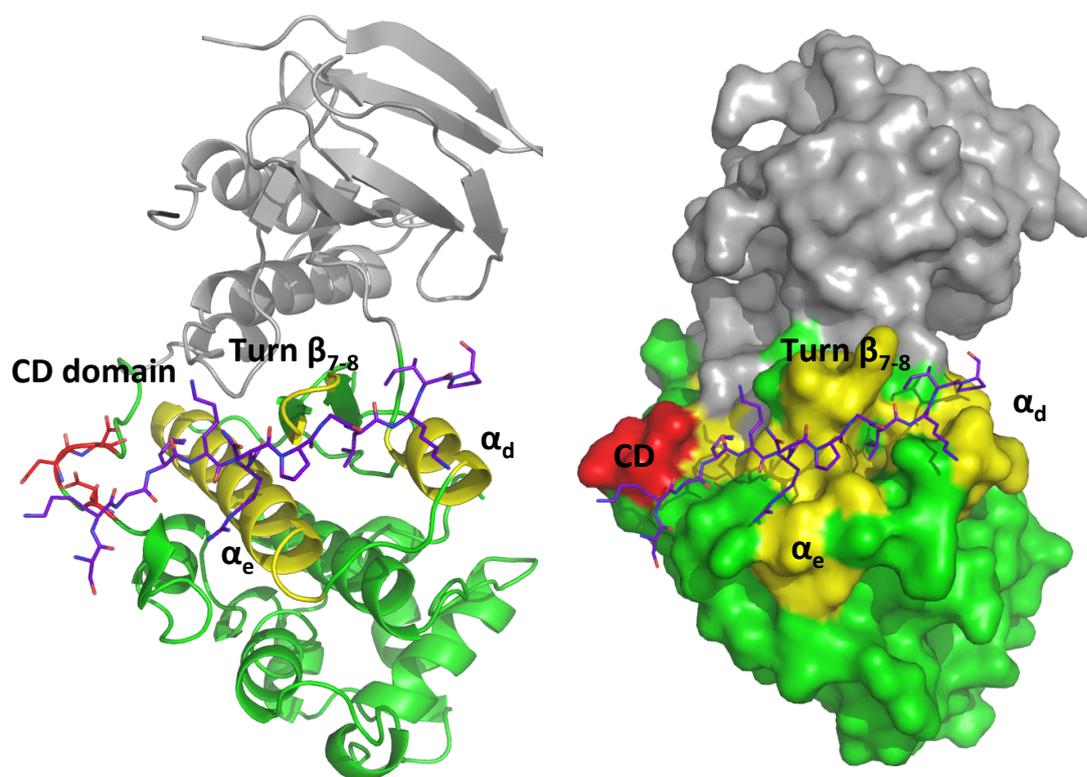


Figure 6.22 MEK6 peptide binds to the p38 α D-motif binding groove. The ribbon and surface representation of pepMEK6-p38 α_{K53R} structure show that the peptide binds in the groove formed by the CD domain, α -helices α_e and α_d , and β_7 - β_8 turn. The MEK6 peptide is in *purple*. The D-motif binding groove is coloured as in figure 6.8.

Among the structures solved of proteins that use the N-terminal tail for recruiting p38 α , this is the longest ordered peptide characterised so far. Analysis of the B-factors shows that the peptide N-terminus is highly flexible, which explains why it is difficult to see in the crystal structures. It is likely that the well-ordered peptide is due to the K53R mutation which locks the protein into a single and much more stable conformation. Surprisingly, the peptide residues modelled in pepMEK6-p38 α_{K53R} have a different position in the recently solved pepMEK6-p38 α_{C162S} structure (PDB code 2Y8O; Garai *et al.*, 2012). In this structure, p38 α carries the mutation C162S. p38 α_{C162S} has been described to increase protein stability, to facilitate the crystallisation process and crystals of p38 α_{C162S} show better diffraction than crystals of WT (Patel *et al.*, 2004). The different peptide conformation is probably due to the mutation C162S, which stabilises

the position of P-2 in the position of R-4 of the MEK6 peptide used in this thesis. As shown in the figure 6.23, S162 is hydrogen bonded to the peptide backbone.

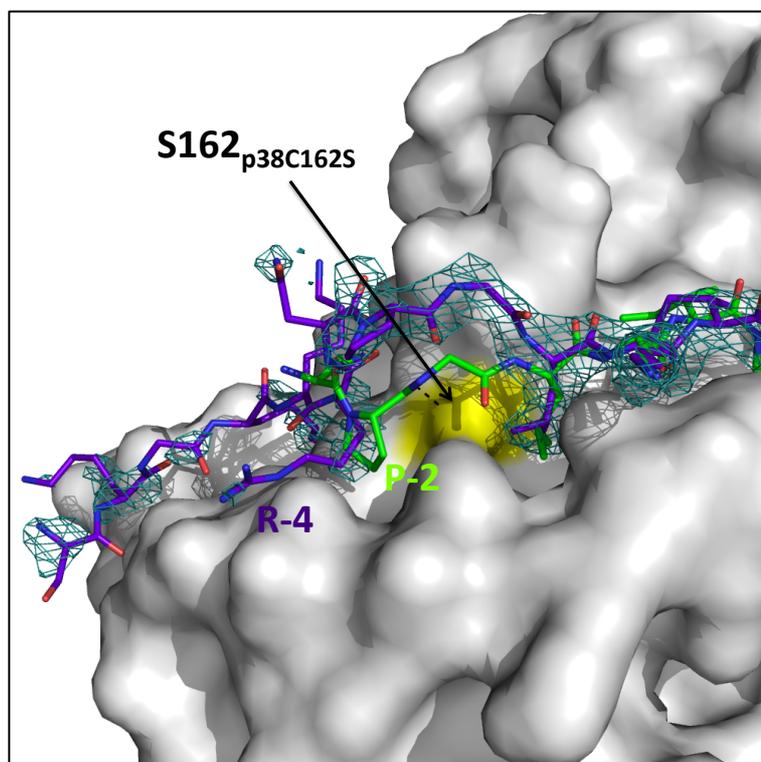


Figure 6.23 Differences in MEK6 peptide conformations. The electron density map of MEK6 peptide from pepMEK6-p38 α _{K53R} is shown overlaid into the structure of pepMEK6-p38 α _{C162S} (PDB code 2Y8O; Garai *et al.*, 2012). Mutated residue is in *yellow*. The H-bond between the Ser162_{p38 α C162S} side chain and the peptide backbone (2.8Å) is shown in *black* dashed line. The electron density map, countered at 1 σ , clearly shows that the conformation proposed does not correspond with the one built in pepMEK6-p38 α _{K53R}.

6.4.2.2 How does the peptide bind specifically to the protein?

6.4.2.2.1 MEK6 binds differently from MKK3b and MEF2A in the hydrophobic docking groove

Three structures of p38 α in complex with docking peptides are reported in literature: pepMKK3b-p38 α , pepMEF2-p38 α and pepMEK6-p38 α _{C162S} (Table 6.6). The mutation

C162S is present only in the pepMEK6-p38 α _{C162S} structure. The description of how the MEK6 peptide, used in this thesis, interacts with p38 α _{K53R} will be performed through the comparison with these published structures.

Table 6.6 Alignment of MEK6, MKK3b and MEF2A peptides. The residues visible in the electron density of each structure are highlighted in *yellow*.

Docking protein	Docking site	Structure
MEK6	SKGKKRNPG-LKI-PKEA	pepMEK6-p38 α _{K53R}
MEK6	SKGKKRNPG-LKI-PK	pepMEK6-p38 α _{C162S}
MKK3b	GKSKRKKD-LRI-SCNS	pepMKK3b-p38 α
MEF2A	RKPDLR-VVI-PPSS	pepMEF2-p38 α

$$R/K_{1-2} - (X)_{2-6} - \phi_A X \phi_B$$

As B-factor analysis and electron density distribution suggested, the docking interaction in pepMEK6-p38 α _{K53R} is mainly through the hydrophobic pocket. Predominance of a hydrophobic anchor has been described in pepMKK3b-p38 α , pepMEF2-p38 α and pepMEK6-p38 α _{C162S}. Surprisingly, the MEK6 peptide does not conserve all the hydrogen bond interactions characterised in these three structures and it assumes a different linear conformation in comparison to MEF2A and MKK3b (Figure 6.24). Indeed, the four peptides are quite different to each other, both in the number and position of the basic residues (Table 6.6).

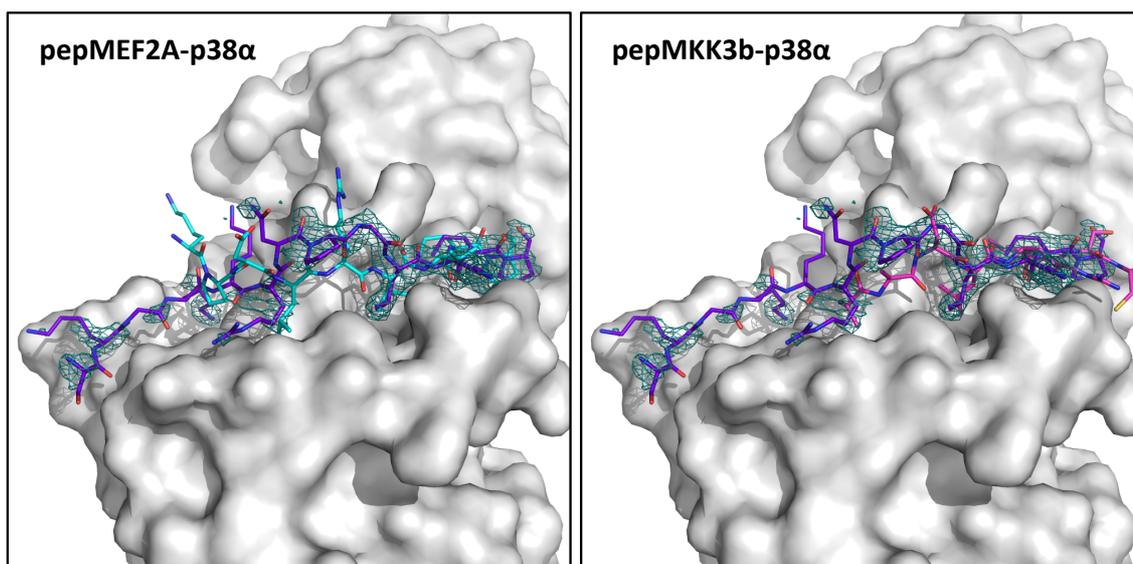


Figure 6.24 Differences in MEK6, MEF2A and MKK3b peptide conformation. The electron density map of MEK6 peptide from pepMEK6-p38 α_{K53R} is shown on the structure of pepMEF2A-p38 α (PDB code 1LEZ; Chang *et al.*, 2002) and pepMKK3b-p38 α (PDB code 1LEZ; Chang *et al.*, 2002). MEF2A peptide is in cyan, while MEK3b peptide is in magenta. The electron density map, countered at 1σ , shows that the MEK6 peptide use in this thesis assumes a different conformation form MKK3b and MEF2A peptides.

The interaction between p38 α and the peptide MEK6 backbone is made by two sequence-independent hydrogen-bonding interactions, involving ϕ_A and the X residue of the motif ϕ_A -X- ϕ_B (**K**), Gln120 in the α_d helix and Glu160 in the β_7 - β_8 reverse turn of p38 α . The backbone amide of **K** forms a hydrogen bond with the side chain oxygen of Gln120-O ϵ_2 (2.7 Å). The backbone amide of the ϕ_A hydrogen bonds to the backbone carbonyl of Glu160 (2.88 Å). In the pepMKK3b-p38 α and pepMEF2-p38 α structures further H-bonds are described, in particular with His126, which in pepMEK6-p38 α_{K53R} is too far from the peptide (Figure 6.25). In the pepMEK6-p38 α_{C162S} , the interaction with His126 is also conserved. However, since His126 is part of the same binding pocket of C162S, it is unclear whether the presence of the interaction is caused by the mutation.

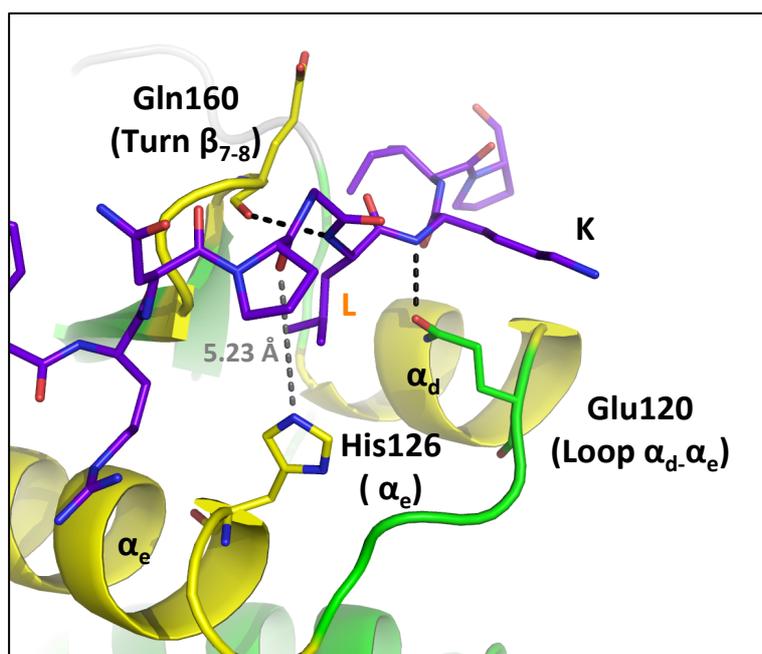


Figure 6.25 Hydrogen bonds in the inter-Pro peptide sequence. The hydrogen bonds between the C-terminal portion of the peptide and the protein are shown as *black* dashed lines. The distance between His126 and the peptide backbone is shown as a *grey* dashed line.

The hydrogen bonds observed are in the inter-Pro peptide sequence, where further stabilisation is achieved by hydrophobic contacts in a similar way to the complexes already described. The side chains of ϕ_A and ϕ_B residues bind to a continuous groove, making extensive contact with Ile116. The side chain of the ϕ_A makes VdW contacts with the side chains of Ile116, Leu122 in α_d , Val158 in β_7 , and Cys162 in β_8 . The ϕ_B residue contacts the side chains of Ala111 and Ile116 in α_d and Val158 in β_7 . Other hydrophobic residues, such as Leu113, Thr123, Val127, Phe129, Leu130, Ile131, Ile 134 and Ala157, contribute to enlarge the hydrophobic interactions.

All the interactions are stabilised by further electrostatic interactions. The side chain of the X residue (**K**) interacts with several water molecules with low B-factors ($17-22 \text{ \AA}^2$), which are hydrogen bonded to the Cys119 backbone oxygen atom. Since this residue interacts with Ile116, the final result is a better stabilisation of the hydrophobic interactions (Figure 6.26).

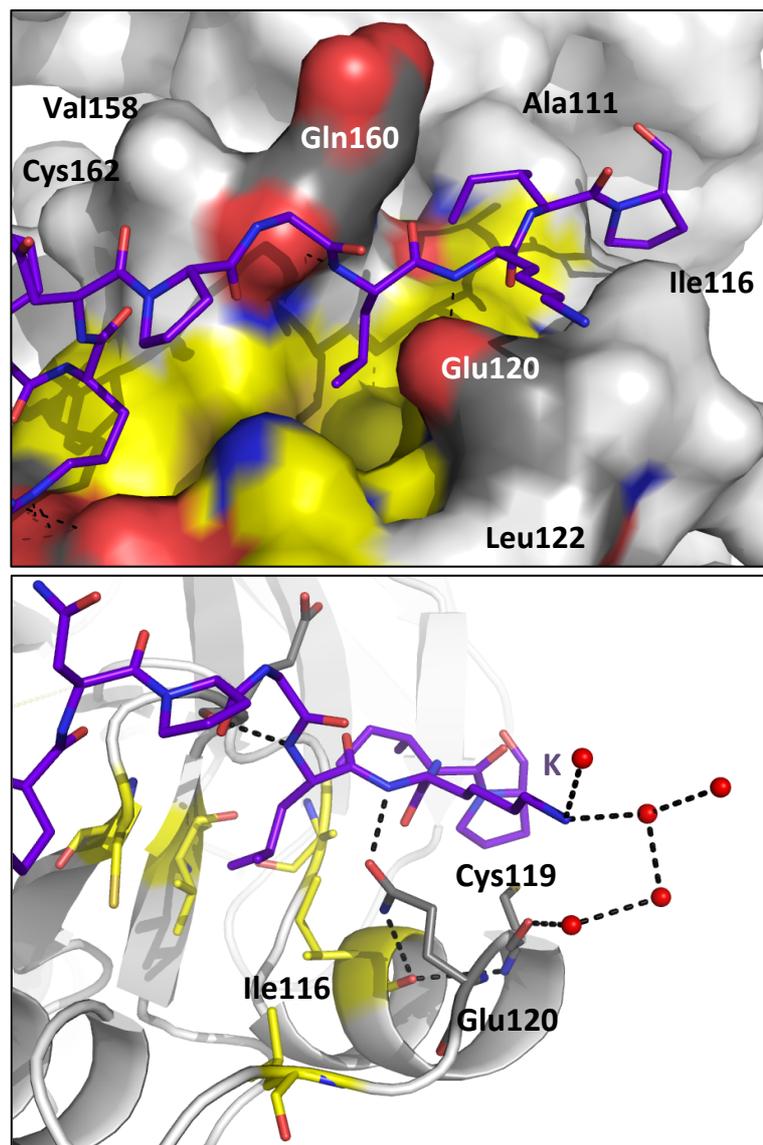


Figure 6.26 Hydrophobic interactions of the inter-Pro peptide sequence. The surface representation shows in *yellow* the residues involved in hydrophobic binding. Only residues that interact directly with ϕ_A or ϕ_B are indicated, for clarity. The bottom panel shows how through the water molecules pattern stabilises the peptide interaction with Ile116. MEK6 peptide is shown in *purple*. Water molecules are in *red* spheres. Dashed *black* lines represent hydrogen bonds.

Such a network of ordered water molecules is also present in the pepMEK6-p38 α_{C162S} structure. In the pepMKK3b-p38 α structure, the X residue is an arginine, whose guanidinium group interacts directly with the Cys119 backbone. In the pepMEF2-p38 α the X residue is a valine, a short hydrophobic residue, which cannot interact directly or indirectly with Cys119.

6.4.2.2.2 MEK6 shows unique binding with Loop16

Further electrostatic interactions are made through the N-terminal portion of the peptide, where 5 basic residues are present. Since the side chain of residues -9, -8 and -4 only are visible in the electron density, the interactions among these residues and the protein will be discussed.

It has been proposed that this basic region should bind to the CD domain. Modelling studies and NMR studies suggest that in solution this is possible for MEF2A and MKK3b peptides (Akella *et al.*, 2010; Chang *et al.*, 2002). However, no interaction is observed between the peptide and the aspartates of the CD domain in pepMEK6-p38 α _{K53R}. The K-8 side chain is too far (6Å) from Asp313. The structure of ERK2 in complex with a docking peptide shows that the concerted binding to the hydrophobic groove and CD domain is allowed by secondary structure elements such as a short α -helix. The electron density map of MEK6 peptide leaves no doubt about the absence of a similar conformation, which is in agreement with model of p38 α activation, elucidated in the introduction.

Interestingly R-4 interacts with residues belonging to Loop16, close to the protein acidic patch. The guanidinium group of R-4 is hydrogen bonded to the hydroxyl group of Tyr311, to the side chain of Gln310 and one of the alternative conformations of Asp125 of helix α_e (Figure 6.27). This binding is further tightened through H-bonds between Gln310 and Ala309 to S-9. The side chain of R-4 has been modelled in a small hydrophobic pocket, where Phe129, Cys162 and His126 are present. Surprisingly, a different residue is modelled inside this pocket for MEF2A or MEK6 in pepMEK6-p38 α _{S162} (a leucine and a proline respectively). The conformation of Asp125 is highly variable in the three structures. In pepMEK6-p38 α _{S162} two conformations are present, while in pepMEF2A-p38 α the conformation binding the R-4 in pepMEK6-p38 α _{C162S} is predominant. An exception is pepMEK3b-p38 α , where a lysine backbone has been modelled inside this pocket and Asp125 conserves the conformation of the apo-protein.

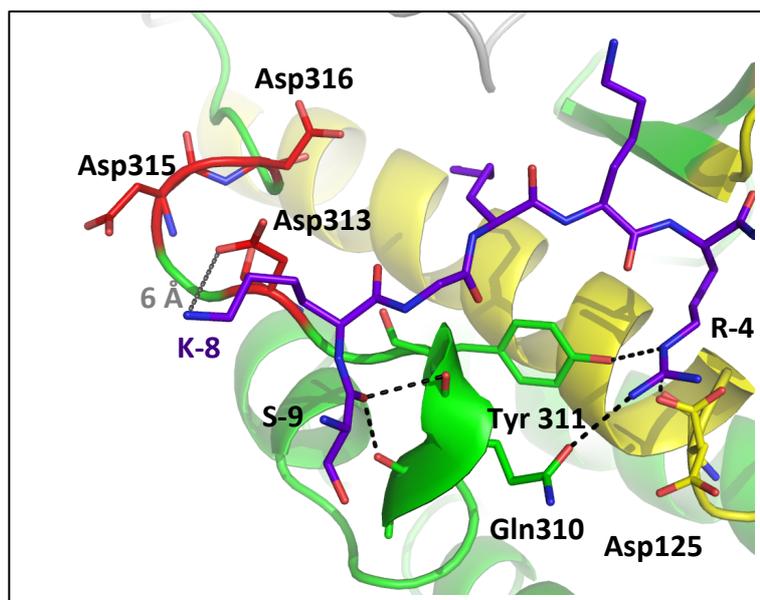


Figure 6.27 Electrostatic interactions of MEK6 N-terminus peptide. Protein residues involved in the interaction with MEK6 peptide are shown in sticks. The distance between K-8 and Asp313 is shown by *grey* dashed line. MEK6 peptide is shown in *purple*. Hydrogen bonds are in *black* dashed lines.

6.4.3 Peptide-induced conformational changes in p38 α K53R

Peptide binding cause several conformational changes, which are expected to prepare the protein for activation. These conformational changes induced by the peptide will be described through the comparison of the apo-structure (p38 α K53R) and the peptide-protein structure (pepMEk6-p38 α K53R) solved in this thesis.

The overlay is based on the C-terminal lobe, as described in the characterisation of the conformational changes induce by MKK3b and MEF2A peptide binding (Chang *et al.*, 2002).

6.4.3.1 Conformational changes at the D-motif binding site

No dramatic conformational changes occur at the D-motif binding site, where the MEK6 peptide is accommodated with the exception of the rearrangement of hydrogen bonds that occurs between L₁₆ and helix α_e . As described above, R-4 enters in a small pocket formed by Tyr311, Gln310 and Asp125. As result of these new interactions, Asp125 is in multiple conformations. Furthermore, the side chain of Asp124 rotates in a way that interacts with Tyr307, from the L₁₆, and Gln128 from α_e (Figure 6.28).

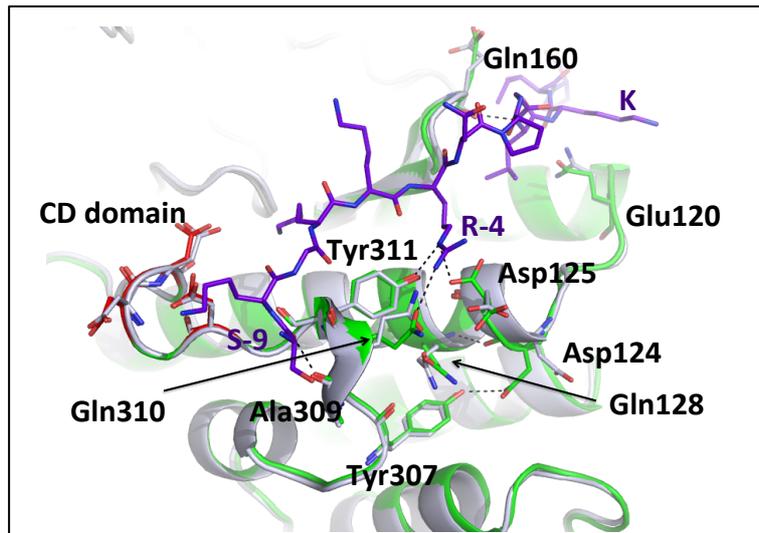


Figure 6.28 Conformational changes at the peptide binding site. The overlay of pepMEK6-p38 α _{K53R} (in *green*, with peptide in *purple*) on p38 α _{K53R} structure (in *light blue*) show the conformational changes induced by entering of R-4. Residues involved in peptide binding are shown in sticks. Hydrogen bonds are in *black* dashed lines.

However, whether Asp24 rotation is part of the activation mechanism is arguable. In the structures of the WT p38 α , Asp124 is already interacting with Tyr107 and Gln128 and this interaction is also maintained upon docking interaction in the pepMEF2A-p38 α and pepMEK3b-p38 α structures.

6.4.3.2 The peptide binding induces conformational changes at the P-Lip, DGF motif and Glycine-rich loop

Macro conformational structural rearrangements are expected upon peptide binding, which should facilitate the phosphorylation of the P-Lip and promote nucleotide binding. The P-Lip, DFG motif and Glycine-rich loop are all involved in these steps. Since in apo-p38 α _{K53R} all these protein regions are ordered, it is now possible to investigate which are the preliminary conformational changes that occur when MEK6 binds p38 α . Initially, peptide binding brings the two lobes closer together. The N-terminal lobe rotates on the C-terminal lobe by approximately 5.2° and a 0.4 Å translation (calculate with Dyn dom; Poornam *et al.*, 2009). During this process the hinge in the N-terminal lobe is broken (Figure 6.29). Residues 0-3 are not visible in the

N-terminal portion of p38 α in the pepMEK6-p38 α_{K53R} structures. However, it is arguable whether this event is caused by lobe rearrangement. For example, it could be caused by the fact that the protein N-terminus no longer interacts with the hydrophobic binding groove of the adjacent symmetry related molecule, which can explain a loss in stability and the absence of electron density for residues 0-3 in pepMEK6-p38 α_{K53R} .

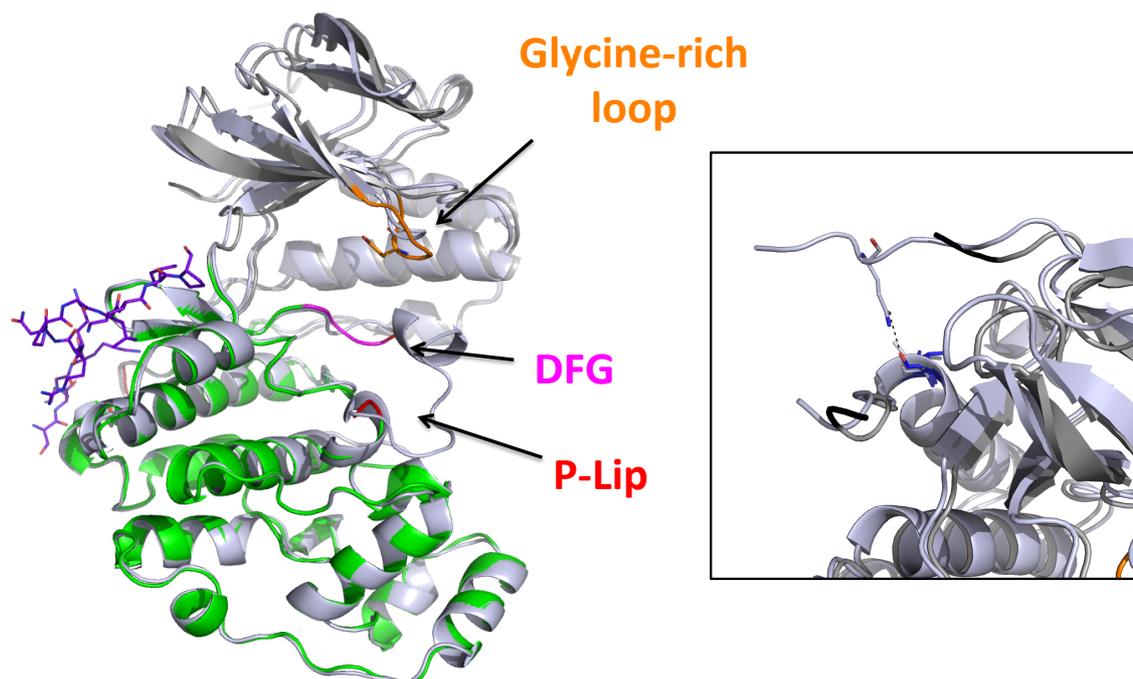


Figure 6.29 Rearrangement of the lobes upon peptide binding. The overlay of pepMEK6-p38 α_{K53R} (N-terminal lobe in *grey*, C-terminal lobe in *green*) on the p38 α_{K53R} structure (in *light blue*) shows that peptide binding at the C-terminal lobe causes rotation of the N-terminal lobe. The panel magnifies the changes at the hinge terminal region. N- and C-terminus of p38 α_{K53R} in pepMEK6-p38 α_{K53R} structure are in *black*. Hydrogen bonds are in *black* dashed lines.

Dramatic conformational changes are then observed (Figure 6.30). As expected, the phosphorylation lip in the C-terminal lobe is the site of the major conformational change (172-183). It becomes disordered upon peptide binding, from residue 171 to residue 183. The adjacent DFG motif (169-171) changes conformation, and the electron density surrounding the side chains, appears less clear. In the phosphate anchor region or Glycine-rich loop, Tyr35 rotates in a way that the hydroxyl group is coordinated by Ser32.

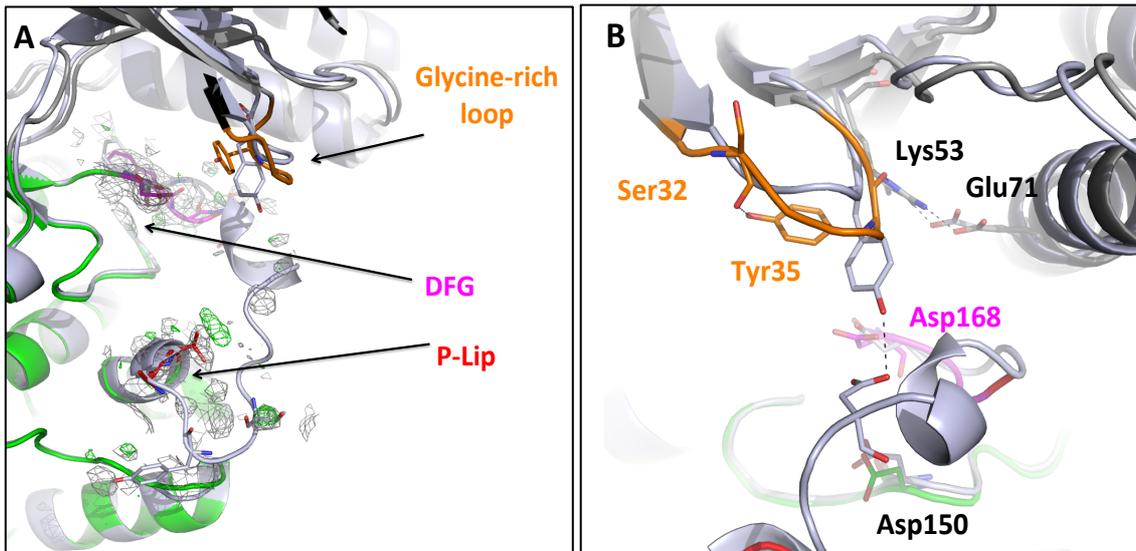


Figure 6.30 Conformational changes induced by peptide binding. The overlay of pepMEK6-p38 α _{K53R} (N-terminal lobe in *grey*, C-terminal lobe in *green*) on the p38 α _{K53R} structure (in *light blue*) show the conformational changes induced by peptide binding in P-Lip (in *red*), the DFG motif (in *magenta*) and the Glycine-rich loop (in *orange*). In the P-Lip the only visible residues are 183-185. Tyr35, DFG and P-Lip residues from pepMEK6-p38 α _{K53R} and DFG residues, Thr180, Tyr182 from p38 α _{K53R} are shown as sticks. A) Electron density map (in *grey*) contoured at 1 σ and difference electron density map (in *green*) contoured at 3 σ from pepMEK6-p38 α _{K53R} structure are shown on DFG motif and P-Lip of p38 α _{K53R}. B) The rotation of Tyr35 and the position of residues involved in catalysis are shown. Hydrogen bonds are in *black dashed lines*.

All these events promote the active conformation of p38. The disordered P-Lip will facilitate the rearrangement of the position of Arg67, Arg70, Arg149, Arg173, Arg186, and Arg189, which are supposed to form two basic grooves for accommodating P-Tyr182 and the P-Thr180 (Bellon *et al.*, 1999) and allow the region to be accommodated in the active site of MEK6.

The DFG motif is proposed to be involved in catalysis and binding of one catalytic magnesium ion. As it has been described for bis-phosphorylated p38 γ when the nucleotide is inside the active site, the catalytic residues are positioned for starting the reaction. Nucleotide entry should then stabilise the flexibility of DFG motif observed, with Asp168 coordinating one of the magnesium ions. Furthermore, Lys53 and Gln74, which participate in catalysis, are expected to move. In the pepMKK3b-p38 α structure conformational changes are also observed in His107 and Met109, where the nucleotide adenine is supposed to be accommodated. However, since p38 α is mutated in Lys53, all

these possible changes are not observed, except for His107, which is present in two conformations.

The most interesting conformational change, not previously observed, is the rotation of the sidechain of Tyr35 in the phosphate anchor region (Figure 6.30). The hydroxyl group of Tyr35 moves from coordinating the P-Lip to be hydrogen bonded by the Glycine-rich loop. The new conformation assumed is the same described in pepMEF2A-p38 α and pepMEK3b-p38 α structures, where it was not highlighted due to the lack of WT p38 α with an ordered Glycine-rich loop.

The rotation of Tyr35 could facilitate the entry of the nucleotide. The Glycine-rich loop is highly conserved among kinases and it is an integral part of the ATP-binding site. Its role is to anchor the ATP and to shield the bound nucleotide from solvent (Hemmer *et al.*, 1997). Comparison of the p38 α K53R structure with pepMek6-p38 α _{K53R} suggests that the movement of the tyrosine is like opening a gate. While in the inactive state the side chain and the interaction with P-Lip prevents the anchoring of the ATP due to steric clashes, the conformational change assumed upon MAPKK binding promotes the ATP recruitment and de-solvation by the glycine residues of the anchor region.

6.5 Conclusions

Two new structures have been presented in this chapter: p38 α K53R and pepMEK6-p38 α K53R. Through the comparison of these structures with multiple crystal structures, several goals have been achieved in this project.

First, the interaction between MEK6 docking peptide and p38 α was described at the molecular level through the pepMEK6-p38 α K53R structure. This structure lacks the mutation in the binding site described in the pepMEK6-p38 α C162S structure (Garai *et al.*, 2012), which makes it a more realistic model.

Second, in the pepMEK6-p38 α K53R structure the peptide N-terminus is partially visible, which has allowed possible CD binding *in crystallo* to be discussed for the first time.

Third, a new structure of p38 α has been presented (p38 α K53R), which provides for the first time a complete snapshot of the ordered un-phosphorylated p38 α . Due to the unique opportunity, conformational changes in the Glycine-rich loop and the P-Lip caused by the docking interaction have been described for the first time.

Two main questions will be addressed in this conclusion:

- How can the described results be helpful in drug design against the D-motif binding site?
- How do the described results improve the model of p38 α activation?

6.5.1 The R-4 binding pocket of p38 α , a third element of the D-motif binding interaction

As described in the introduction, the D-motif binding pocket is under investigation as a specific MAPK target. Through the comparison of pepMEK6-p38 α K53R with the other peptide-p38 α structures (pepMEK6-p38 α C162S, pepMEF2A-p38 α and pepMKK3b-p38 α , overview in figure 6.31), it was highlighted that MEK6 adopts a unique way to bind Loop16.

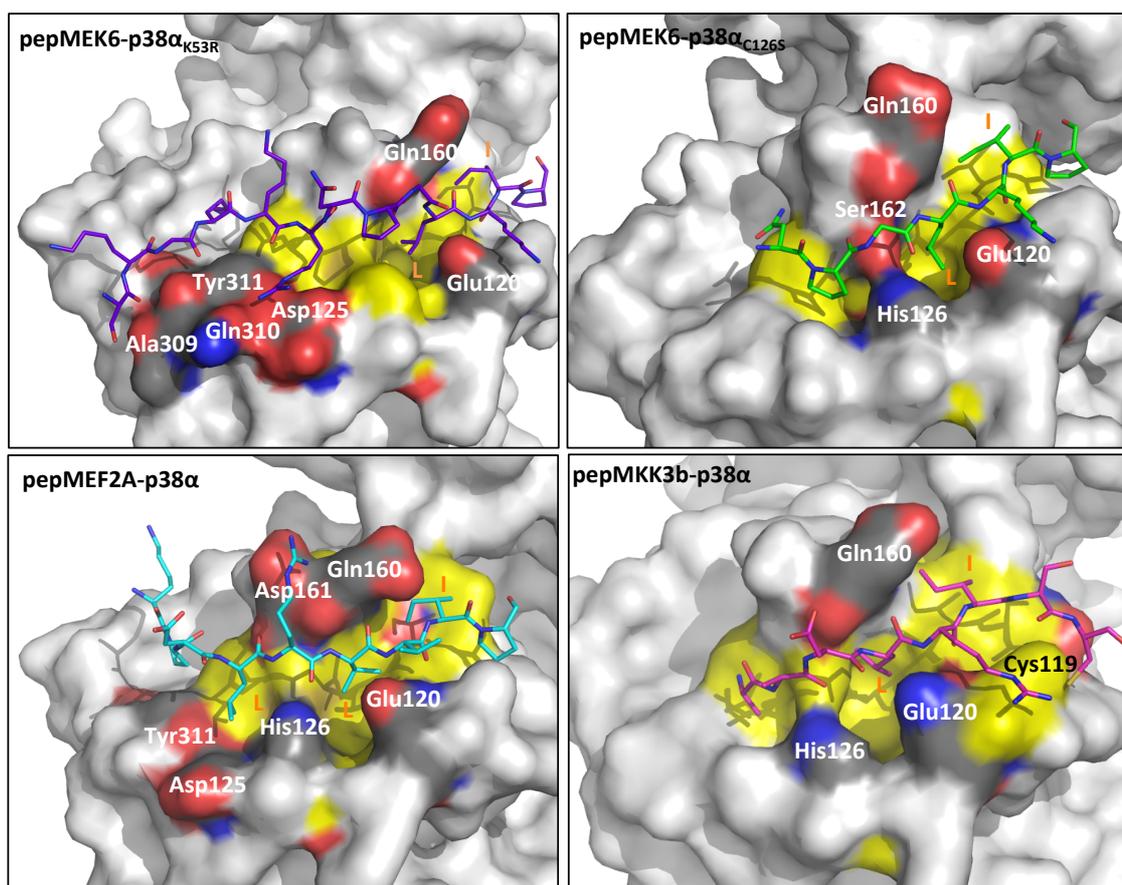


Figure 6.31 Peptide binding in p38 α . The peptide binding is shown in pepMEK6-p38 α_{K53R} , pepMEK6-p38 α_{C126S} , pepMEF2A-p38 α and pepMKK3b-p38 α structures (PDB code 1LEZ, 1LEW, 2Y8O; Chang *et al.*, 2002; Garai *et al.*, 2012) p38 α residues involved in electrostatic interactions and peptide hydrophobic residues are highlighted. His126 does not interact with the peptide in pepMEK6-p38 α_{K53R} structure. Tyr311 and Asp125 are involved in peptide interaction both with MEK6 and MEF2A peptide. The disulphide bond is only present between Cys119 and MKK3b peptide.

The R-4 side chain enters a hydrophobic groove where Phe129, Cys162 and His126 are present. The guanidinium group of R-4 is hydrogen bonded to the hydroxyl group of Tyr311, to the side chain of Gln310 and one of the alternative conformations of Asp125 of helix α_e . This pocket could represent a third element of the D-motif binding domain after the CD domain and the hydrophobic groove. Interactions with Tyr311 and Asp125 have been also observed in pepMEF2A-p38 α (Chang *et al.*, 2002). A similar binding pocket is also present in ERK and JNK. In the alignment shown in figure 6.32, the residues belonging to this pocket for the three human proteins p38 α , ERK2 and JNK1 are highlighted. While the differences between ERK2 and p38 α are minimal,

(Phe169_{p38α} is substituted by tyrosine), the p38α residues are less conserved in JNK1 (4 different residues).

```

p38α MSQ--ERPTFYRQELNKTIWEVPERYQNLSPVGSAGYSVCAAFDTKTGLRVAVKLSRP 58
ERK2 MAA-AAAAGAGPEMVRGQVFDVGPRTNLSYIGEGAYGMVCSAYDNVNKVRVAIKKIS-P 58
JNK1 MSRSKRDNFNYSVEIGDSTFTVLKRYQNLKPIGSGAQGIVCAAYDAILERNVAIKLSRP 60
* : : * * * * : * * * * * * * * * * * * * * * * * * * * * * *
p38α FQSIHAKRTYRELRLKHKHENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVK 118
ERK2 FEHQTYCQRTLREIKILLRFRHENIIGINDIIR-APTIEQMKDVYIVQDLMETDLYKLLK 117
JNK1 FQNQTHAKRAYRELVLKMCVNHKNIIGLLNVFTPKSLEEFQDVYIVMELMDANLCQVIQ 120
* : : * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
p38α CQKLTDDHVQFLIYQILRGLKYIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDD- 177
ERK2 TQHLNSNDHICYFLYQILRGLKYIHSANVLHRDLKPSNLLNTTCDLKICDFGLARVADPD 177
JNK1 M-ELDHERMSYLLYQMLCGIKHLHSAGIIHRDLKPSNIVVKSDC TLKILDFGLARTAGTS 179
* * : : * * * * * * * * * * * * * * * * * * * * * * * * * * * *
p38α -----EMTGYVATRWRAP EIMLNWMHYNQTVDIWSVGCIMAELLTGRTLFPGTDHIDQL 232
ERK2 HDHTGFLTEYVATRWRAP EIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPKGHYLDQL 237
JNK1 ----FMMPYVVTRYRRAPEVILG-MGYKENVDLWSVGCIMGEMVCHKILFPGRDYIDQW 234
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
p38α KLILRLVGTPGAELLKKISSESARNYIQSLTQMPKMNFAVFIGA-----NPL 280
ERK2 NHILGILGSPSQEDLNCIINLKARNYLLSLPHKNKVPWNRLFPNA-----DSK 285
JNK1 NKVIEQLGTPCPEFMKKLQ-PTVRTYVENRPKYAGYSFEKLFDPVLFPPADSEHNKLSAQ 293
: : : * * * * * : : : * * * * * : : : * * * * * : : : * * *
p38α AVDLLEKMLVLVDSKRI TAAQALAHAYFAQYHDPDDEPVAD--PYDQSFESRDLIDEWK 338
ERK2 ALDLLDKMLTFNPHKRIEVEQALAHAYLEQYDPSDEPIAEA-PFKFMELDDLPEKELK 344
JNK1 ARDLLSKMLVIDASKRISVDEALQHPYINWVYDPSAEAPPKIPDKQLDEREHTIEEWK 353
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
p38α SLTYDEVISFVPPP-----LDQEEMES----- 360
ERK2 ELIFEETARFQPGY-----RS----- 360
JNK1 ELIYKEVMDLEERTKNGVIRGQPSPLGAAVINGSQHPSSSSSVNDVSSMSTDP TLASDTD 413
.* :.*. :
p38α -----
ERK2 -----
JNK1 SSLEAAAAGPLGCCR 427

```

Figure 6.32 R-pocket residues in p38α, ERK2 and JNK1. The residues belong to the R-pocket are highlighted in yellow in the structure alignment of p38α, ERK2 and JNK1 MAPK.

Thus, through the analysis of the pepMEK6-p38α_{K53R} structure, a third binding element in the D-binding motif is proposed, which could be a possible determinant of D-motif binding site specificity. In this thesis the name R-pocket will be used to refer to it.

6.5.1 MEK6 might adopt a “scanning process” for recruiting p38α

In the pepMEK6-p38α_{K53R} structure the backbone of 12 of the 16 residues of the peptide are visible in the electron density. The MEK6 peptide binds through hydrogen bonds, which involve S-9 and R-4 side chains in the N-terminus, and the φ_A and K backbone in the hydrophobic C-terminus. Further hydrophobic interactions stabilise the peptide.

Importantly, the electron density of the basic region (SKGKKR) is sufficient to exclude interactions with the CD domain and secondary structural rearrangement. However, this is in contrast with mutational analysis of the CD aspartate residues in p38 α . Mutation of aspartates to alanines prevents the binding of the MKK3b peptide (Chang *et al.*, 2002). Moreover, NMR studies have revealed that the MKK3b peptide binds Asp315, Asp316, and Glu317 in solution, all in or close to the CD domain (Akella *et al.*, 2010).

Two possible explanations can be proposed. First, the peptide could assume a different conformation in solution. In the crystal structure of pep-ERK2 the peptide has a secondary structure arrangement that allows interaction with the CD domain. It is possible that in different buffer conditions, or in the full length protein, the docking region has a similar conformation. However, as described in the introduction, ERK and p38 are activated in different ways, meaning that the docking interactions involved could be different.

A second explanation is that the interactions at the CD domain and at the hydrophobic groove, occur sequentially, rather than simultaneously. The peptide-p38 α structure presented in this work is the first in which the N-terminal peptide is defined showing that no interaction with the CD domain is present. However, the MEK6 docking peptide has a D-motif that is one of the most enriched in basic residues, leaving open the question whether they could assume a role in MAPK binding. B-factor analysis of the peptide reveals that the intra-Pro region has lower B factors (max 30 Å²) while they increase in the basic patch, up to 60 Å². Individual occupancy refinement shows that the occupancy of the basic region is below 100%. Indeed, not all the Asp125 side chain population is coordinating the R-4 guanidinium group. Possibly, the basic N-terminal region is used by MEK6 as “search tool” for CD domain location, followed by the R-pocket binding (formed by Asp125, Gln310 and Tyr311) and eventually the hydrophobic groove, where the interaction will be better stabilised by hydrophobic interactions. A “scanning process” is proposed, where MEK6 uses electrostatic interactions to find the hydrophobic groove. Possibly, crystallisation of p38 α with the basic peptide region alone, will further elucidated this point.

6.5.3 How does the phosphorylation lip become disordered?

It has been proposed that binding of MAPKK to MAPK, induces conformational changes in the active site at the interface of the two lobes. Both the P-Lip and the ATP binding site are involved. The structures solved here show that upon MEK6 peptide binding, the phosphorylation lip becomes disordered. The well ordered structure of mutant p38 α provides a unique opportunity to investigate all the conformational changes occurring in the protein. Upon docking, three main events have been observed:

- the N-terminal lobe gets closer to the C-terminal lobe;
- in the Glycine-rich loop, the side chain of Tyr35 becomes coordinated by Ser32;
- the P-Lip becomes disordered.

How these events are connected to each other is an intriguing question. The K53R mutation has stabilised p38 α in the ordered inactive conformation, in which the Tyr35 side chain is coordinating the P-Lip. Upon binding at the C-terminal lobe, the MEK6 peptide induces conformational changes in the N-terminal lobe, in particular to the Glycine-rich loop, where an evident shift in position is observed. The final result is the breaking of an H-bond between Tyr35 and Asp175. Tyr35 is then stabilised by an interaction with Ser32, and the phosphorylation lip becomes disordered. The order of these events cannot be determined and an alternative scenario is where the peptide induces disorder in the phosphorylation lip. This event causes conformational changes in the Glycine-rich loop of the N-terminal lobe. In both cases the intimate contact between the two lobes is broken which leads to disorder in the P-Lip.

In the WT protein, where both the phosphorylation lip and Glycine-rich loop are unstable, an equilibrium between the two extremes (completely ordered to disordered loop) is expected. In this equilibrium, different Tyr35 conformations correspond to different grades of disorder in the phosphorylation lip. This explains why these residues are disordered (occupancy zero for both Tyr35 side chain) or have high B-factors in the WT p38 α structures. The overlay of all the structures solved so far with and without peptide suggests that upon binding, the Tyr35 moves closer to the Ser32 backbone (Figure 6.33). Since nucleotide binding is expected, this conformational change is likely opening a gate, which then allows the phosphate anchor or Glycine-rich loop to recruit and de-solvate the nucleotide molecule. Mutation analysis of Tyr35 could allow further investigation of the dynamics occurring in the active site.

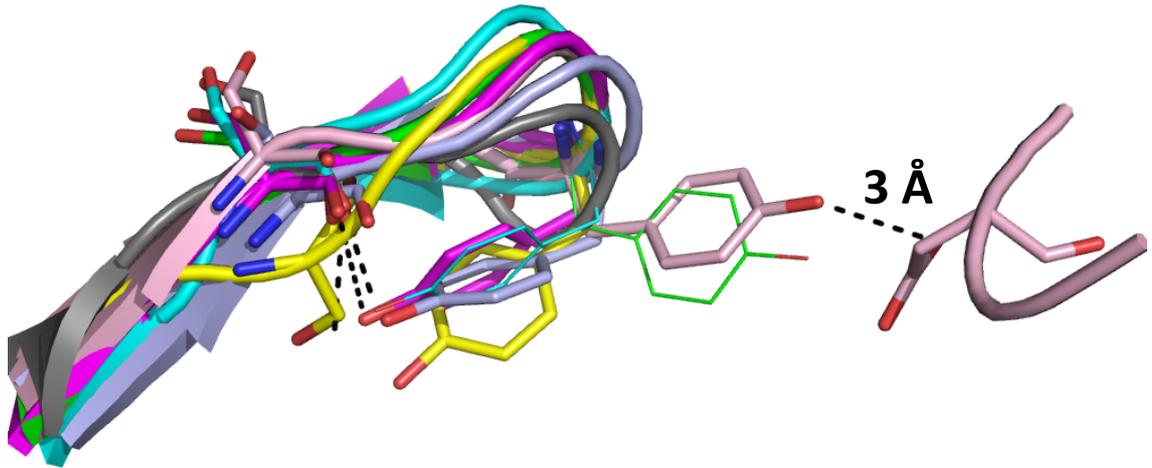


Figure 6.33 Conformational changes of Tyr35. The conformational changes of Tyr35 is shown through the overlay of human p38 α (in *green*), murine p38 α (in *cyan*), p38 α K53R (in *pink*), pepMEK6-p38 α K53R (in *grey*), pepMEK6-p38 α C162S (in *purple*), pepMEF2A-p38 α (in *magenta*)and pepMKK3b-p38 α (in *yellow*) . Residues with occupancy zero are shown in lines.

The results shown in this chapter describe at molecular level the effects of MEK6 docking interaction of MEK6 on p38 α . Three are the new proposals:

- the R-pocket is a third binding element in the D-motif binding groove of MAPK, which can be a determinant of D-motif binding groove specificity;
- binding to the CD domain is relevant in p38 α recruitment (scanning process), but not in p38 α activation;
- the docking interaction with p38 α breaks any possible interaction between the P-Lip and the Glycine-rich loop by promoting the rotation of Tyr35 hydroxybenzene group.

While the first proposal refers to the entire MAPK family, the scanning process and the role of Tyr35 in MAPK activation are currently p38 α exclusive. Indeed, this information could be used for a specific p38 α drug design. A drug, which bound to the p38 α CD domain, could hinder the scanning process, but since the MAPKK-CD domain interaction is a transient interaction, it would not completely prevent the interaction at the hydrophobic groove. Drugs, which target the hydrophobic groove, will be more

efficient, but less specific, due to the high similarity of this groove among MAPKs. Possibly, a drug binding both the hydrophobic groove and the R-pocket will increase the compound specificity. Finally a specific p38 α drug could target the Glycine-rich loop and blocking the Tyr35 in the conformation described in p38 α K53R structure. The structures described here, potentially provide several key information for a successful MAPK-specific drug design and increase the fundamental understanding of activation of p38 α by MEK6.

Chapter 7 The non-canonical MEK6 phosphorylation pathway

7.1 Introduction

7.1.1 How does MAPKK double phosphorylate MAPK?

In the previous chapter a docking interaction between MEK6 and p38 α has been described. This interaction occurs at a location outside of the active site, and it is postulated to be followed by a transient enzyme-substrate interaction at the active site (Tanoue *et al.*, 2001). The goal of the transient reaction is the activation of the MAPK by the double phosphorylation of the P-Lip, at residues Tyr182 and Thr180. How the double-phosphorylation reactions occur is still undefined. In particular, whether protein dissociation occurs between the two reactions is under discussion. Two possible mechanisms have been proposed: the processive and the distributive mechanism. Before discussing the aim of the project, the two mechanisms will be elucidated and open questions addressed.

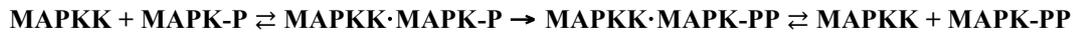
7.1.1.1 Processive vs distributive mechanisms

The first phosphorylation mechanism proposed was the processive mechanism. In this mechanism the active MAPKK protein collides with and binds to the MAPK, phosphorylates it at one site, then in some way moves to align the second phosphorylation site for phosphorylation, phosphorylates MAPK a second time, and then dissociates, yielding a double phosphorylated, active MAPK protein.



The second mechanism is the distributive mechanism. In this mechanism, active MAPKK collides with and binds to MAPK, phosphorylates it once and then completely releases the mono-phosphorylated product. The monophosphorylated MAPK-P then collides with a second molecule of active MAPKK, becomes phosphorylated a second

time, and is released as MAPK-PP.



While the first mechanism is MAPKK concentration independent, the second is concentration dependent. This generates a different stimulus/response curve. For a processive mechanism a hyperbolic curve would be expected, and for a distributive mechanism the stimulus/response curve would have a sigmoidal shape, due to initial accumulation of the mono-phosphorylated MAPK species. Several studies have suggested that the first phosphorylation event occurs on the tyrosine. Threonine phosphorylation of ERK2 MAPK by MAPKK-1 lags substantially behind the tyrosine *in vitro* (Haystead *et al.*, 1992; Scott *et al.*, 1995). Similar results have also been obtained *in vivo* for ERK1 (Robbins & Cobb, 1992). Mechanistic studies of MAPKK phosphorylation of MAPK have been carried out on *Xenopus* oocyte and rat ERK2, by human MEKK1 (Burack & Sturgill, 1997; Ferrell & Bhatt, 1997). It was shown that MAPKK carries out its dual phosphorylation of p42 MAPK by a distributive mechanism, where the amount of mono-phosphorylated MAPK becoming bis-phosphorylated is depended on the MAPKK-1 concentration. Data indicated that the tyrosine is the first phosphorylation site. Similar data on the MAPK cascade has been obtained by Dalby and co-workers (Waas *et al.*, 2001). They demonstrated that p38 α phosphorylates the transcription factor ATF2 on Thr-69 and Thr-71 by a distributive mechanism, with the finding that mono-phosphorylated GST-ATF2 (residues 1–115 of full length ATF2) dissociates from p38 α with a rate constant greater than, or equal to 0.6 s⁻¹ exactly the magnitude of GST-ATF_{21–115} turnover.

7.1.1.2 *In vitro* vs *in vivo*

However, the scenario proposed *in vitro*, could be different from what is occurring *in vivo*. The evidence that MAPKs interact with upstream and downstream proteins through docking-interactions suggests a processive mechanism could be favourable. In

this mechanism, the repositioning of the second phosphorylation site in the active site could be brought about through minimal reorganisation of the intermediate enzyme-substrate complexes, while contact is maintained at the docking site. MAPKK and MAPK could carry on the reaction by simply “sliding” the active site from one phosphorylation site to the next. Moreover, experiments *in vitro* do not consider the participation of additional proteins. In these experiments substrates are freely diffusible with respect to activators and *vice versa*. This causes difficulties in achieving full phosphorylation of MAPK with MEK *in vitro* (Burack & Sturgill, 1997; Haystead *et al.*, 1992). It has been proposed that scaffold proteins could facilitate the interaction between MAPK protein cascades and possibly affect the type of mechanism adopted. A scaffold protein might stabilise a complex of active MEK and MAPK so that it has a lifetime longer than the k_{cat} , causing activation to occur processively (Burack & Sturgill, 1997).

Recently, Matsuda and co-workers have investigated the application of a distributive mechanism *in vivo* to activation of ERK2 by MEK1. Data on MAPK phosphorylation in HeLa cells after EGF activation reveal that the signal-response curve has a more linear gradient than the sigmoidal one proposed (Ferrell & Bhatt, 1997). This is due to the fact that the increase of pY-ERK2 did not precede, but rather accompanied, the increase in pTpY-ERK2. The effect of molecular crowding has been proposed as the reason for the observation of a distributive nature of ERK phosphorylation (Ellis, 2001a; Ellis, 2001b; Minton, 2006). This is because molecular crowding is a critical difference between the environment *in vitro* and *in vivo*. Indeed, the re-analysis of the same phenomenon in diluted cytoplasmic cell extract revealed a distributive mechanism. The processive mechanism was recovered by simply adding 15% PEG (w/v) as crowding agent. In this scenario, scaffolding proteins could participate due to molecular crowding. However, in contrast with the processive mechanism, dissociation between the first and the second phosphorylation has been observed. Eventually a “**quasi-processive phosphorylation**” was proposed, where the diffusion of both MEK and pY-ERK are considerably restricted by the molecular crowding effect. Therefore, MEK rebinds to and phosphorylates pY-ERK with higher probability, with regard to a distributive mechanism (Aoki *et al.*, 2011). These data fit with a previous particle based simulation study, where it was suggested that fast rebinding of MAPK-P by MAPKK could revert

the distributive mechanism to a more processive-like one *in vivo* (Takahashi *et al.*, 2010).

7.2 Aim of this project

The question of which mechanism is occurring *in vivo* is still open. In particular, it will be of interest to investigate whether a different mechanism can be related to the different conformational changes in the phosphorylation lip among MAPKs during the activation process (see Chapter 6). In the case of p38, a distributive mechanism implies a re-binding of the disordered P-Lip, while in a processive mechanism, MEK could bind and stabilise the phosphorylation lip once. However, in a fully processive mechanism, the exchange of the nucleotide will be a greater challenge than in the distributive. The reactivation time of a kinase will depend on the order in which ADP and the modified substrate dissociate and ATP and substrate bind to it. A recent study on a protein kinase revealed that the modified substrate must dissociate from the enzyme before the ADP, and that the ADP/ATP exchange must take place before the enzyme can bind to the substrate again (Aoki *et al.*, 2011). This is in agreement with the fact that kinases use ATP as phosphate donor, while the protein-ADP complex is the inactive state.

In order to investigate these questions, it was decided to study the phosphorylation reaction catalysed by MEK6 using metallofluoride complexes. Published data show that an aluminium fluoride molecule was formed in solution when p38 α was mixed with MEK6 and ADP (Xiaoxia *et al.*, 2011). However, attempts to trap the p38 α -ADP- AlF_4^- -MEK6-TSA in the solid state failed. This is because it was found that the phosphorylation reaction was not completely inhibited. Thus, a series of qualitative and preliminary experiments were performed to investigate how ATP and ADP are involved in the reaction. Surprisingly, it was found that ADP can be used as phosphate donor.

7.3 Experimental procedures

7.3.1 Plasmid and protein expression-purification

Constructs used and protocols for protein expression and purification has been described in the previous chapter, section 6.3.1.

7.3.2 Native PAGE gels

7.3.2.1 Native PAGE gel composition

Native PAGE gels were always freshly prepared. The protein sample was mixed with loading buffer without heat denaturation. All gels were pre-run for at least 20 minutes, in order to remove possible impurities. Gels were run at 20-30 mA under constant current at 4°C, and stained in fresh Coomassie Blue stain solution then destained until the background was colourless (see Chapter 3 for stain and destain solutions composition).

At the beginning of the project, 5% native PAGE gels were used (Figure 7.3). Later it was decided to use exclusively 7.5% gels. Gel compositions are shown in table 7.1:

- Native gel composition table (Table 7.1):

	1.5 M TrisHCl pH8.8	0.5 M TrisHCl pH6.8	10% Ammonium Persulfate (APS)	30% Acrilammide, 0.8% Bis- Acrilammide	TEMED	H ₂ O
RUNNING GEL (7.5%-1 gel)	1.75 ml	-	0.025 ml	2.3 ml	0.060 ml	3.075 ml
STACKING GEL	-	1.25 ml	0.050 ml	0.5 ml	0.005 ml	3.2 ml

- Running buffer 1X: 0.05 M Trizma , 0.38 M Glycine, adjust to pH8.8
- Loading buffer 5X: 5 ml Glycerol 100%, H₂O 2.7 ml, 2.13 ml of 0.5 M TriHCl pH6.8, Bromo Phenol Blue traces.

7.3.2.3 Sample preparation for Native PAGE gel

Figures 7.3, 7.5 and 7.6, 7.7 show qualitative native PAGE gels. Each sample loaded contains proteins in equimolar amounts (1:1) with an excess of nucleotide (10 mM) (Table 7.2). Each sample was incubated for 20 minutes at RT before loading.

Figures 7.4, 7.8, 7.9 and 7.10 report qualitative native PAGE gels, where in the amount of protein was defined based on the criteria of image quality. Activity buffer was further added (see section 7.3.4.1 for the solution composition).

Table 7.2 Native PAGE gel sample composition

Samples	MEK6 (23µM)	p38α (23µM)	MEK6 buffer	p38 buffer	ATP (0.5 µM)	ADP (0.5 µM)	ACTIVITY BUFFER 5X
MEK6	8.7 µl (7µM)	-	-	8.7 µl	-	-	6 µl
MEK6+ATP	8.7 µl (7µM)	-	-	8.7 µl	0.6 µl (10mM)	-	6 µl
MEK6+ADP	8.7 µl (7µM)	-	-	8.7 µl	-	0.6 µl (10mM)	6 µl
p38α	-	8.7 µl (7µM)	8.7 µl	-	-	-	6 µl
p38α+ATP	-	8.7 µl (7µM)	8.7 µl	-	0.6 µl (10mM)	-	6 µl
p38α+ADP	-	8.7 µl (7µM)	8.7 µl	-	-	0.6 µl (10mM)	6 µl
MEK6+p38α	8.7 µl (7µM)	8.7 µl (7µM)	-	-	-	-	6 µl
MEK6+p38α+ATP	8.7 µl (7µM)	8.7 µl (7µM)	-	-	0.6 µl (10mM)	-	6 µl
MEK6+p38α+ADP	8.7 µl (7µM)	8.7 µl (7µM)	-	-	-	0.6 µl (10mM)	6 µl

Reactions were stopped by the addition of 6 µl of loading buffer, containing 100mM EDTA. Only 10 µl of this final composition was loaded.

EDTA was not present in the original loading buffer as it was discovered later that nucleotides were diffusing into the gel and activating the reaction, increasing background phosphorylation.

7.3.3 Western blots of native PAGE gels

Western blot experiments were performed using a polyclonal antibody, able to recognise both mono-phosphorylated and bis-phosphorylated p38α (Cell Signaling technology, Boston, MA, USA). As the secondary antibody, an IgG anti-rabbit was used, conjugated with alkaline phosphatase (Sigma, St.Louis, MO, USA). A SIGMA

premixed buffer was used for development (BCIP/NBT, B6O4-Sigma, St.Louis, MO, USA). The protocol is still in development; table 7.3 shows the major steps. Suggestions were taken from the Cell Signalling Technology website (www.cellsignal.com).

Table 7.3 Western blot protocol

STEPS	METHODS	SOLUTIONS
1. Native gels run	All native gels were run in duplicate.	
2. Protein Blotting	One gel was used for Protein Blotting. Electrotransfer was performed with nitrocellulose membrane, at 250 mM for 45' to 1 hour. The transfer was checked with Red Ponceau staining.	Transfer buffer = 25 mM Tris base , 0.2 M glycine, 20% ethanol (pH8.5)
3. Blocking	The membrane was incubated for one hour in the blocking solution	Blocking/Ab solution = TBS 1X, 5% BSA TBS 10X-1L = 24.2 g Tris base, 80g NaCl (pH 7.6)
4. Primary Antibody incubation	The Ab was diluted 1:10000 in the Blocking/Ab solution, ON, at 4°C, with gentle shaking.	Blocking/Ab solution = TBS 1X, 5% BSA
5. Washing	Washing was performed three times for 5 minutes, on gently shaking, with the washing solution.	Washing solution = TBS 1X, 5% BSA, Tween 0.1%
6. Secondary Antibody incubation	The Ab was diluted 1:30000 in the Blocking/Ab solution, for one hour, at RT, with gentle shaking.	Blocking/Ab solution = TBS 1X, 5% BSA
7. Washing	See step 5.	
8. Protein detection	The membrane was incubated with Tris pH 9 for one minute. Then, incubated with the developer solution (BCIP/NBT) for 1-2'	100 mM Tris pH9

7.3.4 Kinetic assays

For kinetic assays, MEK6 and p38 α buffers were exchanged to 50 mM HEPES pH 6.8, 150 mM NaCl, 2 mM DTT, 10 mM MgCl₂. The details of how the reactions were performed and the results analysed are below.

7.3.4.1 Phosphorylation reactions

The phosphorylation reaction mixture consisted of 1 μ M MEK6 (or MEK6DD), 5 μ M p38 α , 10-10000 μ M nucleotide, 50 mM HEPES pH 7.5, 100 mM NaCl, 10 mM MgCl₂ and 1 mM DTT. Reaction volumes were 100 μ l. The concentrations of the proteins were chosen based on the finding of Ferrell and co-workers, which defined that MAPKK-1 and p42 MAPK work *in vivo* at 1:5 ratio (Ferrell & Bhatt, 1997). The activity buffer, at 5X concentration, was 250 mM HEPES pH 7.5, 50 mM MgCl₂, 500 mM NaCl and 5 mM DTT. Protein concentration was estimated using a thermo Scientific Nanodrop (see Chapter 2, Section 2.4.1).

Reactions were initiated by the addition of nucleotide and terminated every 5 minutes for twenty minutes with 200 mM EDTA. Then samples were loaded on a native PAGE gel. For each time point, 18 μ l of the reaction plus 6 μ l of 1 M EDTA was loaded.

Finally, in order to check the real endpoint of the kinetic experiments, each sample was left ON and run on a native gel the day after (the endpoint sample).

7.3.4.2 Phosphorylation reaction analysis

The bands obtained on native PAGE gels were quantified using ImageJ (Schneider *et al.*, 2012). The procedure used is explained in detail in figure 7.1 for endpoint kinetic experiments.

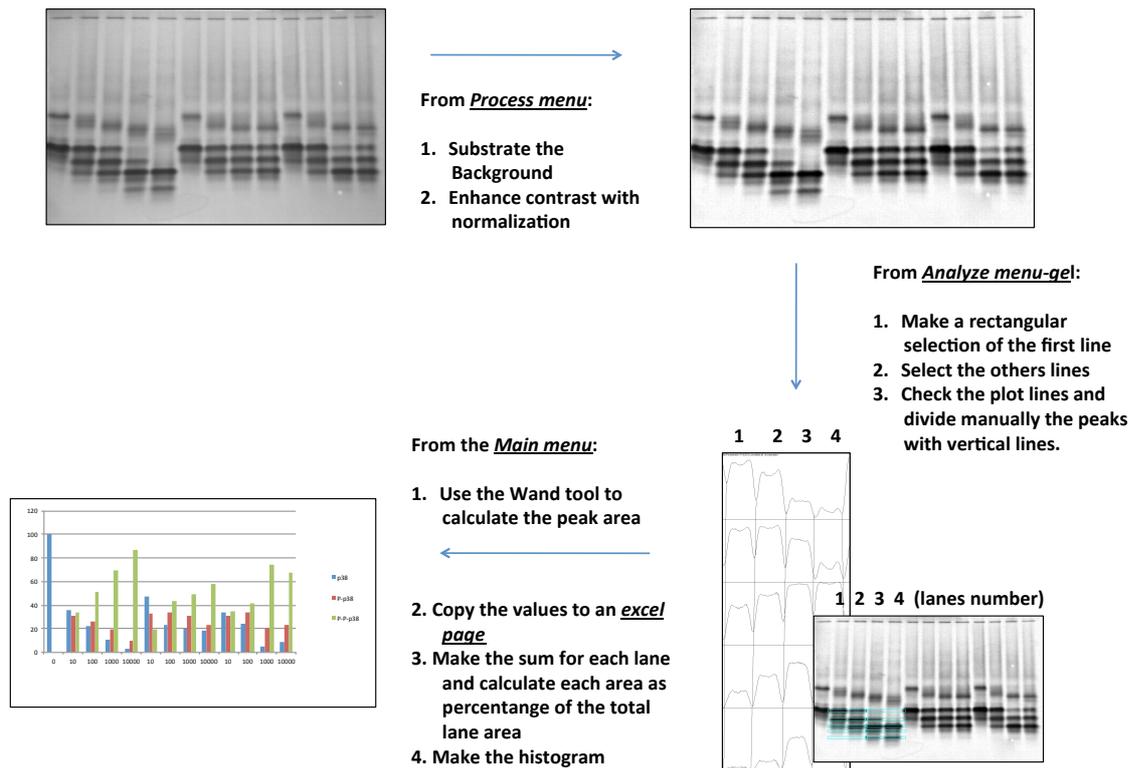


Figure 7.1 Steps performed to quantify gel bands using ImageJ (Schneider *et al.*, 2012). The rectangular selection was performed on maximum of 4 lanes at the same time.

7.3.5 Nucleotide analogues

Several nucleotide analogues were used during the experiments. The figure shows the relative structures (Figure 7.2).

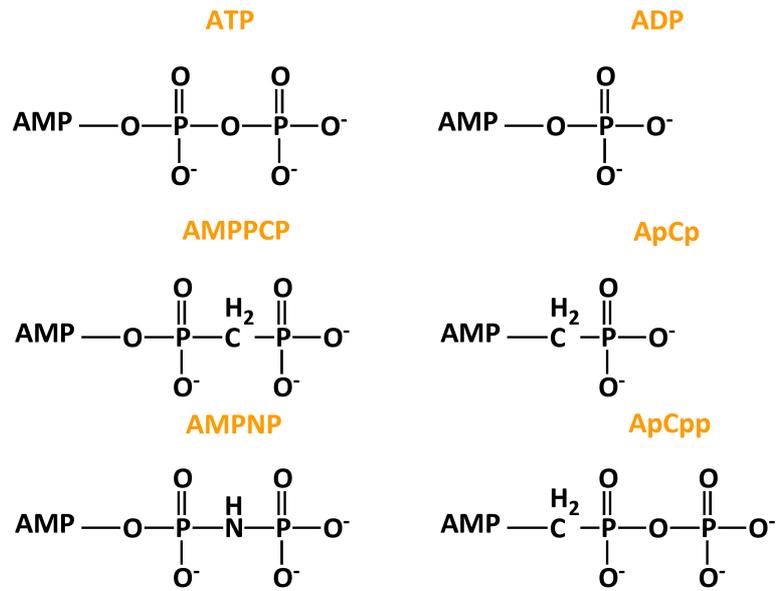


Figure 7.2 Nucleotides and nucleotide analogues.

7.4 Results and Discussion

7.4.1 MEK6-p38 α complex runs as a double band on native PAGE gel

In order to characterise the mechanism of activation of p38 α by MEK6 a crystal structure of the complex is necessary. In experiments to attempt to crystallise p38 α in complex with MEK6-AMPPCP or MEK6-ADP-AlF $_4^-$, several techniques were used to test the formation of the complex in different conditions. As the first goal of the work was to trap in solid state the transition state analogue, initial experiments focussed on the MEK6-p38 α -ADP-AlF $_4^-$ -TSA complex. Initially, MEK6DD, a mutated and constitutively active form of MEK6, and WT p38 α were used.

Classical size exclusion chromatography was not considered as a good candidate for testing complex formation. The two proteins have similar molecular weight and are both dimers in solution. Thus, homodimers and heterodimers were indistinguishable in the elution profile of the size exclusion chromatogram.

Eventually native PAGE gels were revealed to be the most successful method for characterising the formation of the complex. The successful protocol was defined as:

1. Prepare a TSA solution 3X (15 mM ADP, 6 mM AlCl $_3$, 30 mM NH $_4$ F in protein buffer 3X)
2. Mix the TSA solution with equimolar amounts of MEK6DD and p38 α
3. Incubate for one hour on ice (ter Haar *et al.*, 2007)
4. Run a 7.5% basic native PAGE gel in order to visualise complex formation

Analysis by native PAGE gels showed the complex as two new bands, more positively charged with regard to the p38 α band (Figure 7.3). As the p38 α homodimer has the same size as the heterodimer, the difference between p38 α and the complex mobility shift can be related only to the difference in charge.

MEK6DD - TSA	- p38 α TSA	MEK6DD p38 α TSA
--------------------	--------------------------	-------------------------------

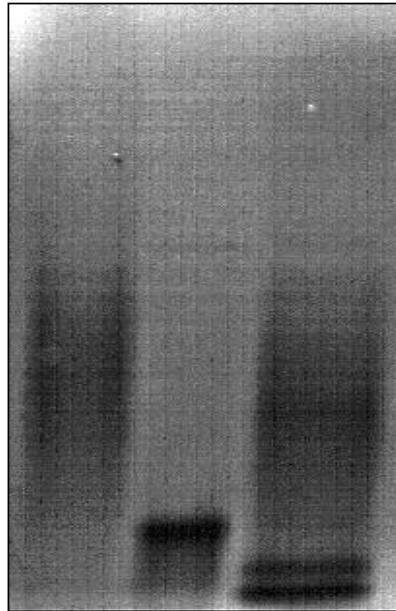


Figure 7.3 Complex detection on a native PAGE gel. On the 5% native PAGE gel shown three samples were loaded: MEK6DD alone plus TSA solution (ADP+AlCl₃+NH₄F), p38 α alone plus TSA solution and the mix of MKK6DD+p38 α +TSA.

The crystallisation trials of this complex at 10 mg/ml did not produce crystals. As the efficiency of complex formation is around 95-100% and the nature of the bands is unknown, the following hypotheses were tested:

- the efficiency of the complex formation is not 100%; this means that monomers or homodimers remain in solution, decreasing the sample homogeneity;
- as there are two bands and p38 α has two phosphorylation sites both exposed to solvent (Wang *et al.*, 1997), the two bands could be related to two different complexes in which MEK6 interacts with each of the phosphorylation sites.

The complex was prepared with an excess of MEK6DD (1:1.2 as p38 α :MEK6), using the construct with an uncleavable His₆-tag. In order to separate the complex from excess MEK6DD, a Ni-NTA column was used. Then an anion exchange column was run with

a NaCl gradient in order to separate the two bands. The results were analysed using native and SDS PAGE. The data showed that the double bands were not separated after anion exchange. Moreover, the SDS PAGE data indicated a small increase in the p38 α mass.

Since the presence of two bands was related to sample heterogeneity, and possibly to unsuccessful crystallisation, whether the nature of the two bands was related to the phosphorylation reaction was investigated.

7.4.2 ADP is a secondary phosphate donor in MEK6 phosphorylation reaction

In the next sections it will be shown that MEK6 can use ADP as second preference phosphate donor after ATP for activating p38 α . Indeed, the protein shift observed in the native PAGE gel of figure 7.3, is caused by phosphorylation at the P-Lip. In the classical phosphorylation pathway a kinase uses two ATP molecules as phosphate donors in order to double phosphorylate the substrate. Here, a non-canonical phosphorylation pathway is proposed, where ADP can remain trapped in the MEK6 nucleotide binding site and be used to activate p38 α .

First, endpoint kinetic experiments of the phosphorylation reaction in the presence of both ATP and ADP were performed. Analysis of the results was performed on 7.5% native PAGE gels. The nature of each band obtained was defined by mass spectrometry analysis (CEA, Grenoble, France) and western blots, in which an antibody able to recognise both the mono- and bis-phosphorylated P-Lip was employed.

Second, several controls were then performed. In order to eliminate the possibility that p38 α auto-phosphorylation was in some way causing the observed results, the end-point kinetic experiments were repeated using the mutant p38 α K53R. Furthermore, in order to eliminate the possibility of the presence of an extra-source of ATP caused by Adenylate kinase, the ADP phosphorylation experiments were run in the presence of an Adenylate kinase inhibitor. Third, a series of kinetic experiments were then performed to investigate the functional role of ADP phosphorylation.

7.4.2.1 Analysis of the native PAGE gels by mass spectrometry

In the first endpoint kinetic experiments, the phosphorylation reaction performed by MEK6 on p38 α was checked in the presence of both ATP and ADP. In these experiments, proteins are in equimolar ratio, (in the μ M range), while nucleotides are in excess (in the mM range).

Two MEK6 and three p38 α constructs were used:

- MEK6, which is the full length form of human MEK6. This construct was used in order to check the activity of WT MEK6. Moreover, it was used to verify if the protein was already active upon the expression-purification process;
- MEK6DD, which is the full length and constitutively active form of human MEK6. This construct was used in order to check whether MEK6DD behaviour is similar to WT MEK6;
- WT p38 α ;
- p38 α K53R, which is an inactive form of p38 α . This construct was used in order to prevent any MEK6 phosphorylation-activation by p38 α (as suggested in Ferrell & Bhatt, 1997) or possible p38 α autophosphorylation;
- p38 α T180A, which is an inactive form of p38 α . This construct was used for checking the ADP reaction when only one phosphorylation site (Tyr182) is available.

The native PAGE gels obtained revealed a similar pattern (Figure 7.4-7.5). Both MEK6 and MEK6DD bands are smeared, while p38 α runs as more defined bands. When the MEK6-p38 α or the MEK6DD-p38 α sample is supplied with ATP, the p38 α band shifts. This occurs for all three p38 α constructs. In order to further characterise the native PAGE gels, the bands were analysed by mass spectrometry. Mass spectrometry experiments were carried out as a service at the CEA, Grenoble (France). The results obtained show that the p38 α band shift or shift one is in accordance with phosphorylation at the P-Lip. Moreover, the mass spec data indicate that the first residue phosphorylated is Tyr182.

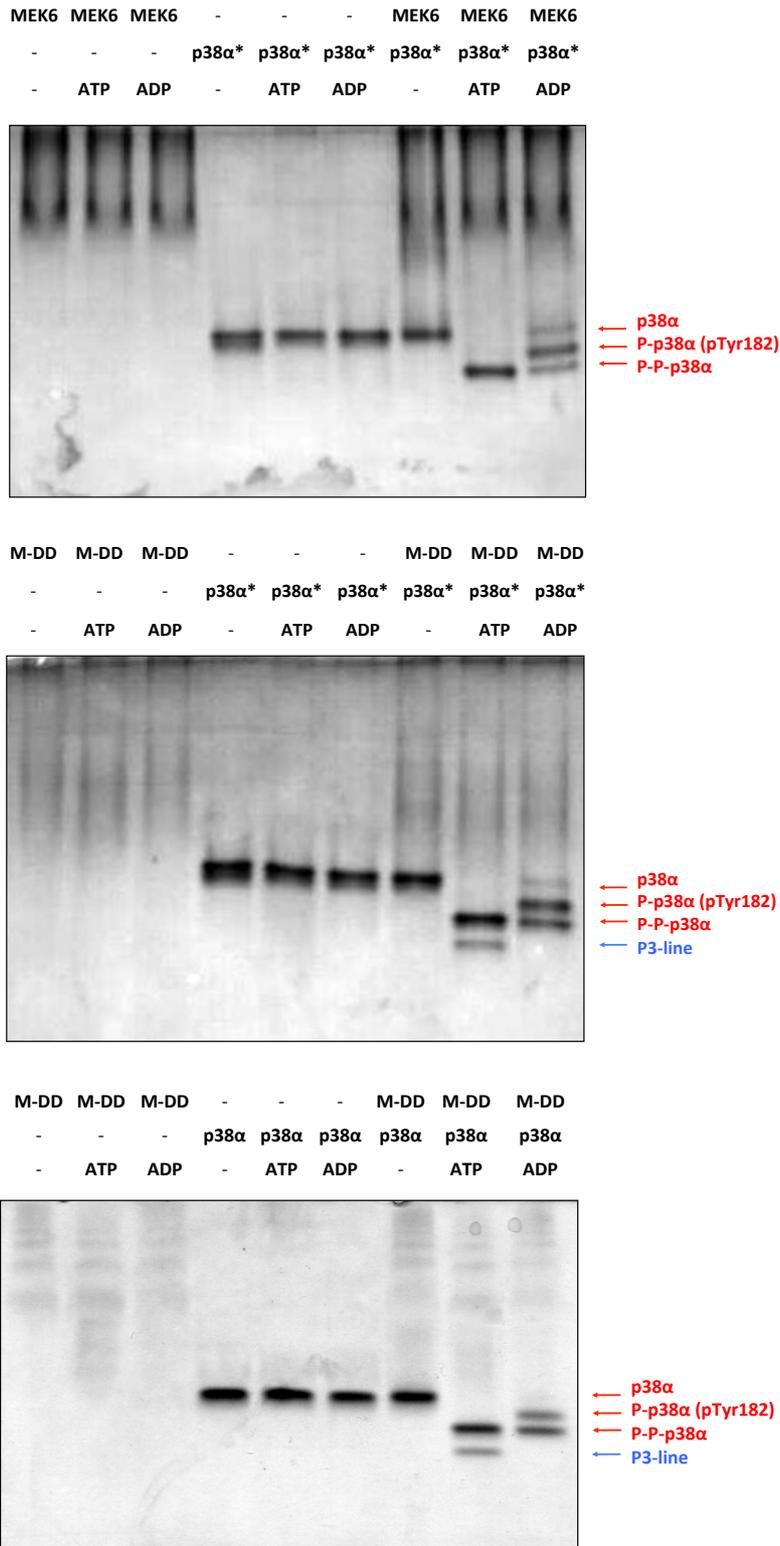


Figure 7.4 Native PAGE gels of MEK6-p38 α K53R, MEK6DD-p38 α K53R, MEK6DD-p38 α in the presence of ATP or ADP. Details are written in the experimental procedures section. The phosphorylation states are indicated by *red* arrows. The *blue* arrow indicates the extra band MEK6DD is named as M-DD. p38 α K53R is named as p38 α^* .

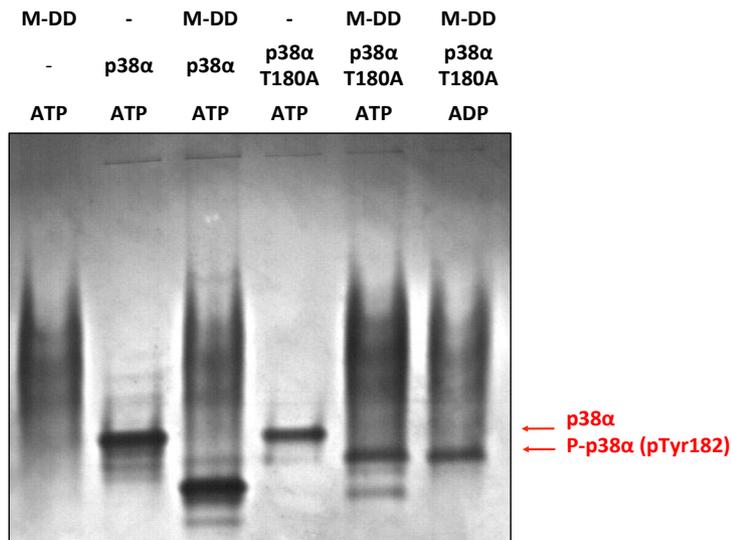


Figure 7.5 Native PAGE gels (7.5%) of MEK6DD-p38 α and MEK6DD-p38 α T180A. The phosphorylation states are indicated by *red* arrows. The phosphorylation legend is shown only for MEK6DD-p38 α T180A for clarity. MEK6DD is named as M-DD.

The native PAGE gels of figures 7.4-7.5 show similar band shifts for both MEK6 and MEK6DD native PAGE gels. This suggests that MEK6 is in an active form. Phosphorylation of MEK6 is carried out *in vivo*, by a specific MAPKKK. However, it is possible that MEK6 is phosphorylated during bacterial expression. To test this hypothesis, mass spectrometry analysis will be performed.

When the MEK6DD construct is used, an extra third phosphorylation band is present (P3) in the native PAGE gels. Mass spectrometry results showed that this band is not linked to additional, non-specific phosphorylation. Two possible explanations are proposed. The first is that MEK6DD is able to phosphorylate p38 α on other sites. The second, that the phosphorylation profile is also affected by the presence of MEK6DD, which means that while the two main bands are linked to a phosphorylation state in which p38 α remains bound to MKK6DD, the band with highest mobility shift is linked to complex dissociation. This phenomenon occurs because the samples are not at chemical equilibrium during the electrophoresis step. Indeed, mass spectrometry analysis revealed that MKK6DD is present in the p38 bands, with a different ratio between the bands. Thus, the difference between the two lines is due to the complex's electrophoretic properties.

Despite this extra band, MEK6DD demonstrates similar behaviour to WT MEK6. Since MEK6DD has higher expression levels than MEK6, the experiments of the next sections were performed using MEK6DD.

7.4.2.2 Analysis of the native PAGE gels by western blots

As a complementary experiment to the mass spectrometry analysis, western blots (WB) were performed, using an antibody specific for the mono-phosphorylated or bis-phosphorylated p38 α . For these experiments, several phosphate donors or phosphate inhibitors were used (ATP, AMP-PNP, AMP-PCP, ADP, ApCp and AMP).

The results obtained are in agreement with the mass spectrometry data (Figure 7.6). This is true also for the P3 line, which corresponds to the double phosphorylated p38 α . The reaction is blocked only when AMP-PCP, non hydrolysable ADP (ApCp) or AMP are used (Figure 7.6). Surprisingly, the reaction was not blocked by AMP-PNP. This is probably due to the fact that the P-N-P bond is relatively liable to nucleophilic attack (Olesen *et al.*, 2007). This information will be useful with the prospective of crystallising the ground state or transition state of the complex MEK6-p38 α , using a nucleotide analogue able to block the phosphorylation reaction.

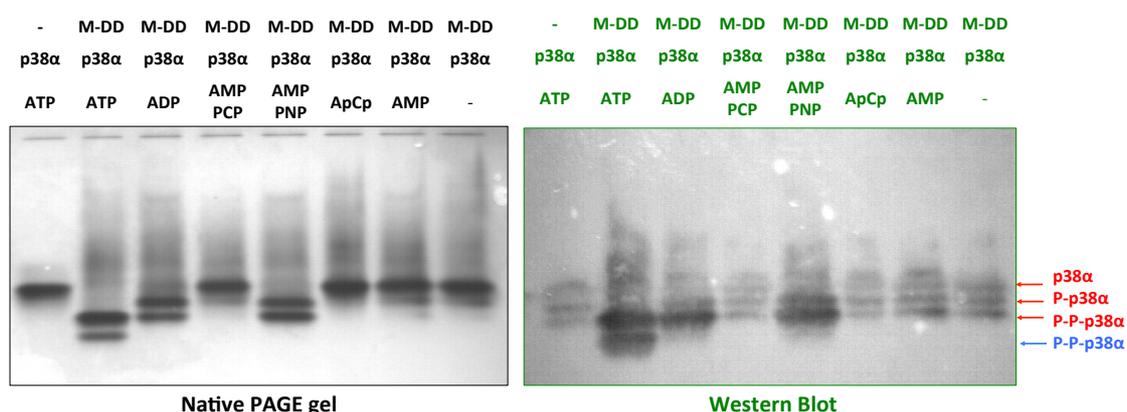


Figure 7.6 Western blot results. MEK6DD and p38 α were mixed with different nucleotides. Phosphorylation states are indicated by *red* and *blue* arrows. MEK6DD is named as M-DD.

7.4.2.3 Alternative explanations to the non-canonical phosphorylation pathway

From these first experiments, MEK6DD appeared to be able to use ADP as a phosphate donor as well as ATP. Due to the novelty of this reaction, a thorough investigation into whether the data observed were caused by other factors was required. Extensive control experiments were performed to eliminate the possibility of extra ATP sources, different from those added, that could account for the effects observed.

As shown in figure 7.4 the reaction was run with p38 α K53R. The mutant obtained is unable to bind nucleotide and is therefore unable to autophosphorylate or phosphorylate MEK6, eliminating the possibility that p38 α itself is in some way causing the observed results. Moreover, this mutant cannot bind ATP, preventing p38 α to be a ATP source for MEK6. The native page gel in figure 7.4, shows no difference between the use of p38 α K53R or p38 α , which excludes the possibility that the phosphorylation observed is dependent on p38 α .

As a second hypothesis, it was proposed that ATP could be generated from a contaminant, such as adenylate kinase. Adenylate kinase can produce ATP and AMP from ADP ($2\text{ADP} \rightleftharpoons \text{ATP} + \text{AMP}$) (Hanks *et al.*, 1988). This enzyme can be efficiently inhibited by P¹-(5'-Adenosyl)P⁵-(5'adenosyl) pentaphosphate (Ap₅A), a nucleotide analogue that acts as competitive antagonist ($K_i = 2 \pm 0.5 \times 10^{-6}$ mM). Thus, MEK6DD and p38 α were mixed as usual and Ap₅A was added in μM -mM concentrations. As the results in figure 7.7 show, Ap₅A does not affect the reaction.

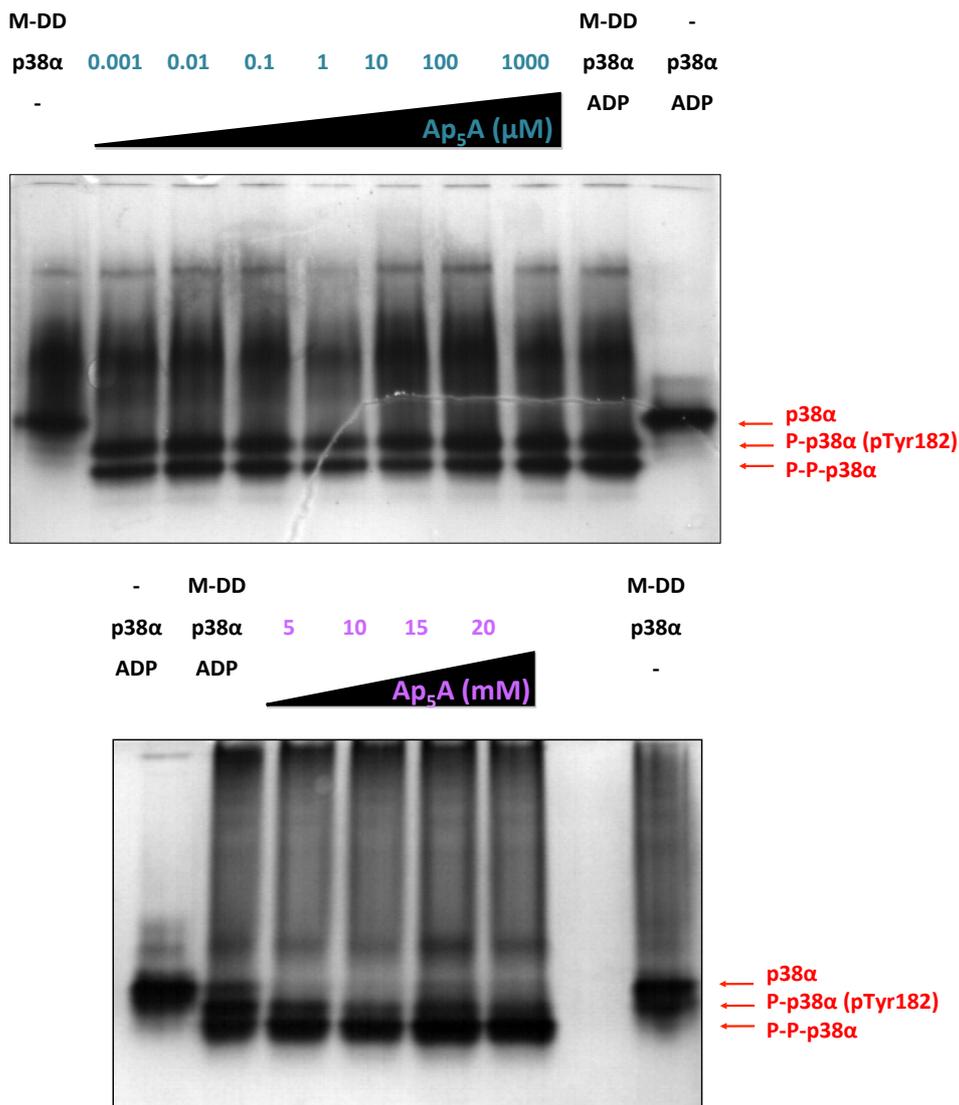


Figure 7.7 Effects of μM or mM amount of Ap_5A on $\text{p38}\alpha$ phosphorylation in presence of ADP. As controls the following samples were added: $\text{p38}\alpha + \text{MEK6DD}$, $\text{p38}\alpha + \text{ADP}$, $\text{p38}\alpha + \text{MEK6DD} + \text{ADP}$. The red arrows indicates the mono- and bis-phosphorylated $\text{p38}\alpha$. ADP is at a concentration of 5 mM in all samples.

7.4.5 The Functional role of MEK6-ADP phosphorylation

The use of ADP in phosphorylation reactions is quite unusual in biology. In this final section a sequence of preliminary experiments are described that try to investigate the functional meaning of the ADP phosphorylation.

The data obtained so far showed that the phosphorylation reaction is more efficient in the presence of ATP than with ADP. Native PAGE gel results show that only when

ATP is used, could phosphorylation of all p38 α in the sample be achieved. Thus, ADP can be an alternative phosphate donor when ATP is absent or the ATP- β P can be used as a second phosphate source after ATP- γ P. The goal of these preliminary experiments was to check the relevance of γ P and β P in the MEK6 reaction. Therefore, the reaction was analysed *in vitro* using three different nucleotides: ATP, ADP and ApCpp, in order to check the reaction efficiency when respectively γ P- β P, β P and γ P are present in solution. In the case where γ P is only used, a similar efficiency is expected in the presence of ATP or ApCpp. In the case when γ P- β P are both used, ADP and ApCpp should show a similar phosphorylation efficiency.

The kinetic experiments were carried out as described in the methods section. MEK6DD and p38 α were used as constructs. The results obtained, and the quantification analyses performed, suggest that ATP γ P and β P are both relevant for the MEK6 phosphorylation reaction. The same experiment was repeated five times and as the endpoint kinetics show different results linked to the nature of the nucleotide, these native PAGE gels were quantified. However, the kinetic experiments provide information about the reaction rate. MEK6 is more efficient in the presence of ATP. The amount of the bis-phosphorylated form of p38 α in the presence of ADP or ApCpp seems to be small and not relevant at physiological nucleotide concentrations (100 μ M and 1 mM). Interestingly, in the presence of both nucleotides (ADP and ApCpp) a constant p38 α amount remains un-phosphorylated (Figure 7.8).

The endpoint kinetic experimental results were analysed and quantified. The ratio between the amount of bis-phosphorylated p38 α in the presence of different nucleotides was calculated. The values indicate that MEK6 is more efficient in the presence of ATP and that there are no differences when ADP or ApCpp are used at physiological nucleotide concentrations (Figure 7.9).

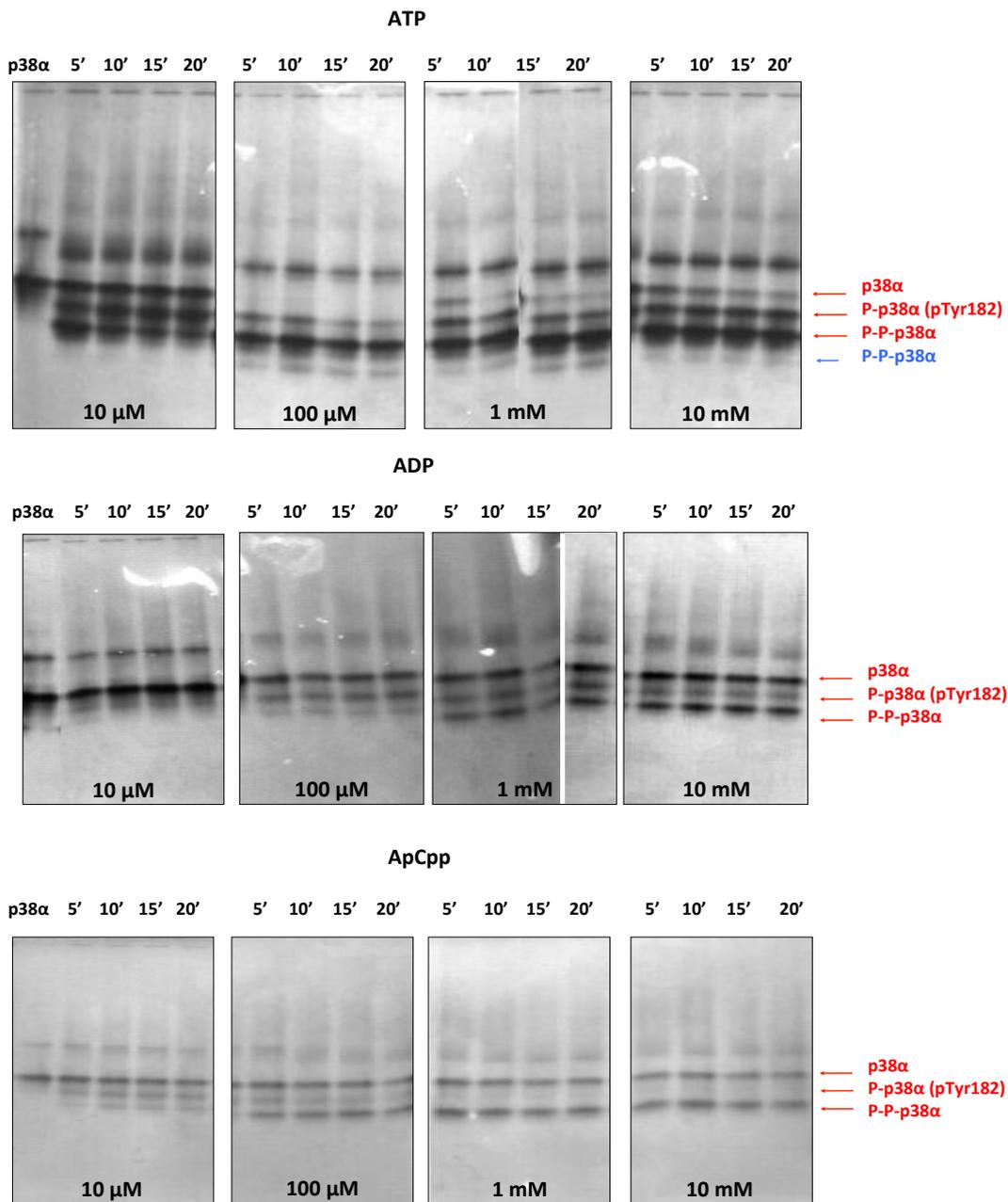


Figure 7.8 Kinetic experiments in the presence of ATP, ADP or ApCpp. As expected, p38α is phosphorylated over the time course. Complete phosphorylation is not achieved in twenty minutes in any condition. The results could be dependent by an error in the experimental set up. Additional bands with lower electrophoretic mobility are also present. These bands correspond to aggregated p38α sample.

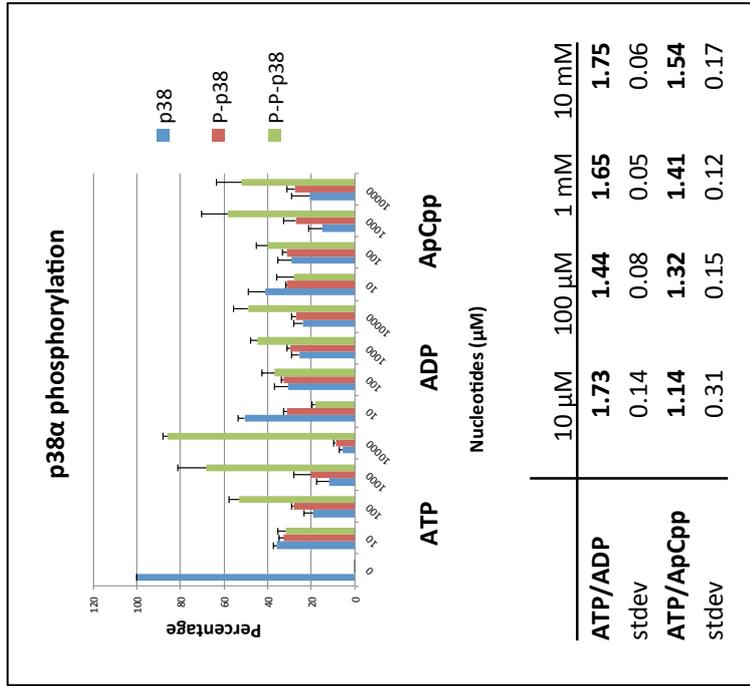
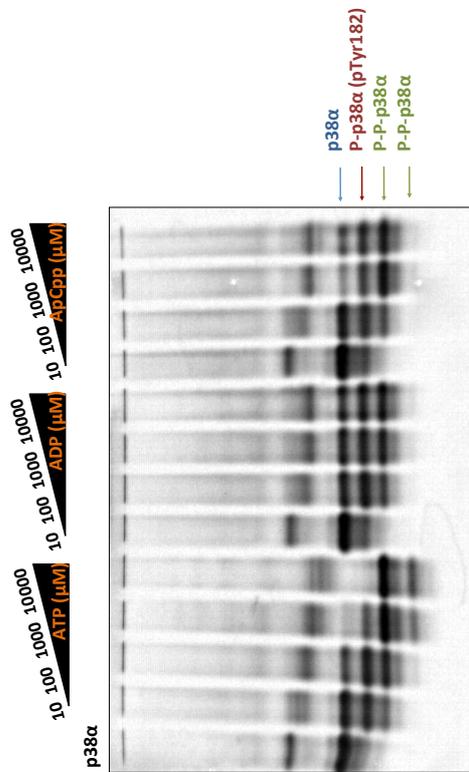
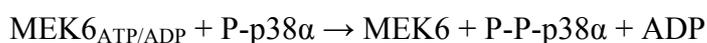


Figure 7.9 - Table 7.4 Endpoint kinetics in presence of ATP, ADP or ApCpP. The Native PAGE gel shows the endpoint of the kinetic assay. This experiment has been repeated 5 times for building up the histogram. The histogram indicates the percentage of mono- or bis-phosphorylated p38α for each concentration of nucleotide used. The phosphorylation state indicate in *blue* is countered inside the double phosphorylated p38α population. The table indicates the efficiency of MEK6DD in bis-phosphorylating p38α in presence of ATP regard to ADP and ApCpP, calculated as ratio between the percentage values.

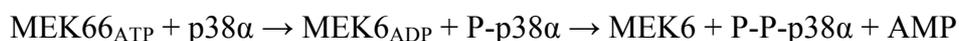
Thus, the data indicate that for MEK6DD activity both the γ P and β P of ATP are relevant.

As both γ P and β P are playing a role, further experiments were proposed to test how each phosphate is involved. It can be proposed that the reaction can be carried out in a canonical way in which the nucleotide is rebound after p38 α is released, or in a non-canonical way in which one ATP molecule is used to carry out both phosphorylation reactions:

- Canonical reaction (nucleotide rebinding):



- Non-canonical reaction (one nucleotide binding):



In the first mechanism either ATP or ADP could be bound by MEK6, in the second it is used exclusively as a phosphate source for Thr180 phosphorylation.

In order to investigate which reaction is occurring, kinetic experiments were repeated mixing ADP and ApCpp in a 1:1 ratio. In this way, in the solution both β P and γ P are present and if the first reaction is correct it would be expected to see a similar pattern to the ATP efficient reaction. On examining the endpoint kinetic results, the efficiency obtained in the presence of both nucleotides is similar to when only ADP or ApCpp are used. These data suggest that the second mechanism (non-canonical reaction) is occurring. Indeed, these results suggest that the remaining nucleotide from the reaction performed with the ADP-ApCpp mix (AMP or ApCp) remains trapped in the complex, instead of being exchanged on a fast timescale as should be expected in a canonical phosphorylation reaction (Figure 7.10).

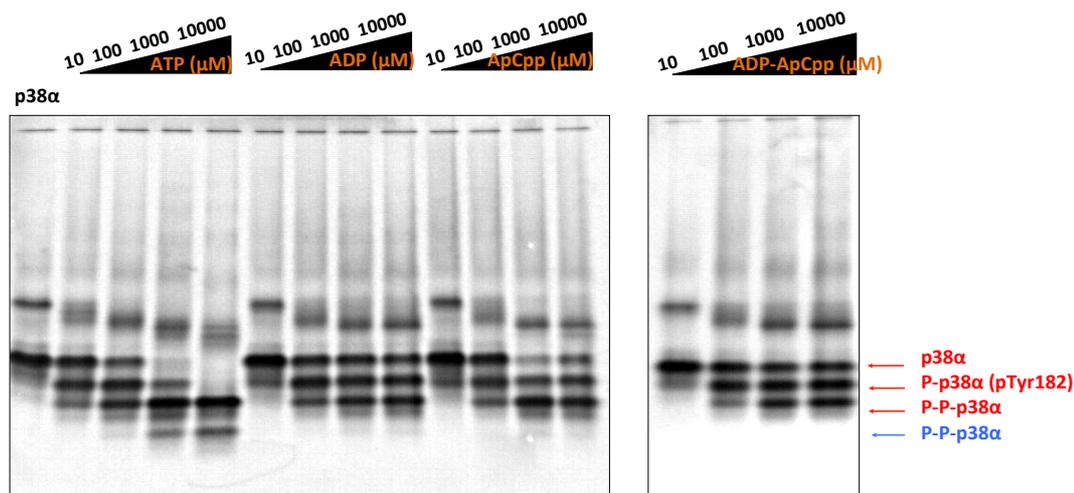


Figure 7.10 Endpoint kinetics for ADP-ApCp. The endpoint kinetic experiment for ADP-ApCp is shown in comparison with experiment of ATP/ADP/ApCp.

While intriguing, these results are not conclusive. MEK6 could have a different affinity for each nucleotide. In particular, ApCp affinity is usually lower than ATP affinity and in this case could affect the results obtained. Moreover differences in affinity for ApCp and ADP could affect the interpretation of the results. Indeed, it will be essential to repeat the same experiments with radioactive nucleotides such as ATP- ^{32}P and ATP- ^{32}P in order to confirm the transfer of the β -phosphate.

7.4.6 Preliminary crystallographic results on Δ MEK6DD

The possibility that MEK6 can use both ATP and ADP to phosphorylate its substrate raises the question of how the protein recognises the nucleotide in the active site in order to perform a second phosphorylation reaction. A proper answer to this question requires the structural characterisation of MEK6 in complex with both nucleotides. For this purpose, the truncated form of MEK6 (Δ MEK6DD) was used. Δ MEK6DD lacks the first 40 residues, which are the ones employed in the D-motif docking interaction and presumably are unstructured without substrate protein.

The structure of a truncated form of MEK6DD has been solved by The Structural Genomics Consortium, Oxford (PDB code 3FME). In this structure, the protein is in complex with the kinase inhibitor staurosporin. From the overlay with other MAPKK

structures it is likely that the staurosporin binding site is the same as the nucleotide-binding site. However, the interaction does not provide any information about nucleotide binding.

Further investigations can be made using the structures of other human MAPKK similar to MEK6. MEK6 has high sequence similarity with MEK3 and MEK4 (80% and 50% identity respectively). While there are no MEK3 structures, the un-phosphorylated structure of MEK4 in complex with ANP has been solved (PDB code 3ALN; Matsumoto *et al.*, 2010). Four molecules are present in the asymmetric unit and the nucleotide is bound differently in each one. Furthermore, it seems that only Lys137 is involved in sugar recognition in two of the molecules, while in the others the ribose is not co-ordinated by a lysine residue. If these observations are not affected by the fact that MEK4 is in an active form, it can be proposed that the MEK4 nucleotide binding site cannot specifically recognise the nucleotide selected, because there is no specific interaction with the sugar moiety or the phosphates (Figure 7.11) rather interactions primarily recognise the hydrophobic adenosine.

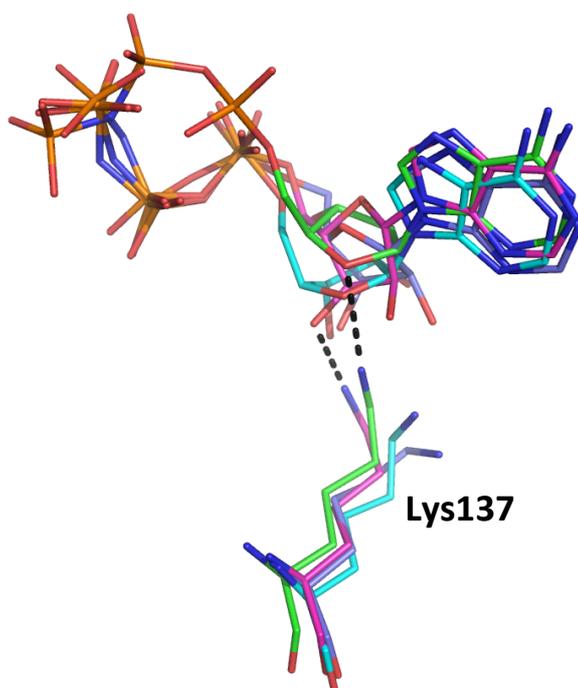
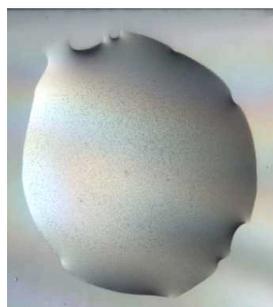


Figure 7.11 Overlay of the four ANP molecules in the MEK4 structure. Each colour corresponds to one molecule in the asymmetric unit. Protein interactions with the sugar moiety are shown. Hydrogen

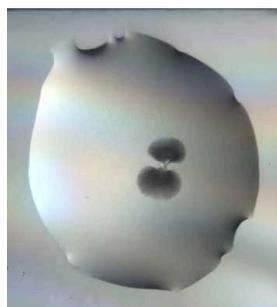
bonds are represented as dashed *black* lines. Lys137 is shown in sticks (PDB code 3ALN; Matsumoto *et al.*, 2010).

Therefore, in order to investigate nucleotide binding/recognition, the truncated form of MEK6DD in the presence of ADP or AMP-PCP was crystallised. The cloning was performed in Sheffield, purification and crystallisation was performed on site. Preliminary crystals were obtained at 4°C, at the HTX laboratory (Table 7.4). Optimisation is now on going.

Table 7.4 Preliminary crystallisation results of Δ MEK6DD:



After 1 week



After 2 weeks

Δ MEK6DD+AppCp

Position: Hampton_5, H10

Screen: Mme 5000

Condition: 0.1M Hepes pH7, 20% PEG MME 5000

Comments: the needles are visible since the second week



After 1 day



After 3 days

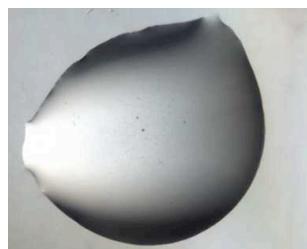
Δ MEK6DD

Position: Hampton_2, C08

Screen: Crystal Screen PEG-Ion, 11

Condition: 0.2 M potassium iodide pH 6.8, 20 %w/vPEG 3350

Comments: the needles are visible since the 3th day



After 3 days



After 1 week

Δ MEK6DD +ADP

Position: Hampton_2, C10

Screen: Crystal Screen PEG-Ion, 27

Condition: 0.2 M sodium acetate trihydrate pH 7.9, 20 % PEG 3350

Comments: the needles are visible since the 1th week

7.5 Conclusions

A novel finding has been described in this chapter: the ability of MEK6 to use ADP as phosphate donor. Several preliminary experiments, which comprise a combination of different techniques (native PAGE gels, western blots and Mass spectrometry), led to this finding. Importantly, relevant controls have been performed, which have allowed the conclusion that the phosphorylation reaction is not dependent on possible p38 α auto-phosphorylation (p38K53R construct) or adenylate kinase contamination. Structural studies and radioactive experiments will hopefully provide a much more convincing proof of the reaction described.

Through a series of kinetic experiments, the physiological meaning of the new finding was investigated. A new proposal was then formulated: MEK6 uses one ATP molecule to catalyse the double phosphorylation of p38 α . After the tyrosine phosphorylation reaction, ADP remains in the active site of MEK6 and is used for the threonine phosphorylation. The ability of a kinase to use ADP as phosphate donor is not novel. So far ADP-dependent phosphorylation in kinases has been described for glucokinase and phosphofructokinase, which are archaeal enzymes (Guixe & Merino, 2009). However, these kinases cannot use ATP, which makes the MEK6 non-canonical phosphorylation different, in that both ADP and ATP can be used.

The ability of MEK6 to use ATP β -phosphate for the second reaction, or in general the ability of using ADP as phosphate donor is a potential advantage:

- *in a processive or quasi-processive mechanism*; in both mechanisms nucleotide exchange will be difficult, because protein dissociation is absent (processive mechanism) or very quick (quasi-processive mechanism). Moreover, it is known that nucleotide exchange can slow down the reaction (Aoki *et al.*, 2011), which implies that an ADP molecule trapped in the MEK6 active site after the first reaction will improve reaction efficiency;
- *in the accommodation of phosphate groups during catalysis*; after Tyr182 has been phosphorylated, the P-Lip remains disordered. This is due to the docking interaction between MEK6 and p38 α . In fact, it has been demonstrated that also when p38 α is double phosphorylated, but the N-terminal tail of MAPKK is bound, the P-Lip remains disordered (Akella *et al.*, 2010). This implies that

during the second reaction the MEK6 active site has to accommodate several phosphate groups, which are P-Tyr82 and the nucleotide phosphate groups. The advantage of using ADP, instead of a second molecule of ATP, is that the number of phosphate groups will be the same in both phosphorylation reactions and a similar way of accommodating the negative charges can be employed. Structures of MEK6 in complex with different nucleotides will be helpful in elucidating this point.

- *when energy saving is required*; the ability of MEK6 to use ADP as phosphate donor could be relevant in situations such as ischemia/hypoxia, where ATP saving is critical. This has been suggested for a homologue of ADP-dependent glucokinase that has been found in *mus musculus* (Ronimus & Morgan, 2004). Since *in vivo* the entire MAPK cascade is supposed to be bound to a unique scaffold protein, MAPK can use the ADP produced by the other reactions for activating p38 α .

Thus, the non-canonical MEK6 phosphorylation pathway is a really novel proposal, which, in the case that it will be confirmed, will lead to the discovery of a new mechanism for protein kinases to perform double-phosphorylation reactions. Structural studies will be important for investigating how the two transition states are accommodated and which are the main elements in catalysis. It is probable that MEK6 uses the so called “minimal structural requirements”, that all protein kinases use to catalyse phosphorylation (Oruganty & Kannan, 2012).

These catalytic requirements are:

- the Glycine-rich loop;
- a lysine/arginine in beta sheet 3 (sub-domain II) that binds ATP (as Lys56 in the double phosphorylated structure of p38 γ , see Chapter6);
- a glutamate in C-helix (sub-domain III) that coordinates with the beta sheet 3 lysine/arginine, (as Glu74 in the double phosphorylated structure of p38 γ , see Chapter 6);
- two aspartates in the active site which belong to the HRD and the DFG motif (one is catalytic base, the other coordinates one of the two ATP magnesium ions).

All these elements are present in MEK6, except that the glutamate of the C-helix is substituted by an aspartate, a residue with similar charge. Thus, it is probable that MEK6 use the usual kinase machinery for catalysing the non-canonical phosphorylation reaction. This suggests the possibility that other kinases could work as MEK6. However, in the case of MEK6 there are particular determinants that promote the non-canonical phosphorylation: the necessity of a double phosphorylation, the unstructured P-Lip and the fast rate of dissociation occurring *in vivo* for the complex. Possibly, similar systems will work in a similar way, such as MEK3, the other p38 activator, or JNK's MAPKK which also has been proposed to induce disorder in the P-Lip upon the docking interaction (Garai *et al.*, 2012; Heo *et al.*, 2004).

In conclusion, as the evolution theory defines, the MEK6 non-canonical phosphorylation pathway occurs not because "it can", but because it is a great advantage in the system to which it belongs.

Chapter 8 Perspectives

In this thesis a combination of ^{19}F -NMR, X-ray crystallography and biochemical studies has been used to investigate enzyme catalysis, specifically in three different systems:

- the bacterial mutase, β -Phosphoglucomutase from *L. lactis* (β PGM);
- the human small G protein RhoA, with its GTPase activating protein (RhoGAP);
- the human MAPK p38 α , with one of its MAPKK activators (MEK6).

The use of aluminium and magnesium fluoride as analogues of the “in flight” phosphorus of the transition state has furthered understanding of each system. The exception is the work on the p38 α -MEK6 complex, where a TSA structure was not obtained. However, the experimental attempts for trapping the transition state has led to other relevant findings.

Eventually for each subject novel relevant aspects of enzyme catalysis from substrate recognition to transition state formation have been defined.

In this final chapter, the general issues that have been addressed in this thesis will be discussed, specifically:

- How the study of transition state complexes, which comprise fluoride analogues, can benefit from the combination of ^{19}F -NMR and X-ray crystallography?
- Is the charge balance hypothesis universal?
- How substrate recognition is involved in catalysis?

8.1 How can the study of transition state complexes, which comprise fluoride analogues, benefit from the combination of ^{19}F -NMR and X-ray crystallography?

The combination of ^{19}F -NMR and X-ray crystallography was crucial for the characterisation of the major and minor conformers in the $\beta\text{PGM-G6P-MgF}_3^-$ -TSA complex, the selection of the proper phosphonate analogue of βG1P and the investigation of charge balance in RhoA-GDP-AlF_3 and $\text{RhoA-GDP-AlF}_4^-/\text{MgF}_3^-$ - $\text{RhoGAP}_{\text{R85A}}$ transition states.

The characterisation of the major/minor conformer in $\beta\text{PGM-G6P-MgF}_3^-$ -TSA, has shown how different conformers can exist through the combination of experiments in solution and the solid state. The successful combination of room temperature data collection and ^{19}F -NMR overcame restrictions at 100 K, where only one single model could be defined. Further efforts that go in the same direction can be found in the scientific community, such as the software tool Ringer, which systematically samples electron density around the dihedral angles of protein side chains looking for alternative conformations (Lang *et al.*, 2010). Magnesium fluoride was described for the first time in octahedral geometry rather than the classical TBP, opening the intriguing question of whether phosphorus could be similarly coordinated in the active site or if the rearrangements observed, such as the Ser114 backbone flipping, could be relevant in catalysis. For example, both phenomena could be involved in the releasing or binding of βG16BP . Indeed, the molecular mechanism by which βPGM releases and binds βG16BP in a different orientation is still under investigation. The structural changes that cause the observation of the major and minor conformers could be an explanation where the affinity of the active site changes.

The combination of ^{19}F -NMR and X-ray crystallography was also relevant in the selection of phosphonate analogues of βG1P . In this project, the most modern guidelines for designing and evaluation of a possible lead compound in academia environment were followed. First, the drug design was based on the structure of the natural enzyme substrate. This led to the design of phosphonate or fluorophosphonate compounds, which usually have the same or higher affinity as the natural substrate. Second, the most recent guidelines for the synthesis of phosphonates and fluorophosphonates compounds were followed. Third, ^{19}F -NMR was used as a

screening tool for testing the potential of fluorine phosphonate and phosphonate candidates to form a trifluoromagnesate or a tetrafluoroaluminate complex. This use of ^{19}F -NMR has been already suggested (Vulpetti & Dalvit, 2012). Fourth, crystal structures were fundamental for interpreting the ^{19}F -NMR and biochemical data. Indeed, the structures provided explanations for the inability of 1- α -hydroxyl phosphonates and the preference for the *S*-isomer to form $\beta\text{PGM-TSA}$ complexes, which led to the conclusion that attempts to produce further fluorophosphonate molecules, such as $\beta\text{CF}_2\text{G1P}$, would be useless. Importantly, from this work a possible new lead compound for combating bacterial infections (*S*)- βCHFG1P , has been proposed.

^{19}F -NMR and X-ray crystallography were also used for investigating how charge is balanced and catalysis is achieved in RhoA-GDP-AlF_3 and $\text{RhoA-GDP-AlF}_4^-/\text{MgF}_3^-$ - $\text{RhoGAP}_{\text{R85A}}$ complexes. Two main conclusions can be highlighted in this discussion. First, the solid and solution states can be different. Indeed, the crystallisation of RhoA-GDP-AlF_3 failed due to maintenance of a non-binding conformation in the crystal lattice, while ^{19}F -NMR clearly demonstrated the formation of the complex in solution. Secondly, both crystal structures and ^{19}F -NMR spectra of the $\text{RhoA-GDP-RhoGAP}_{\text{R85A-TSA}}$ complexes revealed that AlF_4^- or MgF_3^- were present in the mutant metallofluoride complex rather than the expected AlF_3 and MgF_2 . This is relevant because the interpretation of the moiety of transition state species in the electron density map can be ambiguous, in particular at resolutions lower than 2-2.5 Å. This implies that the recording of ^{19}F -NMR spectra is good practice for unambiguously assigning the TSA species.

8.2 Is the charge balance hypothesis universal?

In the general introduction, the important consideration that charge balance, rather than the native geometry, for phosphoryl transfer enzymes is essential for TS stabilisation during catalysis was discussed. In this thesis, the study in solution and solid state of $\beta\text{PGM}_{\text{K145A}}\text{-G6P-MgF}_2\text{-TSA}$, RhoA-GDP-AlF_3 and $\text{RhoA-GDP-AlF}_4^-/\text{MgF}_3^-$ - $\text{RhoGAP}_{\text{R85A}}$ were primarily performed to investigate the charge balance hypothesis in a

mutase and in a small G protein system. However, it resulted that charge balance interpretation in these mutated systems was more complicated than expected.

Based on the charge balance hypothesis, the loss of a positive charge should bring the loss of a negative charge in the bound substrate in order to satisfy the zero charge required in the active site (Baxter *et al.*, 2008; Cliff *et al.*, 2010). In the case of PGK_{K219A}, it has been found that one of the fluorine ions of AlF_4^- was substituted by a water molecule, without any change in the protein structure or in the metal fluoride features (Cliff *et al.*, 2010). This occurs in β PGM_{K145A}-G6P-MgF₂-TSA too, where magnesium is coordinated by two fluorine atoms and one water molecule. However, a change in geometry was also observed, both in the magnesium fluoride geometry and the β PGM active site. In the case of the small G protein studied, both fluorine NMR and crystal structures confirm that the mutation R85A did not induce any change in the transition state analogue, but caused the substitution of the arginine finger with a tyrosine side chain.

The complexity of these systems is probably related to specific features of the proteins studied, which prevent prediction with a simple rule of the consequences of charge mutation in the active site. In the case of the PGK-TSA, ¹⁹F-NMR spectra collected on both TSA of the WT and mutant show a clearly interpretable peak pattern; moreover, the mutant PGK_{K219A} is a catalytically dead enzyme. The mutase and small G protein systems are less simple models: more than one conformer was present in the β PGM-TSA spectra and the mutation performed in β PGM promotes one of them; in the case of RhoA and RhoAGAP_{R85A} an intrinsic activity and residual activity is still present. Thus, how charge balance is achieved in these systems is still under investigation.

Finally, it is an intriguing question how charge is balanced in MEK6 double phosphorylation of p38 α . In this thesis, it has been proposed that MEK6 use one ATP molecule to catalyse both reactions. As discussed in section 7.5, this implies that the number of phosphorus groups inside the active site will be the same for both reactions. This scenario is a great advantage in term of charge balance, because the charge of the phosphorus groups could be accommodated in a similar way for both reactions. A different charge balance accommodation is indeed expected if for the second reaction where a second ATP molecule will be used, and, thus a fourth phosphorus would be

present in the active site. Indeed, charge balance is a further argument for the MEK6 non-canonical phosphorylation reaction here proposed.

8.3 How is substrate recognition involved in catalysis?

Enzymes lower the activation energy through binding energy, which implies that the molecular characterisation of enzyme substrate recognition at both ground and transition state is a relevant goal in chemical biology, which leads to the comprehension of how an enzyme mechanism or enzyme region could become a possible pharmaceutical target.

Substrate recognition has been studied in all the three systems described in this thesis. Firstly, it has been characterised for the first time how β PGM distinguishes between β G1P and G6P through the comparison of high resolution structures (1.1-1.3 Å). It was found that β PGM has probably evolved to recognise β G1P, which is the sugar that defines α/β anomer specificity.

Secondly, the study on RhoA-GDP-AlF₄⁻/MgF₃⁻-RhoGAP_{R85A} provides a structural proof of the RhoGAP binding contribution in enhancing the RhoA GTP hydrolysis. RhoGAP binding promotes the corrected positioning and polarisation of the attacking water molecule, and occludes it in the active site.

Thirdly, a structural characterisation of how MEK6 recognises p38 α from the docking interaction has been provided for the first time. A dynamic model, the “the scanning process”, which requires further investigation, has been proposed, where MEK6 uses electrostatic interactions to find the hydrophobic groove on the p38 α D-motif binding site. The electrostatic interactions are within two main elements of the D-motif binding site: the well known CD domain, and the R-pocket, which has been proposed in this thesis to be the third element of the D-motif binding site in MAPKs.

Eventually, several new findings have been discussed, which could have a remarkable impact from the comprehension of phosphoryl transfer enzyme catalysis to drug design on one of the system analysed. Indeed it is hoped that the results obtained in this thesis, will be the basis for new and exciting research projects.

The work done on β PGM has led to the finding of a possible lead compound. Experiments *in vivo* are now required to define where the compound could work as a

possible antibiotic, which is one of the most important goals of medical science in a world where bacteria start to be resistant to many antibiotics.

The work done on p38 α -MEK6 has led to several new proposals, among which the most novel is the MEK6 non-canonical phosphorylation pathway. The confirmation of this proposal is only the first step to a new field of investigation. The comprehension of how a protein that uses ATP as phosphate donor, could also use ADP, will indeed change the general way of thinking of the enzyme-ADP complex as the inactive complex or to the dissociation of enzyme-ADP complex as the reaction limiting step.

Thus, the hope is that all the results reported in this thesis will have a positive impact on the efforts that both academia and the pharmaceutical industry are putting into the understanding of how phosphoryl transfer enzymes work and in the search for new targets and lead compounds in cancer, inflammatory diseases and infectious diseases, where all the systems described in this thesis are involved.

Appendix A Crystallographic theory

A.1 Introduction

Crystals have fascinated people since prehistoric times. With the discovery of X-ray diffraction in 1912 by Max von Laue, the possibility of investigating the internal structure of inorganic and organic crystals became possible.

Today, protein crystal production has become one of the most relevant steps in biochemistry. Protein structures solved from X-ray diffraction provides information about protein folds and function. In this chapter, the principles of determining a protein structure from a diffraction experiment will be discussed. As main reference the book “*Principles of protein x-ray crystallography*”, by Drenth J will be followed (Drenth, 1999).

First, the hardware required for the experiment will be described.

A.2 Hardware

The main pieces of hardware required for the collection of the X-ray diffraction data are an X-ray source and an X-ray detector. Since in this thesis most of the diffraction experiments were performed at the ESRF, the next paragraphs will refer to synchrotron radiation and, more specifically, the ESRF.

A.2.1 Synchrotrons

Synchrotrons are devices for circulating electrically charged particles at nearly the speed of light. Electrons or protons are first injected into an accelerator device, such as a linear accelerator or a booster synchrotron. Then the particle beam enters a vacuum tube in the storage ring, with a diameter from 10 to a few hundred meters. The pseudo-circular trajectory is then determined by the particle’s energy and by the magnetic field. The magnetic field acts through bending magnets, that are used to guide the particles in their orbit and causes changes in the direction followed. At each direction change the

particles lose energy. This phenomenon produces X-ray electromagnetic radiation (wavelength of 10^{-7} - 10^{-11} m).

Synchrotrons are described by their energy of operation (E) and the magnetic strength of the bending magnets. The ESRF is an electron synchrotron that has a circumference of 8444.39 meters, is operated with energy of 6 GeV and has bending magnets with a 0.86 -T field strength. In the storage-ring the beam is guided onto the pseudo-circular orbit by 64 bending magnets and is focused by 320 quadrupole magnets. A further 224 sextupole magnets control the effect linked to the energy dispersion of the electrons (chromatic aberration). The storage-ring components are arranged in 16 super-period cells with the same magnet distribution, which is named the Chasman-Green structure (Lindley, 1999). Several devices, such as wavelength shifters and undulators are used in the super-period cell to increase locally synchrotron beam intensity. The wavelength shifter is similar to bending magnets, but with a sharper curvature and a stronger local magnetic field, which produces high-intensity X-rays and a shorter wavelength. Usually, a series of wavelength shifters or multiple wigglers are used, which is easily tuneable to the desired wavelength. Undulators are similar to the multiple wiggler, but with moderate magnetic fields and a large numbers of poles close together; moreover they produce radiation at more specific wavelength than wigglers (Figures A.1-A.2).

Synchrotrons have several advantages with regard to home X-ray sources. The X-ray diffraction intensity is much higher (at least two orders of magnitude stronger) and the beam is less divergent. The higher intensity is relevant, because crystals of macromolecules tend to have a low scattering power, due to a low number of molecules inside the crystal and the high percentage of low diffracting elements. A less divergent beam results in sharper and better separated diffraction spots.

Synchrotrons have also other relevant features, which are missing in the X-ray home source:

- it is a multiple-wavelength source (essential for MAD experiments);
- the tuneability can be regulated by selection of a specific spectral range with monochromators;
- the radiation is highly polarised, which can affect the anomalous X-ray scattering signal.

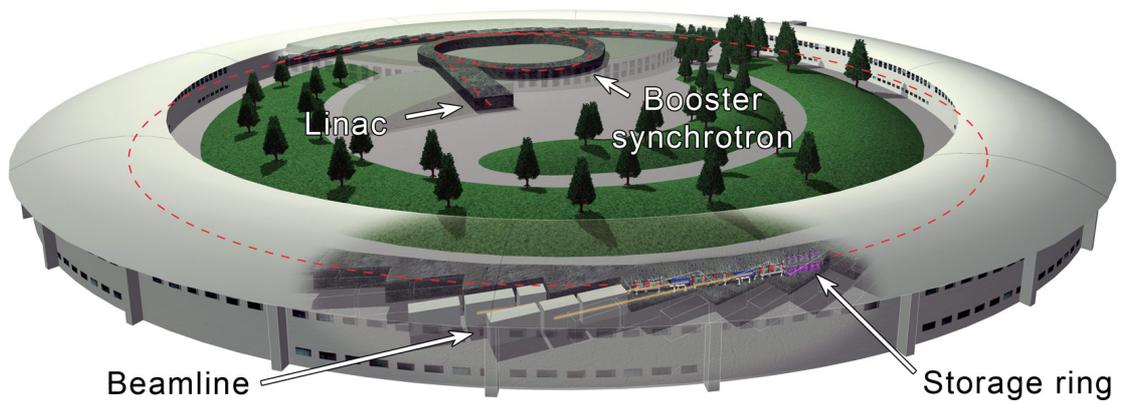


Figure A.1 The European Synchrotron Radiation facility (ESRF). The structure of the ESRF is shown with the main components. The injection system for the ESRF comprises a 200 MeV linear Pre-Injector (LINAC) and a full energy fast cycling Booster synchrotron. One of the 16 super-period cells is shown. Each one contains two undulators. The synchrotron X-ray beam leaves the main storage ring at a point a few metres after the undulator. Upon leaving the storage ring, the X-rays enter one of 41 beamlines, each an ensemble of laboratory blocks or “hutches” where the actual research takes place (www.esrf.eu).

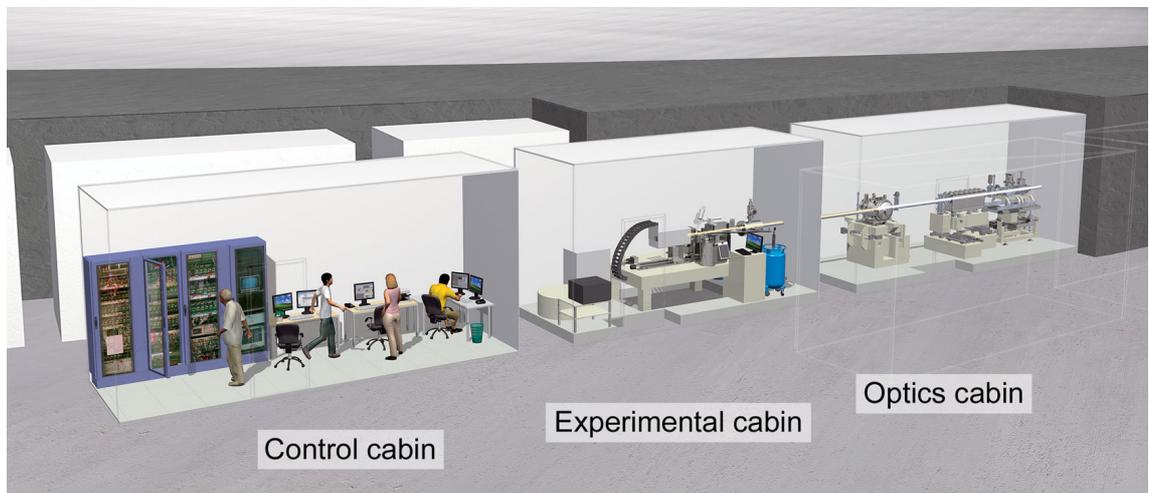


Figure A.2 The beamline experimental hutches. The schematic representation of an ESRF beamline shows that it is comprised of three main hatched or cabins: the **optic cabin**, which contains the mirror and monochromators which allow to focus and to tune the beam; the **experimental cabin**, which contains the hardware necessary for a specific experiment; the **control cabin**, from which the other hutches are controlled (www.esrf.eu).

A.2.2 Detectors

For a crystal structure determination, the intensities of all (or a great many) reflections must be measured. This requires at least two more pieces of equipment: a mechanical tool, that will rotate the crystal in the X-ray beam to allow diffraction from all the lattices points and centre the crystal to the beam (see paragraphs A.2-A.3) and a detecting device for measuring the position and the intensity of the diffraction spots.

At the ESRF, the following detectors were used: the CCD (charge coupled device detector) and PILATUS (Single photon counting pixel detector).

A.2.2.1 CDD detector

CCD detectors have extremely fast readout times, low background levels and a high maximum count rate (particularly useful with the intense beams produced by synchrotrons). A CCD is a two-dimensional surface that processes the signal immediately after photon detection.

The three main components of the CCD detector are a phosphor screen, a fiber-optic taper, and a CCD chip. X-ray photons interact with the phosphor screen. In the ADSC Quantum detectors this is gallium arsenide doped with europium ions. The screen converts X-rays photons to light. Then the fiber-optic taper demagnifies the light image down to the size of the CCD chip. The CCD chip detects the light image as an electric charge.

The CCD chip is made of silicon, which has a valence band and a conducting band available for the electrons. When subjected to photons the electrons acquire enough energy to go from the valence to the conducting band, leaving a hole behind. Electron-hole pairs are formed in a pixel, which is gated by a local voltage. Once the exposure is over, the gate electrodes are changed to release the charge, which is read for each pixel sequentially and produces a charge, associated with each pixel. This charge is amplified and converted to an image that, after geometric and intensity correction, represents the number of X-ray photons incident on the detector surface (Figure A.3).

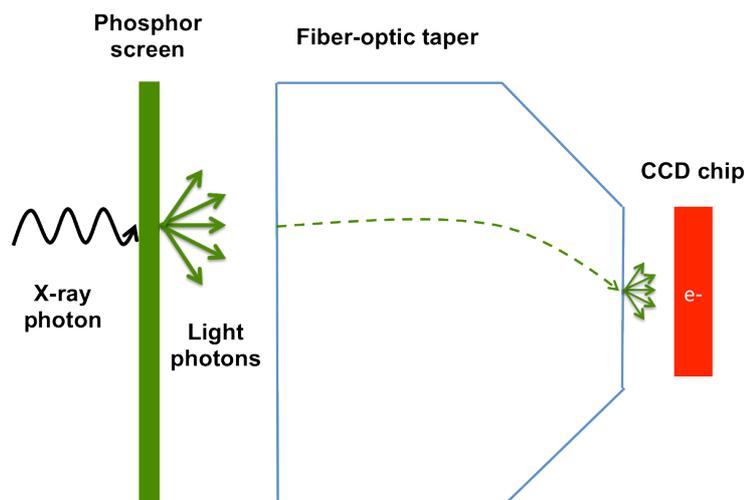


Figure A.3 CCD detector. The three main components of the CCD detector are shown as described in the main text.

A.2.2.2 PILATUS detector

The PILATUS detector is a recent type of X-ray detector, which has been developed at the Paul Scherrer Institut (PSI) for the Swiss Light Source (SLS).

PILATUS detectors are two-dimensional hybrid pixel array detectors, which operate in single-photon counting mode. It comprises a preamplifier, a comparator and a counter. The preamplifier enforces the charge generated in the sensor by the incoming X-ray; the comparator produces a digital signal if the incoming charge exceeds a predefined threshold and thus, together with the counter, one obtains a complete digital storage and read-out of the number of detected X-rays per pixel without any readout noise, which is an advantage with regard to CDD detectors (Figure A.4).

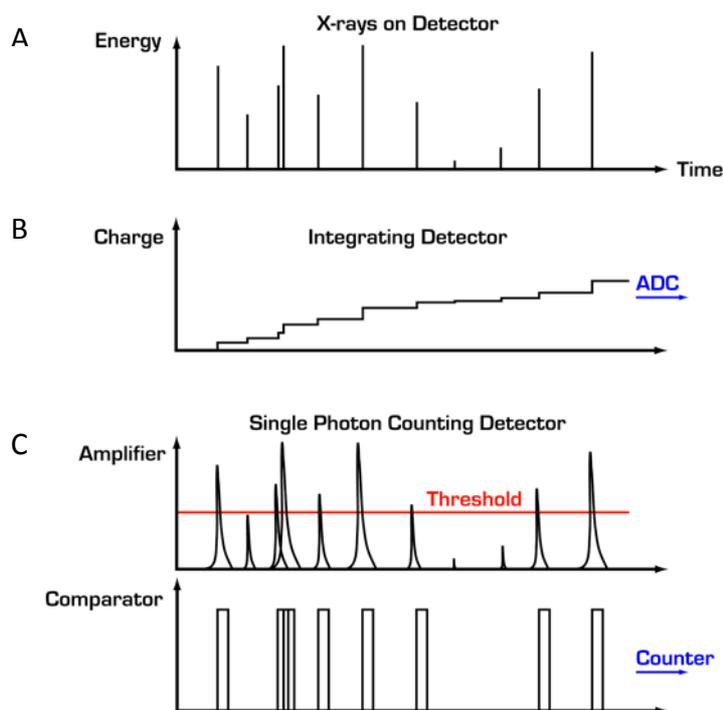


Figure A.4 The PILATUS detector detects X-rays without readout noise. A) X-rays of different energies are impinging on the detector. B) In the case of the integrating detectors such as CCD cameras, charge is accumulated and then converted. C) In the case of single photon counting an energy threshold is apply for each pixel and a comparator sends only those pulses to the counter which are higher than the energy threshold. Therefore, no readout noise is collected.

Further advantages include: superior signal-to-noise ratio, readout time of 5 ms (instead of 1-120 s), a dynamic range of 20 bit, high detective quantum efficiency and the possibility to suppress fluorescence by an energy threshold that is set individually for each pixel. Practically, the PILATUS is a much faster data collection system, which provides a better background subtraction. The short readout and fast framing time allow diffraction data to be collected in continuous mode without opening and closing the shutter for each frame.

A.3 Crystal properties

A crystal is a three-dimensional periodical arrangement of molecules. When materials precipitate from a solution, the molecules attempt to reach the lowest free-energy state, which can lead to packing in a regular way. This phenomenon often leads to a symmetric relationship between the molecules.

In crystal packing, three repeating vectors (a, b, c) with angles (α, β, γ) between them can be recognised. The three vectors define the **unit cell** in the **crystal lattice**. Thus, the crystal is a three dimensional repetition of the same unit cell. The edges of the unit cell form a grid, which is called crystal lattice. The crystal lattice is described by right hand coordinate system, in which the axis $x, y,$ and z have the direction of a, b, c respectively (Figure A.5).

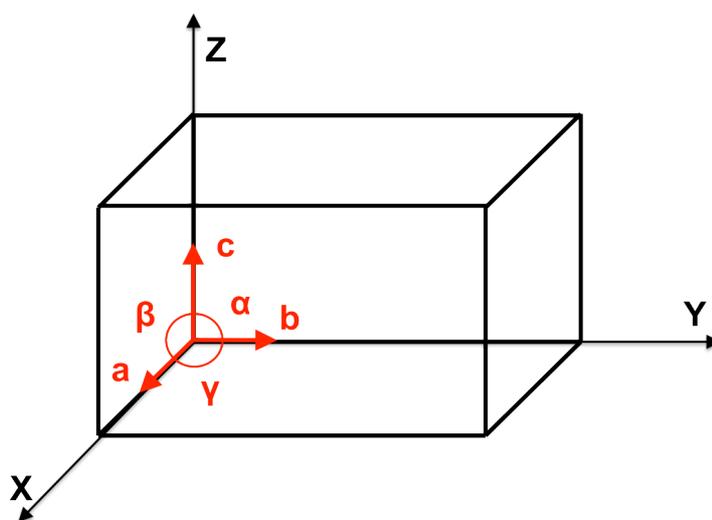


Figure A.5 A unit cell within a crystal lattice. Angles and vectors defining the unit cell are in *red*. The crystal lattice system is in *black*.

The repetition of the unit cell is called three-dimensional translational symmetry. Often, other symmetry operators can be found in the crystal. Protein crystals have rotation axes (2, 3, 4 and 6 fold axes) and rotation inversion axes which combine rotation with translation (screw axes). The combination of all the possible symmetry operators in a protein crystal leads to 32 possible point groups. The term point group refers to the fact

that all the symmetry operations pass through one point. The 32 point groups are assigned to a maximum of 7 crystallographic systems: Triclinic, Monoclinic, Orthorombic, Tetragonal, Trigonal, Hexagonal and Cubic (Table A.1).

Table A.1 The seven crystal systems

Crystal lattice	Symmetry requirement	Axes	Angles and length constrains
Triclinic	None	No constrains	None
Monoclinic	One 2-fold axes	b parallel to 2-fold axis, a and c perpendicular to the 2-fold axis	$\alpha = \gamma = 90^\circ$
Orthorombic	Three perpendicular 2-fold axes	a, b and c parallel to 2-fold axis	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	One 3-fold axes	c parallel to 3-fold axis, a and b perpendicular to the 3-fold axis	$\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$ $a = b$
Tetragonal	One 4-fold axis	c parallel to 4-fold axis, a and b perpendicular to the 4-fold axis	$\alpha = \beta = \gamma = 90^\circ$ $a = b$
Hexagonal	One 6-fold axis	c parallel to 6-fold axis, a and b perpendicular to the 6-fold axis	$\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$ $a = b$
Cubic	Four 3-fold axis	a, b and c related by 3-fold axis	$\alpha = \beta = \gamma = 90^\circ$ $a = b = c$

If the lattice has a level of symmetry higher than triclinic (no symmetry), each particle in the cell will be repeated a number of times as a consequence of the symmetry operator. For example, if the space group is $P2_12_12_1$, at least four equal particles are expected for the unit cell. In other words, the unit cell has four asymmetric units. The asymmetric unit is the smallest unit that can be rotated or translated to generate the unit cell, following the symmetry operators described by the space group. The number of molecules in the asymmetric unit can be higher than one. Since these molecules are not related by crystallographic symmetry, they can have different conformations. However,

they can be related by other symmetry operators, which is called non-crystallographic symmetry (NCS).

Most crystals cannot be considered ideal single crystals because the regular repetition of the unit cells is interrupted by lattice defects (Figure A.6). The diffraction pattern of such crystals can be regarded as the sum of the diffraction patterns originating from mosaic blocks with slightly different orientations, this is called mosaicity or the mosaic spread of the crystal. The mosaic spread for a good quality protein crystals is between 0.2-0.5°. Mosaicity can be increased during the cryocooling process and incorrect mosaicity estimation during data processing and reduction can decrease the quality of the statistical data.

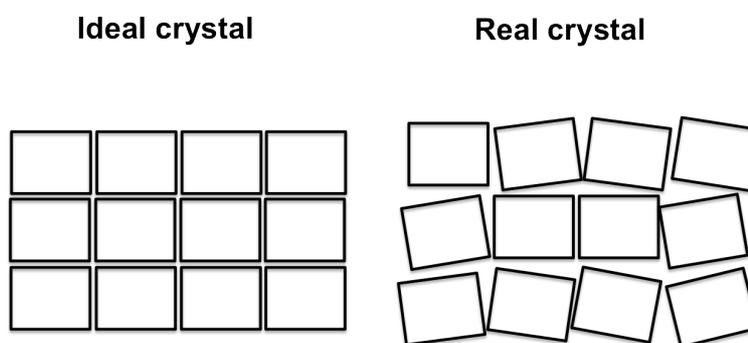


Figure A.6 Schematic of the mosaicity in the protein crystals. The image shows the difference between an ideal crystal and a real one, which can be thought of as a series of mosaic blocks with slightly different orientations.

A.4 Diffraction from protein crystals

The X-ray diffraction from a crystal derives from the structure of the molecule from which the crystal has grown. X-ray detectors collect this diffraction and a sequence of diffraction images is collected. The intensity of the spots depends on the electrons that scatter the X-rays, while the position of the spot arises from the positions of the molecules within the crystal lattice. Whether a spot will be observed or no is defined by Bragg's law and visualised in the Ewald sphere construction.

The diffraction of a crystal can be described as diffraction from its **lattice planes**, which are constructed through the lattice points. Parallel and equidistant planes, with

perpendicular distance d , form a set of planes. Each set is described by three indices, hkl , which are the Miller indices. Each index defines how many times the lattice plane cuts the axis a , b and c . Each set of planes reflects parallel X-rays in a way that the angle of incidence is equal to the angle of reflection and the beam is reflected by an angle of 2θ relative to each incident direction. Constructive interference will occur between reflected X-rays when the difference in the path length (the distance travelled by the X-rays), is an integral number of wavelengths. The difference in the distance travelled by reflected X-rays is $2d\sin\theta/\lambda$ and it can be applied to all reflected X-rays. This gives the Bragg's Law:

$$2d \sin \theta = n\lambda$$

where d is the distance between planes in the crystal, θ is the angle of incidence of the X-rays, n is an integer and λ is the wavelength of the X-rays (Figure A.7).

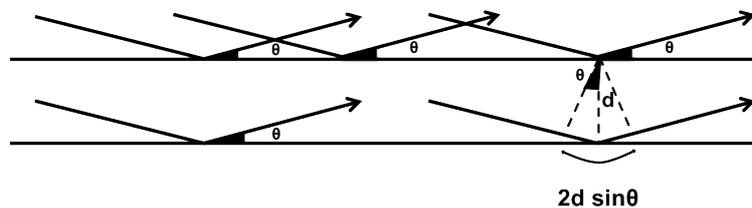


Figure A.7 The Bragg's law in the diffraction pattern. Two lattice planes are drawn separated by a distance d . Incident and reflected X-rays form an angle of θ with the lattice planes. The geometrical relationship between θ and d is shown.

The X-ray diffraction pattern can be explained through the visual construction of the Ewald sphere and how the reciprocal lattice interacts with it (Figure A.8). The Ewald sphere is a sphere whose radius is reciprocal to the X-ray wavelength ($1/\lambda$) and whose centre is the crystal position. The reciprocal lattice, is an imaginary lattice formed by planes, which are perpendicular to the set of planes in the crystal lattice. Thus, the unit cell dimensions are reciprocally related to the unit cell dimension of the crystal lattice and the crystallographic system is the same. Specifically, there is an inverse

relationship. In the Ewald sphere construction, diffraction only occurs when the diffraction spots of the reciprocal lattice planes are on the sphere. Spots from different planes can be brought to diffraction by rotating the crystal, which implies the rotation of the reciprocal lattice. Thus, the direction of the diffracted beam depends on the X-ray wavelength, which determines the sphere radius, and by the unit cell distance of the real space, from which the unit cell distances of the reciprocal lattice are derived.

Bragg's law and the Ewald construction are related each other. As described in figure A.8, S is the distance between the diffracting spot and the reciprocal lattice with origin O . Based on the Ewald construction is length is $2\sin\theta/\lambda$. When $2\sin\theta/\lambda$ is equal to $1/d$, the Bragg's law is satisfied.

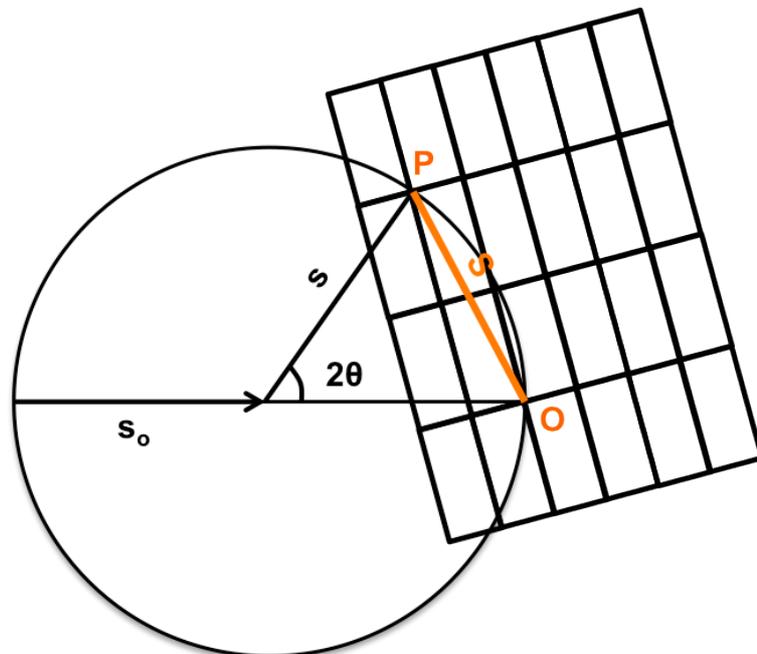


Figure A.8 A two-dimensional representation of the Ewald sphere. The sphere has a radius of $1/\lambda$; s_0 is the direction of the incident X-rays and s the direction of the scattered beam. The sphere has its origin at the crystal position. The origin of the reciprocal lattice is O , where the X-ray beam exits the Ewald sphere. The lattice point P is in contact with the sphere and will produce a reflection. OP is equal to S , whose amplitude is $s-s_0$, which in turn is $2\sin\theta/\lambda$.

Furthermore, as d is the distance between sets of planes in the lattice, S is perpendicular to these planes, which are the reflecting planes characterised by hkl indices.

Constructive scattering from the crystal can occur only when h/S , k/S , l/S are equal to cell dimensions a , b , c . These conditions are called the Laue conditions.

A.5 Structure solution

A.5.1 From Structure Factor to Electron Density

The intensity of the diffracted beam is proportional to the square of the amplitude of the Structure Factor, $F(S)$.

The Structure Factor is a function of the electron density distribution in the unit cell:

$$F(s) = \sum_j f_j e^{2\pi i \mathbf{r}_j \cdot \mathbf{s}}$$

where f_j is the Structure Factor for a specific atom j and \mathbf{r} is its position in respect of the unit cell with origin in O ($\mathbf{r} = ax+by+cz$). Combining with the Laue conditions, the Structure Factor for each hkl is defined by:

$$F_{hkl} = \sum_j^n f_j e^{2\pi i (hx_j + ky_j + lz_j)}$$

Thus, the intensity of each of the spots in a diffraction pattern depends on the distribution of the electron density within the unit cell. The goal of the protein X-ray crystallography is to calculate the electron density ρ at every position x , y , z from the diffraction pattern collected.

Since the F_{hkl} is a summation over all the atoms j in the unit cell, it can be also express as integration over all the electrons in the unit cell:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) e^{2\pi i (hx+ky+lz)} dx dy dz$$

where ρ is the electron density and V the cell volume. Since $\rho(xyz)$ is the Fourier transformation of $F(hkl)$:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx+ky+lz)}$$

where $\rho(xyz)$ is expressed as summation because diffraction occurs only under Laue conditions.

F can be represented as a vector $F=|F|e^{i\alpha}$ (Figure A.9), which means that final the equation becomes:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)}$$

where $\rho(xyz)$ is the variable while $|F(hkl)|$ and the *phase angle* are the data which required to be defined experimentally. Although the module of F can be derived from the intensities, the phase angle cannot be derived straightforwardly from the diffraction pattern. This is the so called phase problem.

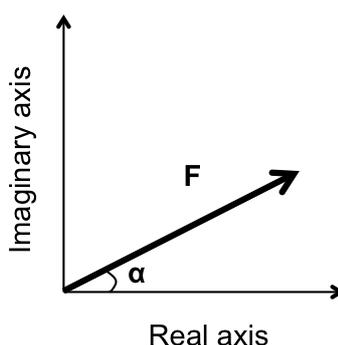


Figure A.9 The Structure Factor. The Structure Factor can be represented as a vector F on the Argand diagram. The module of the vector is F amplitude, while the inclination is the phase.

A.5.2 The phase problem

Different methods have been derived to overcome the phase problem. Isomorphous replacement, multiple and single anomalous diffraction method (MAD and SAD) are used when no information of the phases is known. Molecular replacement is used when phases can be calculated from a homologous model.

All the methods used in protein crystallography for solving structures require the knowledge of the Patterson function. The Patterson function $P(uvw)$ is a Fourier summation with intensities as coefficients and without phase angles or, rather, with all phase angles equal to zero:

$$P(uvw) = \frac{1}{V} |F(hkl)|^2 \cos 2\pi(hu + kv + lw)$$

where uvw are the relative coordinates to xyz in the unit cell. The dimensions of the real crystal lattice and the Patterson are the same. Since no phase knowledge is required, the Patterson map can be calculated from the intensities. If a real unit cell contains N atoms, the corresponding Patterson map will show N^2 peaks, which correspond to the vectors used from going from one atom to all the others, including the one with length 0 that goes from an atom to the same atom (Figure A.10). This kind of representation is useful when a limited number of atoms are considered.

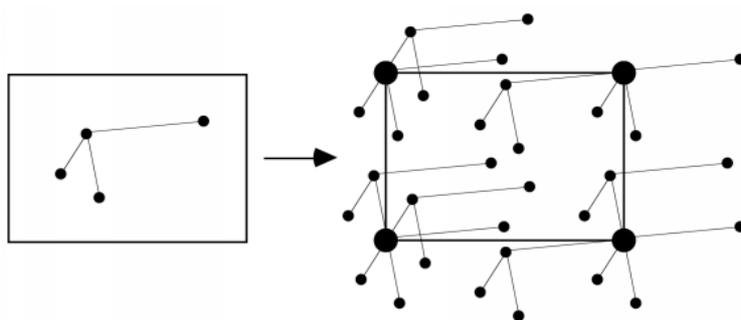


Figure A.10 The Patterson map. The figure illustrates a Patterson map corresponding to a cell with one molecule. The N self-peaks which correspond to the vector for going from one atom to the same one, overlap each other in the origin of the unit cell (self-Patterson map). The others are cross-vectors, which are the vectors for going from one atom to another. They are $N(N-1)$ for a molecule with N atoms (cross-Patterson map).

A.5.2.1 Isomorphus replacement and MAD/SAD

Isomorphus Replacement and the MAD-SAD methods for phase calculation are used when no homologous structure is available.

The first method developed was isomorphous replacement (Green, 1954). In this method, heavy metal atoms (such as mercury, platinum or gold) are introduced into the crystals that bind to specific residues (mercury binds to cysteine residues, platinum to methionine residues), naturally present or introduced into the protein, to create a heavy atom derivative (i.e. crystals with the same unit cell and orientation of protein). As heavy atoms are electron rich, the diffraction of the incident X-rays will differ from the diffraction from the native crystals. Thus, two $F(S)$ will be measured: the one of the native crystal F_p , and the one of the derivative F_{PH} . Referring to F vectoral representations, the difference vector is F_H , which amplitude depends on the heavy atoms position (Figure A.11).

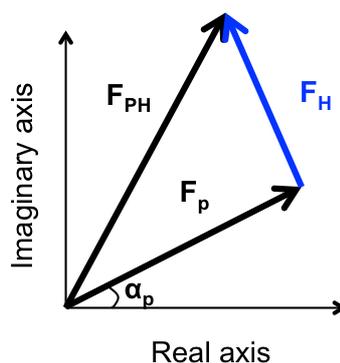


Figure A.11 Structure Factor of the heavy atom structure. F_H is the difference vector between the Structure Factor of the native molecule and the Structure Factor of the derivative one.

The heavy atom position is calculated using the Patterson map. The Patterson summation with the coefficients $(|F_{PH}| - |F_p|)^2$ gives the Patterson map of the heavy atom arrangement in the unit cell. Since the Patterson unit cell and the real space unit cell are related each other, the coordinates of the heavy atom in the lattice can be calculated. F_H amplitude is then estimated, including sign and phase angle.

Now the protein phase angles can be determined. Since F_H is known in amplitude and

phase, the phase problem is reduced to a phase ambiguity. The Harker construction provides an intuitive way of explaining how this phase solution is achieved (Figure A.12). A second derivative or more is required for solving the ambiguity. However, as errors are introduced in the F amplitude estimation and by poor isomorphism, the accuracy of the phase solution increases by increasing the number of derivatives used. For this reason the technique is better called Multiple Isomorphus Replacement (MIR).

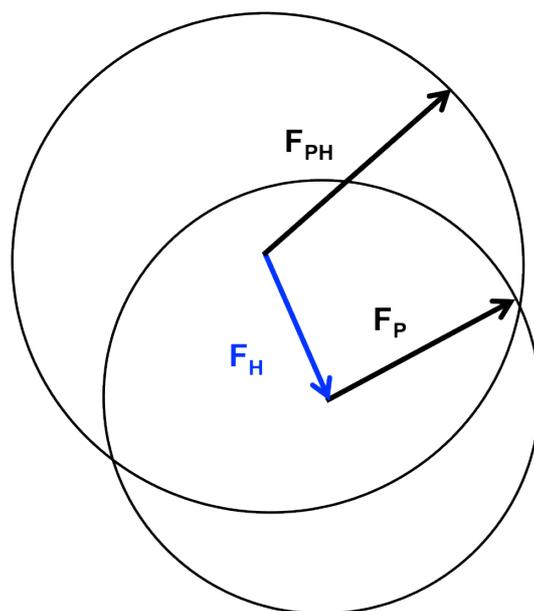


Figure A.12 The Harker construction. The Harker construction has been obtained by repeating the following instructions: 1) Draw a circle with radius $|F_P|$. 2) From the centre of the circle vector the calculated $-F_H$ is drawn. 3) A second circle with module $|F_{PH}|$ and centre in the end of the vector $-F_H$ is drawn. Two phase solutions are then obtained.

Today, MIR is not the technique of choice as there are several difficulties in practice. A long trial and error procedure is followed to obtain derivatives and the binding of the heavy atoms could cause loss in resolution and non-isomorphism. It is preferable to use heavy atoms which are, or can be naturally included in the protein, such as Zn, Fe and Cu in metalloproteins. Moreover, if the construct used contain methionines, it is possible to genetically modify the protein to incorporate a heavy atom such as Selenium. Se-Met (selenium-methionine) is a modified residue, which is equivalent to a

methionine where the sulphur is replaced by a selenium atom. It is easily absorbed by cells and bacteria during protein expression, and used to replace methionine during protein synthesis in minimal media that lacks the natural residue.

To overcome the problem of the number of derivatives required, the anomalous scattering phenomenon is considered. The presence of a heavy atom breaks Friedel's Law where $I_{hkl} = I_{-h-k-l}$ and leads to different intensities of Friedel pairs of reflections. Thus, for each heavy atom, two derivatives are obtained: F_{hkl} and F_{-h-k-l} . The use of the anomalous scattering, is the principle of the SAD or MAD method (Hendrickson, 1991).

In a typical SAD or MAD experiment, a fluorescent scan of the heavy atom introduced in the crystal performed. The scan measures the photon emission of the heavy atom around the X-ray wavelength corresponding to its absorption edge. The plot of the absorption as a function of the X-ray energy, will show a typical curve with a sharp change at a specific λ , which is called the absorption edge. The change refers to ejection of an electron caused by a specific X-ray energy, which is different for each atom and can be affected by the experimental environment. In the diffraction pattern, Friedel's law is broken and the maximum difference between the pairs $hkl/-h-k-l$ is achieved at the peak of the absorption edge. Thus, in order to solve the phase problem, data collection is performed at the wavelength of the absorption peak. During the data processing and reduction two Structure Factors are then estimated: $F(hkl)$ and $F(-h-k-l)$. This experiment is called SAD, because only one wavelength is used. The experiment requires Synchrotron radiation, where the X-ray wavelength is tuneable.

In a MAD experiment, up to four wavelengths are used, providing a much more accurate phase calculation. Four data sets can be collected, at the absorption peak (as in SAD), the point of inflection of the absorption curve, a remote wavelength and a high energy remote data set (Figure A.13). However, despite the accuracy that could be achieved, the experiment is time consuming and increase the probability of radiation damage in the crystal.

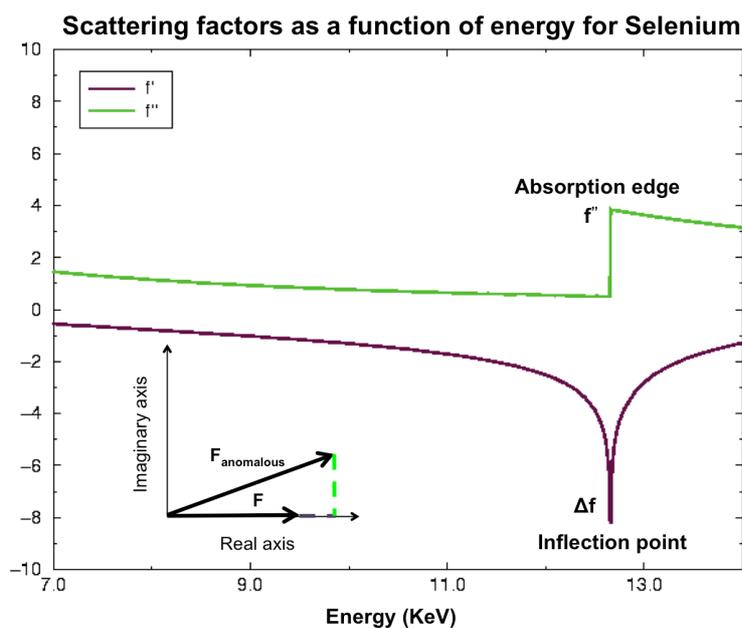


Figure A.13 The selenium k absorption edge. Anomalous scattering has two contributions: Δf and f'' . The curves are shown respectively in purple and green (www.esrf.eu). The minimum achieved by Δf curve is the inflection point or $\lambda 1$; the maximum achieved by f'' curve is the absorption edge or $\lambda 2$; $\lambda 3$ is the remote wavelength, which is far from both $\lambda 1$ and $\lambda 2$. At $\lambda = \lambda 2$ the anomalous difference is the largest.

Thus, in SAD or MAD data are collected on a single crystal, without any problem of isomorphism. The scattering factors of the anomalously scattering atoms are different, and therefore, the intensities of reflections will be different for each wavelength at which a data set is taken.

A.5.2.2 Molecular replacement

If the structure of a homologous protein is known the model can be used to estimate the phases. This process is called molecular replacement (MR). The goal of the molecular replacement process is to transfer the known protein molecular structure from its crystalline arrangement to the crystal of the protein of unknown structure. The process is a two-step procedure, which comprises the calculation of a rotation and a translation function. In the rotation function, the orientation of the new molecule in the unit cell will be found. As first described by Rossmann and Blow (Rossmann, 1990), the rotation function uses the Patterson maps to find the orientation, by looking for the function that

will overlap the *self-Patterson* map (see figure A.10). The two maps are expected to be very similar, apart from having a different orientation. The final result of the first step is a list of Eulerian angles or Polar angles, which described the rotation to apply to the Patterson function around an axis through its origin.

Then the translation function is applied, which defines the translation required to overlap one molecule in real space. The simplest way for calculating this is by trial and error. The known molecule is moved through the asymmetric unit and the structure factors are calculated and compared to the observed ones, by calculating a correlation coefficient. Alternatively, the translation function can be calculated by comparing the cross-Patterson vectors map, which is made by vectors between atoms of two different molecules that are related each other in the model structure by crystallographic symmetry.

A.5.3 Model building, Refinement and Validation

After phases have been estimated and electron density maps calculated, the model can be built. The electron density is interpreted and a first molecular model is built to fit the density. The structure factors calculated on the basis of this model (F_{calc}) are generally in poor agreement with the observed Structure Factors (F_{obs}). This is clearly shown by the difference Fourier map. The map shows positive electron density peaks at the site of atoms that were not present in the real structure and it shows negative peaks at the positions of the atoms present only in the model. The refinement is the process of adjusting the model to find a closer agreement between the calculated and the observed structure factors. The agreement index between the calculated and observed structure factors is usually represented by a R_{factor} , defined by the equation:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

where k is the scale factor.

It has been shown that R_{factor} could reach very small values for a protein structure that later appear to be incorrect. This is due to the high model parameters that are

considered. To avoid this pitfall, the *Free R_{factor}* is also used, which is unbiased from the refinement process. The reflection set is divided in two sets: the working-set (~95%) and the test-set (usually 5%). Refinement is run only on the working-set and the *Free R_{factor}* is calculated only on the test set with exactly the same equation. If a structure is really improving during the refinement step both *R_{factor}* will decrease. This process is called cross-validation.

During refinement, both atomic coordinates and temperature factors or B-factors of the molecule are refined against the observed amplitudes. The refinement of the temperature factors considers the phenomenon that X-rays do not interact with identical atoms on exactly the same position in successive unit cells. This is due to the fact that atoms vibrate around their equilibrium position. Where the components of the vibration are the same in all directions, the vibration is called isotropic. At high resolution, differences can be measured and the temperature factor is refined as anisotropic. However, refinement of a full anisotropic model, which comprises six parameters, can be challenging. TLS refinement is a way of modelling anisotropic displacements using only a few parameters. The structure is divided in groups, which are supposed to move as rigid bodies; the rigid-body motion is described by translation (T), libration (L) and screw (S) tensors using a total of 20 parameters for each group. Then anisotropic parameters for each atom are derived (Murshudov *et al.*, 2011). A correct refinement of the B-factors decreases the *R_{factor}* values by providing a better model.

The refinement process is usually run in a restrained way. This implies that limits for stereochemical parameters such as bond lengths, torsion angles and Van de Waals contacts, are introduced based on values determined from small molecules. Further NCS restraints can be added, which force same molecules related each other by non-crystallographic symmetry to be equal.

The function of minimising the difference between the *F_{calc}* and the *F_{obs}* can be compared to a trajectory towards a minimum value. If the distance between the model and the real structure is too large, refinement can be trapped in a local minimum. In this situation *simulating annealing* can be applied in order to ‘push’ the model out of a local minimum and let it reach the real minimum. The basic idea of this method belongs to the molecular dynamic system, where the behaviour of a particle can be simulated at specific temperature and pressure. In the case of a molecular structure, the idea is to

raise the temperature sufficiently high for the atoms to overcome energy barriers. The increase in energy increases the freedom in the stereochemical restraints, which lets the structure achieve conformations that would not be allowed in a normal restrained refinement. Indeed, the system can jump the local energy minima and cool down to the real minimum.

After the molecular model is refined, the structure needs to be validated. This is performed by checking the agreement between the structure stereochemistry and the conventional geometrical-chemical restraints. The stereochemistry of the main chain folding can be investigated with a Ramachandran plot, in which the dihedral angles φ and ψ allowed for each residue are plotted in a square matrix. Furthermore, unusual ω angles, eclipsed dihedral angles, unsatisfied high B-factors, unpaired charged residues should be checked carefully.

References

- Adams JL, Badger AM, Kumar S, Lee JC (2001) p38 MAP kinase: molecular target for the inhibition of pro-inflammatory cytokines. *Progress in medicinal chemistry* **38**: 1-60
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography* **66**: 213-221
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Echols N, Headd JJ, Hung LW, Jain S, Kapral GJ, Grosse Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner RD, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2011) The Phenix software for automated determination of macromolecular structures. *Methods* **55**: 94-106
- Admiraal SJ, Herschlag D (1995) Mapping the transition state for ATP hydrolysis: implications for enzymatic catalysis. *Chemistry & biology* **2**: 729-739
- Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica Section D, Biological crystallography* **68**: 352-367
- Akella R, Min X, Wu Q, Gardner KH, Goldsmith EJ (2010) The third conformation of p38alpha MAP kinase observed in phosphorylated p38alpha and in solution. *Structure* **18**: 1571-1578
- Akella R, Moon TM, Goldsmith EJ (2008) Unique MAP Kinase binding sites. *Biochimica et biophysica acta* **1784**: 48-55
- Alberts B. (2008) Molecular biology of the cell. Garland Science,, New York ; Abingdon.
- Aleshin AE, Kirby C, Liu X, Bourenkov GP, Bartunik HD, Fromm HJ, Honzatko RB (2000) Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation. *Journal of molecular biology* **296**: 1001-1015
- Alhambra C, Wu L, Zhang ZY, Gao JL (1998) Walden-inversion-enforced transition-state stabilization in a protein tyrosine phosphatase. *J Am Chem Soc* **120**: 3858-3866
- Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta crystallographica Section B, Structural science* **58**: 380-388
- Allen WE, Zicha D, Ridley AJ, Jones GE (1998) A role for Cdc42 in macrophage chemotaxis. *The Journal of cell biology* **141**: 1147-1157
- Alphey MS, Burton A, Urbaniak MD, Boons GJ, Ferguson MA, Hunter WN (2006) Trypanosoma brucei UDP-galactose-4'-epimerase in ternary complex with NAD⁺ and the substrate analogue UDP-4-deoxy-4-fluoro-alpha-D-galactose. *Acta crystallographica Section F, Structural biology and crystallization communications* **62**: 829-834
- Anderson CM, Stenkamp RE, McDonald RC, Steitz TA (1978) A refined model of the sugar binding site of yeast hexokinase B. *Journal of molecular biology* **123**: 207-219
- Anderson NG, Maller JL, Tonks NK, Sturgill TW (1990) Requirement for integration of signals from two distinct phosphorylation pathways for activation of MAP kinase. *Nature* **343**: 651-653

- Andersson U, Radstrom P (2002a) Beta-glucose 1-phosphate-interconverting enzymes in maltose- and trehalose-fermenting lactic acid bacteria. *Environmental microbiology* **4**: 81-88
- Andersson U, Radstrom P (2002b) Physiological function of the maltose operon regulator, MalR, in *Lactococcus lactis*. *BMC microbiology* **2**: 28
- Aoki K, Yamada M, Kunida K, Yasuda S, 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**: 12675-12680
- Asthagiri D, Dillet V, Liu TQ, Noodleman L, Van Etten RL, Bashford D (2002) Density functional study of the mechanism of a tyrosine phosphatase: 1. Intermediate formation. *J Am Chem Soc* **124**: 10225-10235
- Bøhm H-J, Schneider G (2003) *Protein-ligand interactions from molecular recognition to drug design*, Weinheim: Wiley-VCH.
- Bailey JM, Fishman PH, Pentchev PG (1970) Anomalous mutarotation of glucose 6-phosphate. An example of intramolecular catalysis. *Biochemistry* **9**: 1189-1194
- Bardwell AJ, Flatauer LJ, Matsukuma K, Thorner J, Bardwell L (2001) A conserved docking site in MEKs mediates high-affinity binding to MAP kinases and cooperates with a scaffold protein to enhance signal transmission. *The Journal of biological chemistry* **276**: 10374-10386
- Bardwell AJ, Frankson E, Bardwell L (2009) Selectivity of docking sites in MAPK kinases. *The Journal of biological chemistry* **284**: 13165-13173
- Bardwell L (2006) Mechanisms of MAPK signalling specificity. *Biochemical Society transactions* **34**: 837-841
- Barnette WE (1984) The synthesis and biology of fluorinated prostacyclins. *CRC critical reviews in biochemistry* **15**: 201-235
- Barrett T, Xiao B, Dodson EJ, Dodson G, Ludbrook SB, Nurmahomed K, Gamblin SJ, Musacchio A, Smerdon SJ, Eccleston JF (1997) The structure of the GTPase-activating domain from p50rhoGAP. *Nature* **385**: 458-461
- Barsyte-Lovejoy D, Galanis A, Sharrocks AD (2002) Specificity determinants in MAPK signaling to transcription factors. *The Journal of biological chemistry* **277**: 9896-9903
- Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica Section D, Biological crystallography* **67**: 271-281
- Baxter NJ, Blackburn GM, Marston JP, Hounslow AM, Cliff MJ, Bermel W, Williams NH, Hollfelder F, Wemmer DE, Waltho JP (2008) Anionic charge is prioritized over geometry in aluminum and magnesium fluoride transition state analogs of phosphoryl transfer enzymes. *Journal of the American Chemical Society* **130**: 3952-3958
- Baxter NJ, Bowler MW, Alizadeh T, Cliff MJ, Hounslow AM, Wu B, Berkowitz DB, Williams NH, Blackburn GM, Waltho JP (2010) Atomic details of near-transition state conformers for enzyme phosphoryl transfer revealed by MgF₃ rather than by phosphoranes. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 4555-4560

- Baxter NJ, Olguin LF, Golicnik M, Feng G, Hounslow AM, Bermel W, Blackburn GM, Hollfelder F, Waltho JP, Williams NH (2006) A Trojan horse transition state analogue generated by MgF₃-formation in an enzyme active site. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14732-14737
- Behr J-BE, C. M.; Phung, N.; Guillermin, G. (1997) Synthesis of (difluoromethyl)phosphonate azasugras designed as inhibitors for glycosyl transferase. *J Chem Soc, Perkin Trans 1*
- Bellon S, Fitzgibbon MJ, Fox T, Hsiao HM, Wilson KP (1999) The structure of phosphorylated p38gamma is monomeric and reveals a conserved activation-loop conformation. *Structure* **7**: 1057-1065
- Berg JM, Tymoczko JL (2002) *Biochemistry*, New York: Freeman, W.H.
- Bergfors T (2003) Seeds to crystals. *Journal of structural biology* **142**: 66-76
- Berkowitz DB, Bose M, Pfannenstiel TJ, Doukov T (2000) alpha-fluorinated phosphonates as substrate mimics for glucose 6-phosphate dehydrogenase: the CHF stereochemistry matters. *The Journal of organic chemistry* **65**: 4498-4508
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic acids research* **28**: 235-242
- Bi X, Corpina RA, Goldberg J (2002) Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**: 271-277
- Bishop AL, Hall A (2000) Rho GTPases and their effector proteins. *The Biochemical journal* **348 Pt 2**: 241-255
- Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. *Journal of medicinal chemistry* **53**: 5061-5084
- Blackburn GM, England, D. A. and Kolkman, F. (1981) Monofluoro- and difluoro-methylenebisphosphonic acids: isopolar analogues of pyrophosphoric acid. *J Chem Soc, Chem Commun*: 930-932
- Boehm JC, Smietana JM, Sorenson ME, Garigipati RS, Gallagher TF, Sheldrake PL, Bradbeer J, Badger AM, Laydon JT, Lee JC, Hillegass LM, Griswold DE, Breton JJ, Chabot-Fletcher MC, Adams JL (1996) 1-substituted 4-aryl-5-pyridinylimidazoles: a new class of cytokine suppressive drugs with low 5-lipoxygenase and cyclooxygenase inhibitory potency. *Journal of medicinal chemistry* **39**: 3929-3937
- Bohm HJ, Banner D, Bendels S, Kansy M, Kuhn B, Muller K, Obst-Sander U, Stahl M (2004) Fluorine in medicinal chemistry. *Chembiochem : a European journal of chemical biology* **5**: 637-643
- Bokoch GM (2000) Regulation of cell function by Rho family GTPases. *Immunologic research* **21**: 139-148
- Bondi A (1964) Van Der Waals Volumes + Radii. *J Phys Chem-U.S* **68**: 441-&
- Bos JL, Rehmann H, Wittinghofer A (2007) GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **129**: 865-877

- Bourne HR, Sanders DA, McCormick F (1990) The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**: 125-132
- Bourne N, Williams AT (1983) The question of concerted or stepwise mechanisms in phosphoryl group (-PO₃²⁻) transfer to pyridines from isoquinoline-N-phosphonate. *J Am Chem Soc* **105**: 3357-3358
- Bourne N, Williams AT (1984) Evidence for a single transition state in the transfer of the phosphoryl group (-PO₃²⁻) to nitrogen nucleophiles from pyridino-N-phosphonates. *J Am Chem Soc* **106**: 7591-7596
- Bowler MW, Cliff MJ, Waltho JP, Blackburn GM (2010a) Why did Nature select phosphate for its dominant roles in biology? *New J Chem* **34**: 784-794
- Bowler MW, Guijarro M, Petitdemange S, Baker I, Svensson O, Burghammer M, Mueller-Dieckmann C, Gordon EJ, Flot D, McSweeney SM, Leonard GA (2010b) Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. *Acta crystallographica Section D, Biological crystallography* **66**: 855-864
- Brock J, Midwinter K, Lewis J, Martin P (1996) Healing of incisional wounds in the embryonic chick wing bud: characterization of the actin purse-string and demonstration of a requirement for Rho activation. *The Journal of cell biology* **135**: 1097-1107
- Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475
- Buchwald SL, Friedman JM, Knowles JR (1984) Stereochemistry of nucleophilic displacement on two phosphoric monoesters and a phosphoguanidine: the role of metaphosphate. *J Am Chem Soc* **106**: 4911-4916
- Buchwald SL, Hansen DE, Hassett A, Knowles JR (1982) Chiral [16O, 17O, 18O]phosphoric monoesters as stereochemical probes of phosphotransferases. *Methods in enzymology* **87**: 279-301
- Burack WR, Sturgill TW (1997) The activating dual phosphorylation of MAPK by MEK is nonprocessive. *Biochemistry* **36**: 5929-5933
- Bustelo XR, Sauzeau V, Berenjano IM (2007) GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *BioEssays : news and reviews in molecular, cellular and developmental biology* **29**: 356-370
- Caffrey DR, O'Neill LA, Shields DC (1999) The evolution of the MAP kinase pathways: coduplication of interacting proteins leads to new signaling cascades. *Journal of molecular evolution* **49**: 567-582
- Canagarajah BJ, Khokhlatchev A, Cobb MH, Goldsmith EJ (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* **90**: 859-869
- Cantrell D (1998) Lymphocyte signalling: a coordinating role for Vav? *Current biology : CB* **8**: R535-538
- Caron E, Hall A (1998) Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases. *Science* **282**: 1717-1721
- Cassano AG, Anderson VE, Harris ME (2002) Evidence for direct attack by hydroxide in phosphodiester hydrolysis. *J Am Chem Soc* **124**: 10964-10965

Chang CI, Xu BE, Akella R, Cobb MH, Goldsmith EJ (2002) Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b. *Molecular cell* **9**: 1241-1249

Chayen NE, Cianci M, Olczak A, Raftery J, Rizkallah PJ, Zagalsky PF, Helliwell JR (2000) Apocrustacyanin A1 from the lobster carotenoprotein alpha-crustacyanin: crystallization and initial X-ray analysis involving softer X-rays. *Acta crystallographica Section D, Biological crystallography* **56**: 1064-1066

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010a) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* **66**: 12-21

Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu B, Wright A, Vanderbilt C, Cobb MH (2001) MAP kinases. *Chemical reviews* **101**: 2449-2476

Chen Z, Medina F, Liu MY, Thomas C, Sprang SR, Sternweis PC (2010b) Activated RhoA binds to the pleckstrin homology (PH) domain of PDZ-RhoGEF, a potential site for autoregulation. *The Journal of biological chemistry* **285**: 21070-21081

Clader JW (2004) The discovery of ezetimibe: a view from outside the receptor. *Journal of medicinal chemistry* **47**: 1-9

Cleland WW, Hengge AC (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chemical reviews* **106**: 3252-3278

Cliff MJ, Bowler MW, Varga A, Marston JP, Szabo J, Hounslow AM, Baxter NJ, Blackburn GM, Vas M, Waltho JP (2010a) Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *J Am Chem Soc* **132**: 6507-6516

Cliff MJ, Bowler MW, Varga A, Marston JP, Szabo J, Hounslow AM, Baxter NJ, Blackburn GM, Vas M, Waltho JP (2010b) Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *Journal of the American Chemical Society* **132**: 6507-6516

Cobb MH, Goldsmith EJ (1995) How MAP kinases are regulated. *The Journal of biological chemistry* **270**: 14843-14846

Coleman DE, Lee E, Mixon MB, Linder ME, Berghuis AM, Gilman AG, Sprang SR (1994) Crystallization and preliminary crystallographic studies of Gi alpha 1 and mutants of Gi alpha 1 in the GTP and GDP-bound states. *Journal of molecular biology* **238**: 630-634

Colicelli J (2004) Human RAS superfamily proteins and related GTPases. *Science's STKE : signal transduction knowledge environment* **2004**: RE13

Condeelis J (2001) How is actin polymerization nucleated in vivo? *Trends in cell biology* **11**: 288-293

Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica Section D, Biological crystallography* **62**: 1002-1011

Crowther RA (1972) *The Molecular Replacement Method*, New York: Gordon & Breach.

- D'Arcy A, Mac Sweeney A, Habera A (2004) Modified microbatch and seeding in protein crystallization experiments. *Journal of synchrotron radiation* **11**: 24-26
- Dai J, Wang L, Allen KN, Radstrom P, Dunaway-Mariano D (2006) Conformational cycling in beta-phosphoglucosyltransferase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate. *Biochemistry* **45**: 7818-7824
- Dai JB, Liu Y, Ray WJ, Jr., Konno M (1992) The crystal structure of muscle phosphoglucosyltransferase refined at 2.7-angstrom resolution. *The Journal of biological chemistry* **267**: 6322-6337
- Dalvit C, Vulpetti A (2011) Fluorine-protein interactions and (1)(9)F NMR isotropic chemical shifts: An empirical correlation with implications for drug design. *ChemMedChem* **6**: 104-114
- Dalvit C, Vulpetti A (2012) Intermolecular and intramolecular hydrogen bonds involving fluorine atoms: implications for recognition, selectivity, and chemical properties. *ChemMedChem* **7**: 262-272
- Dauter Z, Dauter M, Dodson E (2002) Jolly SAD. *Acta crystallographica Section D, Biological crystallography* **58**: 494-506
- Davis RJ (1995) Transcriptional regulation by MAP kinases. *Molecular reproduction and development* **42**: 459-467
- De Bondt HL, Rosenblatt J, Jancarik J, Jones HD, Morgan DO, Kim SH (1993) Crystal structure of cyclin-dependent kinase 2. *Nature* **363**: 595-602
- Del Campo M, Lambowitz AM (2009) Crystallization and preliminary X-ray diffraction of the DEAD-box protein Mss116p complexed with an RNA oligonucleotide and AMP-PNP. *Acta crystallographica Section F, Structural biology and crystallization communications* **65**: 832-835
- Deng Q, Liao R, Wu BL, Sun P (2004) High intensity ras signaling induces premature senescence by activating p38 pathway in primary human fibroblasts. *The Journal of biological chemistry* **279**: 1050-1059
- Dickinson RJ, Keyse SM (2006) Diverse physiological functions for dual-specificity MAP kinase phosphatases. *Journal of cell science* **119**: 4607-4615
- Drenth J (1999) *Principles of protein x-ray crystallography*, 2nd edn. New York ; London: Springer.
- Ebnet K, Suzuki A, Horikoshi Y, Hirose T, Meyer Zu Brickwedde MK, Ohno S, Vestweber D (2001) The cell polarity protein ASIP/PAR-3 directly associates with junctional adhesion molecule (JAM). *The EMBO journal* **20**: 3738-3748
- Ellis RJ (2001a) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current opinion in structural biology* **11**: 114-119
- Ellis RJ (2001b) Macromolecular crowding: obvious but underappreciated. *Trends in biochemical sciences* **26**: 597-604
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* **60**: 2126-2132
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta crystallographica Section D, Biological crystallography* **66**: 486-501
- Engel R (1977) Phosphonates as analogues of natural phosphates. *Chem Rev* **77**: 349-367

- Enslin H, Brancho DM, Davis RJ (2000) Molecular determinants that mediate selective activation of p38 MAP kinase isoforms. *The EMBO journal* **19**: 1301-1311
- Etienne-Manneville S, Hall A (2002) Rho GTPases in cell biology. *Nature* **420**: 629-635
- Evans P (2006) Scaling and assessment of data quality. *Acta crystallographica Section D, Biological crystallography* **62**: 72-82
- Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica Section D, Biological crystallography* **67**: 282-292
- Feig LA (1999) Tools of the trade: use of dominant-inhibitory mutants of Ras-family GTPases. *Nature cell biology* **1**: E25-27
- Fera MT, Giannone M, Pallio S, Tortora A, Blandino G, Carbone M (2001) Antimicrobial activity and postantibiotic effect of flurithromycin against *Helicobacter pylori* strains. *International journal of antimicrobial agents* **17**: 151-154
- Ferrell JE, Jr., Bhatt RR (1997) Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *The Journal of biological chemistry* **272**: 19008-19016
- Flot D, Mairs T, Giraud T, Guijarro M, Lesourd M, Rey V, van Brussel D, Morawe C, Borel C, Hignette O, Chavanne J, Nurizzo D, McSweeney S, Mitchell E (2010) The ID23-2 structural biology microfoc beamline at the ESRF. *Journal of synchrotron radiation* **17**: 107-118
- French S, Wilson K (1978) On the treatment of negative intensity observations. *Acta crystallographica Section A, Foundations of crystallography* **34**: 517-525
- Fukata Y, Amano M, Kaibuchi K (2001) Rho-Rho-kinase pathway in smooth muscle contraction and cytoskeletal reorganization of non-muscle cells. *Trends in pharmacological sciences* **22**: 32-39
- Gabadinho J, Beteva A, Guijarro M, Rey-Bakaikoa V, Spruce D, Bowler MW, Brockhauser S, Flot D, Gordon EJ, Hall DR, Lavault B, McCarthy AA, McCarthy J, Mitchell E, Monaco S, Mueller-Dieckmann C, Nurizzo D, Ravelli RB, Thibault X, Walsh MA, Leonard GA, McSweeney SM (2010) MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *Journal of synchrotron radiation* **17**: 700-707
- Garai A, Zeke A, Gogl G, Toro I, Fordos F, Blankenburg H, Barkai T, Varga J, Alexa A, Emig D, Albrecht M, Remenyi A (2012) Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. *Science signaling* **5**: ra74
- Garman E (1999) Cool data: quantity AND quality. *Acta crystallographica Section D, Biological crystallography* **55**: 1641-1653
- Garman EF, Schneider TR (1997) Macromolecular Cryocrystallography. *JApplCryst* **30**: 211-237
- Gavin AC, Nebreda AR (1999) A MAP kinase docking site is required for phosphorylation and activation of p90(rsk)/MAPKAP kinase-1. *Current biology* : **CB 9**: 281-284
- Gerhart JC, Pardee AB (1962) The enzymology of control by feedback inhibition. *The Journal of biological chemistry* **237**: 891-896
- Geyer M, Wittinghofer A (1997) GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins. *Current opinion in structural biology* **7**: 786-792

- Gibbs CS, Zoller MJ (1991) Identification of electrostatic interactions that determine the phosphorylation site specificity of the cAMP-dependent protein kinase. *Biochemistry* **30**: 5329-5334
- Golicnik M, Olguin LF, Feng G, Baxter NJ, Waltho JP, Williams NH, Hollfelder F (2009) Kinetic analysis of beta-phosphoglucomutase and its inhibition by magnesium fluoride. *Journal of the American Chemical Society* **131**: 1575-1588
- Gomez del Pulgar T, Benitah SA, Valeron PF, Espina C, Lacal JC (2005) Rho GTPase expression in tumourigenesis: evidence for a significant link. *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**: 602-613
- Gotta M, Abraham MC, Ahringer J (2001) CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. *Current biology : CB* **11**: 482-488
- Graham DL, Eccleston JF, Lowe PN (1999) The conserved arginine in rho-GTPase-activating protein is essential for efficient catalysis but not for complex formation with Rho.GDP and aluminum fluoride. *Biochemistry* **38**: 985-991
- Graham DL, Lowe PN, Grime GW, Marsh M, Rittinger K, Smerdon SJ, Gamblin SJ, Eccleston JF (2002) MgF(3)(-) as a transition state analog of phosphoryl transfer. *Chemistry & biology* **9**: 375-381
- Green DW (1954) The structure of Haemoglobin IV. *ProcRoy Soc A* 287-387
- Griffin J (2011) Investigation of the metal fluoride transition state and ground state analogue complexes of HAD superfamily proteins by nuclear magnetic resonance spectroscopy. Department of Molecular Biology and Biotechnology, The University of Sheffield,
- Griffin JL, Bowler MW, Baxter NJ, Leigh KN, Dannatt HR, Hounslow AM, Blackburn GM, Webster CE, Cliff MJ, Waltho JP (2012) Near attack conformers dominate beta-phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 6910-6915
- Grzyska PK, Czyryca PG, Purcell J, Hengge AC (2003) Transition state differences in hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions. *Journal of the American Chemical Society* **125**: 13106-13111
- Guixe V, Merino F (2009) The ADP-dependent sugar kinase family: kinetic and evolutionary aspects. *IUBMB life* **61**: 753-761
- Gum RJ, Young PR (1999) Identification of two distinct regions of p38 MAPK required for substrate binding and phosphorylation. *Biochemical and biophysical research communications* **266**: 284-289
- Hakem A, Sanchez-Sweetman O, You-Ten A, Duncan G, Wakeham A, Khokha R, Mak TW (2005) RhoC is dispensable for embryogenesis and tumor initiation but essential for metastasis. *Genes & development* **19**: 1974-1979
- Hall A (1998) Rho GTPases and the actin cytoskeleton. *Science* **279**: 509-514
- Hall A (2012) Rho family GTPases. *Biochemical Society transactions* **40**: 1378-1382
- Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* **100**: 57-70

- Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**: 42-52
- Harrop SJ, Helliwell JR, Wan TC, Kalb AJ, Tong L, Yariv J (1996) Structure solution of a cubic crystal of concanavalin A complexed with methyl alpha-D-glucopyranoside. *Acta crystallographica Section D, Biological crystallography* **52**: 143-155
- Hassett A, Blattler W, Knowles JR (1982) Pyruvate kinase: is the mechanism of phospho transfer associative or dissociative? *Biochemistry* **21**: 6335-6340
- Haystead TA, Dent P, Wu J, Haystead CM, Sturgill TW (1992) Ordered phosphorylation of p42mapk by MAP kinase kinase. *FEBS letters* **306**: 17-22
- Heasman SJ, Ridley AJ (2008) Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nature reviews Molecular cell biology* **9**: 690-701
- Hemmer W, McGlone M, Tsigelny I, Taylor SS (1997) Role of the glycine triad in the ATP-binding site of cAMP-dependent protein kinase. *The Journal of biological chemistry* **272**: 16946-16954
- Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51-58
- Hengge AC (2002) Isotope effects in the study of phosphoryl and sulfuranyl transfer reactions. *Accounts of chemical research* **35**: 105-112
- Hengge AC, Edens WA, Elsing H (1994) Transition-state structures for phosphoryl-transfer reactions of p-nitrophenyl phosphate. *J Am Chem Soc* **116**: 5045-5049
- Heo YS, Kim SK, Seo CI, Kim YK, Sung BJ, Lee HS, Lee JI, Park SY, Kim JH, Hwang KY, Hyun YL, Jeon YH, Ro S, Cho JM, Lee TG, Yang CH (2004) Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125. *The EMBO journal* **23**: 2185-2195
- Herschlag D, Jencks WP (1987) The Effect of Divalent Metal-Ions on the Rate and Transition-State Structure of Phosphoryl-Transfer Reactions. *J Am Chem Soc* **109**: 4665-4674
- Holtz KM, Stec B, Kantrowitz ER (1999) A model of the transition state in the alkaline phosphatase reaction. *The Journal of biological chemistry* **274**: 8351-8354
- Huang M, Prendergast GC (2006) RhoB in cancer suppression. *Histology and histopathology* **21**: 213-218
- Ihara K, Muraguchi S, Kato M, Shimizu T, Shirakawa M, Kuroda S, Kaibuchi K, Hakoshima T (1998) Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue. *J Biol Chem* **273**: 9656-9666
- Incardona MF, Bourenkov GP, Levik K, Pieritz RA, Popov AN, Svensson O (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation* **16**: 872-879
- Ip YT, Davis RJ (1998) Signal transduction by the c-Jun N-terminal kinase (JNK)--from inflammation to development. *Current opinion in cell biology* **10**: 205-219
- Jacinto A, Martinez-Arias A, Martin P (2001) Mechanisms of epithelial fusion and repair. *Nature cell biology* **3**: E117-123

- Jencks WP (1972) Structure-reactivity correlations and general acid-base catalysis in enzymic transacylation reactions. *Cold Spring Harbor symposia on quantitative biology* **36**: 1-11
- Jencks WP (1987) *Catalysis in chemistry and enzymology*, New York: over publication, Inc.
- Jin Y (2012) Metal Fluorides as Probes for Enzyme Catalysed Phosphoryl Transfer. Department of Molecular Biology and Biotechnology, The University of Sheffield
- Jin Y, Cliff MJ, Baxter NJ, Dannatt HR, Hounslow AM, Bowler MW, Blackburn GM, Waltho JP (2012) Charge-Balanced Metal Fluoride Complexes for Protein Kinase A with Adenosine Diphosphate and Substrate Peptide SP20. *Angewandte Chemie* **51**: 12242-12245
- Johnson GL, Lapadat R (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**: 1911-1912
- Johnson KA (2008) Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *The Journal of biological chemistry* **283**: 26297-26301
- Johnson LN, Noble ME, Owen DJ (1996) Active and inactive protein kinases: structural basis for regulation. *Cell* **85**: 149-158
- Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J (2011) Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Molecular cell* **42**: 9-22
- Just I, Selzer J, Hofmann F, Green GA, Aktories K (1996) Inactivation of Ras by Clostridium sordellii lethal toxin-catalyzed glucosylation. *The Journal of biological chemistry* **271**: 10149-10153
- Kabsch W (2010) Xds. *Acta crystallographica Section D, Biological crystallography* **66**: 125-132
- Kallunki T, Deng TL, Hibi M, Karin M (1996) c-jun Can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell* **87**: 929-939
- Kaminska B (2005) MAPK signalling pathways as molecular targets for anti-inflammatory therapy--from molecular mechanisms to therapeutic benefits. *Biochimica et biophysica acta* **1754**: 253-262
- Kamps MP, Sefton BM (1986) Neither arginine nor histidine can carry out the function of lysine-295 in the ATP-binding site of p60src. *Molecular and cellular biology* **6**: 751-757
- Karnoub AE, Weinberg RA (2008) Ras oncogenes: split personalities. *Nature reviews Molecular cell biology* **9**: 517-531
- Kay AJ, Hunter CP (2001) CDC-42 regulates PAR protein localization and function to control cellular and embryonic polarity in *C. elegans*. *Current biology : CB* **11**: 474-481
- Kim EE, Wyckoff HW (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *Journal of molecular biology* **218**: 449-464
- Kimura K, Ito M, Amano M, Chihara K, Fukata Y, Nakafuku M, Yamamori B, Feng J, Nakano T, Okawa K, Iwamatsu A, Kaibuchi K (1996) Regulation of myosin phosphatase by Rho and Rho-associated kinase (Rho-kinase). *Science* **273**: 245-248
- Kirby AJ, Jencks WP (1965) Reactivity of nucleophilic reagents toward p-nitrophenyl phosphate dianion. *J Am Chem Soc* **87**: 3209-3216
- Kirby AJ, Varvoglis AG (1967) Reactivity of phosphate esters. Monoester hydrolysis. *J Am Chem Soc* **89**: 415-423

Kirchhausen T, Rosen FS (1996) Disease mechanism: unravelling Wiskott-Aldrich syndrome. *Current biology : CB* **6**: 676-678

Kitago Y, Watanabe N, Tanaka I (2005) Structure determination of a novel protein by sulfur SAD using chromium radiation in combination with a new crystal-mounting method. *Acta crystallographica Section D, Biological crystallography* **61**: 1013-1021

Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**: 407-414

Knowles JR (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry* **49**: 877-919

Knust E (1997) Drosophila morphogenesis: movements behind the edge. *Current biology : CB* **7**: R558-561

Kornev AP, Haste NM, Taylor SS, Eyck LF (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17783-17788

Kornev AP, Taylor SS (2010) Defining the conserved internal architecture of a protein kinase. *Biochimica et biophysica acta* **1804**: 440-444

Kornev AP, Taylor SS, Ten Eyck LF (2008) A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 14377-14382

Kovensky J, Duchaussoy P, Bono F, Salmivirta M, Sizun P, Herbert JM, Petitou M, Sinay P (1999) A synthetic heparan sulfate pentasaccharide, exclusively containing L-iduronic acid, displays higher affinity for FGF-2 than its D-glucuronic acid-containing isomers. *Bioorganic & medicinal chemistry* **7**: 1567-1580

Kraynov VS, Chamberlain C, Bokoch GM, Schwartz MA, Slabaugh S, Hahn KM (2000) Localized Rac activation dynamics visualized in living cells. *Science* **290**: 333-337

Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica Section D, Biological crystallography* **60**: 2256-2268

Kumar S, McLaughlin MM, McDonnell PC, Lee JC, Livi GP, Young PR (1995) Human mitogen-activated protein kinase CSBP1, but not CSBP2, complements a hog1 deletion in yeast. *The Journal of biological chemistry* **270**: 29043-29046

Kyriakis JM, Avruch J (1996) Protein kinase cascades activated by stress and inflammatory cytokines. *BioEssays : news and reviews in molecular, cellular and developmental biology* **18**: 567-577

Lad C, Williams NH, Wolfenden R (2003) The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 5607-5610

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

- Lahiri SD, Wang PF, Babbitt PC, McLeish MJ, Kenyon GL, Allen KN (2002a) The 2.1 Å structure of *Torpedo californica* creatine kinase complexed with the ADP-Mg(2+)-NO(3)(-)-creatine transition-state analogue complex. *Biochemistry* **41**: 13861-13867
- Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN (2002b) Caught in the act: the structure of phosphorylated beta-phosphoglucosyltransferase from *Lactococcus lactis*. *Biochemistry* **41**: 8351-8359
- Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* **299**: 2067-2071
- Laidler KJ, King CM (1983) Development of transition-state theory. *J Phys Chem* **87**: 2657-2664
- Lamarche N, Hall A (1994) GAPs for rho-related GTPases. *Trends in genetics : TIG* **10**: 436-440
- Lamzin VS, Wilson KS (1993) Automated refinement of protein models. *Acta crystallographica Section D, Biological crystallography* **49**: 129-147
- Lancaster CA, Taylor-Harris PM, Self AJ, Brill S, van Erp HE, Hall A (1994) Characterization of rhoGAP. A GTPase-activating protein for rho-related small GTPases. *The Journal of biological chemistry* **269**: 1137-1142
- Lang PT, Ng HL, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T (2010) Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein science : a publication of the Protein Society* **19**: 1420-1431
- Lassila JK, Zalatan JG, Herschlag D (2011) Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annual review of biochemistry* **80**: 669-702
- Laughlin JD, Nwachukwu JC, Figuera-Losada M, Cherry L, Nettles KW, LoGrasso PV (2012) Structural mechanisms of allostery and autoinhibition in JNK family kinases. *Structure* **20**: 2174-2184
- Lee SJ, Zhou T, Goldsmith EJ (2006) Crystallization of MAP kinases. *Methods* **40**: 224-233
- Lee T, Winter C, Marticke SS, Lee A, Luo L (2000) Essential roles of *Drosophila* RhoA in the regulation of neuroblast proliferation and dendritic but not axonal morphogenesis. *Neuron* **25**: 307-316
- Leslie AG (2006) The integration of macromolecular diffraction data. *Acta crystallographica Section D, Biological crystallography* **62**: 48-57
- Li Z, Van Aelst L, Cline HT (2000) Rho GTPases regulate distinct aspects of dendritic arbor growth in *Xenopus* central neurons in vivo. *Nature neuroscience* **3**: 217-225
- Lindley PF (1999) Macromolecular crystallography with a third-generation synchrotron source. *Acta crystallographica Section D, Biological crystallography* **55**: 1654-1662
- Liu S, Sun JP, Zhou B, Zhang ZY (2006) Structural basis of docking interactions between ERK2 and MAP kinase phosphatase 3. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 5326-5331
- Longenecker K, Read P, Lin SK, Somlyo AP, Nakamoto RK, Derewenda ZS (2003) Structure of a constitutively activated RhoA mutant (Q63L) at 1.55 Å resolution. *Acta crystallographica Section D, Biological crystallography* **59**: 876-880

Lowry TH, Richardson KS (1987) *Mechanism and theory in organic chemistry*, 3rd edn. New York: Harper & Row.

Lu Q, Longo FM, Zhou H, Massa SM, Chen YH (2009) Signaling through Rho GTPase pathway as viable drug target. *Current medicinal chemistry* **16**: 1355-1365

Lu Y, Settleman J (1999) The Drosophila Pkn protein kinase is a Rho/Rac effector target required for dorsal closure during embryogenesis. *Genes & development* **13**: 1168-1180

Luo L (2000) Rho GTPases in neuronal morphogenesis. *Nature reviews Neuroscience* **1**: 173-180

Mabe S, Eller J, Champney WS (2004) Structure-activity relationships for three macrolide antibiotics in Haemophilus influenzae. *Current microbiology* **49**: 248-254

Machacek M, Hodgson L, Welch C, Elliott H, Pertz O, Nalbant P, Abell A, Johnson GL, Hahn KM, Danuser G (2009) Coordination of Rho GTPase activities during cell protrusion. *Nature* **461**: 99-103

Mackay DJ, Hall A (1998) Rho GTPases. *The Journal of biological chemistry* **273**: 20685-20688

Madaule P, Axel R (1985) A novel ras-related gene family. *Cell* **41**: 31-40

Madigan MT, Martinko JM, Brock TD (2006) *Brock Biology of microorganisms*, 11th edn. Upper Saddle River, NJ: Pearson Prentice Hall.

Maegley KA, Admiraal SJ, Herschlag D (1996) Ras-catalyzed hydrolysis of GTP: a new perspective from model studies. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 8160-8166

Manning AM, Davis RJ (2003) Targeting JNK for therapeutic benefit: from junk to gold? *Nature reviews Drug discovery* **2**: 554-565

Mansour SJ, Resing KA, Candi JM, Hermann AS, Gloor JW, Herskind KR, Wartmann M, Davis RJ, Ahn NG (1994) Mitogen-activated protein (MAP) kinase phosphorylation of MAP kinase kinase: determination of phosphorylation sites by mass spectrometry and site-directed mutagenesis. *Journal of biochemistry* **116**: 304-314

Matsumoto T, Kinoshita T, Kirii Y, Yokota K, Hamada K, Tada T (2010) Crystal structures of MKK4 kinase domain reveal that substrate peptide binds to an allosteric site and induces an auto-inhibition state. *Biochemical and biophysical research communications* **400**: 369-373

Melander LCS, Saunders WH (1980) *Reaction rates of isotopic molecules*, 2nd edn. New York &c.,

Meller N, Merlot S, Guda C (2005) CZH proteins: a new family of Rho-GEFs. *Journal of cell science* **118**: 4937-4946

Mesak LR, Dahl MK (2000) Purification and enzymatic characterization of PgcM: a beta-phosphoglucomutase and glucose-1-phosphate phosphodismutase of Bacillus subtilis. *Archives of microbiology* **174**: 256-264

Milburn MV, Tong L, Devos AM, Brunger A, Yamaizumi Z, Nishimura S, Kim SH (1990) Molecular Switch for Signal Transduction - Structural Differences between Active and Inactive Forms of Protooncogenic Ras Proteins. *Science* **247**: 939-945

Minton AP (2006) How can biochemical reactions within cells differ from those in test tubes? *Journal of cell science* **119**: 2863-2869

- More O'Ferrall RA (1970) Relationships between E2 and E1cB mechanisms of β -elimination. *JChem Soc B*: 274-277
- Moriarty NW, Grosse-Kunstleve RW, Adams PD (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta crystallographica Section D, Biological crystallography* **65**: 1074-1080
- Muller K, Faeh C, Diederich F (2007) Fluorine in pharmaceuticals: looking beyond intuition. *Science* **317**: 1881-1886
- Mulloy JC, Cancelas JA, Filippi MD, Kalfa TA, Guo F, Zheng Y (2010) Rho GTPases in hematopoiesis and hemopathies. *Blood* **115**: 936-947
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica Section D, Biological crystallography* **67**: 355-367
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255
- Nadif Kasri N, Van Aelst L (2008) Rho-linked genes and neurological disorders. *Pflugers Archiv : European journal of physiology* **455**: 787-797
- Nassar N, Hoffman GR, Manor D, Clardy JC, Cerione RA (1998) Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nature structural biology* **5**: 1047-1052
- Nelson DL, Cox MM, Lehninger AL (2008) *Lehninger principles of biochemistry*, 5th edn. New York ; Basingstoke: W.H. Freeman.
- Nilsson U, Radstrom P (2001) Genetic localization and regulation of the maltose phosphorylase gene, malP, in *Lactococcus lactis*. *Microbiology* **147**: 1565-1573
- Nishida E, Gotoh Y (1993) The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem Sci* **18**: 128-131
- Nossaman BD, Kadowitz PJ (2009) The role of the RhoA/rho-kinase pathway in pulmonary hypertension. *Current drug discovery technologies* **6**: 59-71
- O'Brien LE, Jou TS, Pollack AL, Zhang Q, Hansen SH, Yurchenco P, Mostov KE (2001) Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. *Nature cell biology* **3**: 831-838
- O'Brien PJ, Herschlag D (2002) Alkaline phosphatase revisited: hydrolysis of alkyl phosphates. *Biochemistry* **41**: 3207-3225
- OHagan D, Rzepa HS (1997) Some influences of fluorine in bioorganic chemistry. *Chem Commun*: 645-652
- Olesen C, Picard M, Winther AM, Gyruup C, Morth JP, Oxvig C, Moller JV, Nissen P (2007) The structural basis of calcium transport by the calcium pump. *Nature* **450**: 1036-1042
- Olson MF, Ashworth A, Hall A (1995) An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science* **269**: 1270-1272

- Olson MF, Pasteris NG, Gorski JL, Hall A (1996) Faciogenital dysplasia protein (FGD1) and Vav, two related proteins required for normal embryonic development, are upstream regulators of Rho GTPases. *Current biology : CB* **6**: 1628-1633
- Olson MF, Paterson HF, Marshall CJ (1998) Signals from Ras and Rho GTPases interact to regulate expression of p21Waf1/Cip1. *Nature* **394**: 295-299
- Opperdoes FR (1987) Compartmentation of carbohydrate metabolism in trypanosomes. *Annual review of microbiology* **41**: 127-151
- Oruganty K, Kannan N (2012) Design principles underpinning the regulatory diversity of protein kinases. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **367**: 2529-2539
- Ostanin K, Van Etten RL (1993) Asp304 of Escherichia coli acid phosphatase is involved in leaving group protonation. *The Journal of biological chemistry* **268**: 20778-20784
- Ozdinler PH, Erzurumlu RS (2001) Regulation of neurotrophin-induced axonal responses via Rho GTPases. *The Journal of comparative neurology* **438**: 377-387
- Pai SY, Kim C, Williams DA (2010) Rac GTPases in human diseases. *Disease markers* **29**: 177-187
- Paithankar KS, Garman EF (2010) Know your dose: RADDOS. *Acta crystallographica Section D, Biological crystallography* **66**: 381-388
- Pan X, Eathiraj S, Munson M, Lambright DG (2006) TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature* **442**: 303-306
- Park BK, Kitteringham NR, O'Neill PM (2001) Metabolism of fluorine-containing drugs. *Annual review of pharmacology and toxicology* **41**: 443-470
- Park HO, Bi E (2007) Central roles of small GTPases in the development of cell polarity in yeast and beyond. *Microbiology and molecular biology reviews : MMBR* **71**: 48-96
- Parri M, Chiarugi P (2010) Rac and Rho GTPases in cancer cell motility control. *Cell communication and signaling : CCS* **8**: 23
- Patel SB, Cameron PM, Frantz-Wattley B, O'Neill E, Becker JW, Scapin G (2004) Lattice stabilization and enhanced diffraction in human p38 alpha crystals by protein engineering. *Biochimica et biophysica acta* **1696**: 67-73
- Patrick D. Shaw Stewart SAK, Richard A. Briggs , Naomi E. Chayen , and Peter F. M. Baldock (2011) Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization. *Cryst Growth Des* **11**: 3432-3441
- Pauling L (1948) Nature of forces between large molecules of biological interest. *Nature* **161**: 107-709
- Pauling L (1960) The Resonating-Valence-Bond Theory of Metals. *Acta Crystallogr* **13**: 983-984
- Paulini R, Muller K, Diederich F (2005) Orthogonal multipolar interactions in structural chemistry and biology. *Angewandte Chemie* **44**: 1788-1805
- Payne DM, Rossomando AJ, Martino P, Erickson AK, Her JH, Shabanowitz J, Hunt DF, Weber MJ, Sturgill TW (1991) Identification of the regulatory phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase). *The EMBO journal* **10**: 885-892

- Pellegrini E, Piano D, Bowler MW (2011) Direct cryocooling of naked crystals: are cryoprotection agents always necessary? *Acta crystallographica Section D, Biological crystallography* **67**: 902-906
- Penning TD, Talley JJ, Bertenshaw SR, Carter JS, Collins PW, Docter S, Graneto MJ, Lee LF, Malecha JW, Miyashiro JM, Rogers RS, Rogier DJ, Yu SS, Anderson Gd, Burton EG, Cogburn JN, Gregory SA, Koboldt CM, Perkins WE, Seibert K, Veenhuizen AW, Zhang YY, Isakson PC (1997) Synthesis and biological evaluation of the 1,5-diarylpyrazole class of cyclooxygenase-2 inhibitors: identification of 4-[5-(4-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benzenesulfonamide (SC-58635, celecoxib). *Journal of medicinal chemistry* **40**: 1347-1365
- Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nature structural biology* **6**: 458-463
- Pierce KL, Lefkowitz RJ (2001) Classical and new roles of beta-arrestins in the regulation of G-protein-coupled receptors. *Nature reviews Neuroscience* **2**: 727-733
- Pinxteren JA, O'Sullivan AJ, Larbi KY, Tatham PE, Gomperts BD (2000) Thirty years of stimulus-secretion coupling: from Ca(2+) to GTP in the regulation of exocytosis. *Biochimie* **82**: 385-393
- Poornam GP, Matsumoto A, Ishida H, Hayward S (2009) A method for the analysis of domain movements in large biomolecular complexes. *Proteins* **76**: 201-212
- Popov AN, Bourenkov GP (2003) Choice of data-collection parameters based on statistic modelling. *Acta crystallographica Section D, Biological crystallography* **59**: 1145-1153
- Pruyne D, Bretscher A (2000) Polarization of cell growth in yeast. *Journal of cell science* **113 (Pt 4)**: 571-585
- Pulido R, Zuniga A, Ullrich A (1998) PTP-SL and STEP protein tyrosine phosphatases regulate the activation of the extracellular signal-regulated kinases ERK1 and ERK2 by association through a kinase interaction motif. *The EMBO journal* **17**: 7337-7350
- Purser S, Moore PR, Swallow S, Gouverneur V (2008) Fluorine in medicinal chemistry. *Chemical Society reviews* **37**: 320-330
- Qian N, Stanley GA, Bunte A, Radstrom P (1997) Product formation and phosphoglucomutase activities in *Lactococcus lactis*: cloning and characterization of a novel phosphoglucomutase gene. *Microbiology* **143 (Pt 3)**: 855-865
- Qian N, Stanley GA, Hahn-Hagerdal B, Radstrom P (1994) Purification and characterization of two phosphoglucomutases from *Lactococcus lactis* subsp. *lactis* and their regulation in maltose- and glucose-utilizing cells. *Journal of bacteriology* **176**: 5304-5311
- Rahlfs S, Koncarevic S, Iozef R, Mailu BM, Savvides SN, Schirmer RH, Becker K (2009) Myristoylated adenylate kinase-2 of *Plasmodium falciparum* forms a heterodimer with myristoyltransferase. *Molecular and biochemical parasitology* **163**: 77-84
- Raich WB, Agbunag C, Hardin J (1999) Rapid epithelial-sheet sealing in the *Caenorhabditis elegans* embryo requires cadherin-dependent filopodial priming. *Current biology : CB* **9**: 1139-1146
- Rall TW, Sutherland EW (1958) Formation of a cyclic adenine ribonucleotide by tissue particles. *The Journal of biological chemistry* **232**: 1065-1076

- Ramos A, Boels IC, de Vos WM, Santos H (2001) Relationship between glycolysis and exopolysaccharide biosynthesis in *Lactococcus lactis*. *Applied and environmental microbiology* **67**: 33-41
- Read RJaS, A.J. (1988) A phased translational function. *Journal of Applied Crystallography* **21**: 490-495
- Regan J, Capolino A, Cirillo PF, Gilmore T, Graham AG, Hickey E, Kroe RR, Madwed J, Moriak M, Nelson R, Pargellis CA, Swinamer A, Torcellini C, Tsang M, Moss N (2003) Structure-activity relationships of the p38alpha MAP kinase inhibitor 1-(5-tert-butyl-2-p-tolyl-2H-pyrazol-3-yl)-3-[4-(2-morpholin-4-yl-ethoxy)naphthalen-1-yl]urea (BIRB 796). *Journal of medicinal chemistry* **46**: 4676-4686
- Reid TW, Wilson IB (1971) Conformational isomers of alkaline phosphatase in the mechanism of hydrolysis. *Biochemistry* **10**: 380-387
- Remenyi A, Good MC, Bhattacharyya RP, Lim WA (2005) The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. *Molecular cell* **20**: 951-962
- Resnick L, Fennell M (2004) Targeting JNK3 for the treatment of neurodegenerative disorders. *Drug discovery today* **9**: 932-939
- Rittinger K, Walker PA, Eccleston JF, Smerdon SJ, Gamblin SJ (1997) Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* **389**: 758-762
- Robbins DJ, Cobb MH (1992) Extracellular signal-regulated kinases 2 autophosphorylates on a subset of peptides phosphorylated in intact cells in response to insulin and nerve growth factor: analysis by peptide mapping. *Molecular biology of the cell* **3**: 299-308
- Robbins DJ, Zhen E, Owaki H, Vanderbilt CA, Ebert D, Geppert TD, Cobb MH (1993) Regulation and properties of extracellular signal-regulated protein kinases 1 and 2 in vitro. *The Journal of biological chemistry* **268**: 5097-5106
- Roberts PJ, Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**: 3291-3310
- Robinson MJ, Harkins PC, Zhang J, Baer R, Haycock JW, Cobb MH, Goldsmith EJ (1996) Mutation of position 52 in ERK2 creates a nonproductive binding mode for adenosine 5'-triphosphate. *Biochemistry* **35**: 5641-5646
- Rojas AM, Fuentes G, Rausell A, Valencia A (2012) The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *The Journal of cell biology* **196**: 189-201
- Romanenko VD, Kukhar VP (2006) Fluorinated phosphonates: synthesis and biomedical application. *Chemical reviews* **106**: 3868-3935
- Ronimus RS, Morgan HW (2004) Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochemical and biophysical research communications* **315**: 652-658
- Rossmann MG (1990) The molecular replacement method. *Acta crystallographica Section A, Foundations of crystallography* **46 (Pt 2)**: 73-82
- Russi S, Juers DH, Sanchez-Weatherby J, Pellegrini E, Mossou E, Forsyth VT, Huet J, Gobbo A, Felisaz F, Moya R, McSweeney SM, Cusack S, Cipriani F, Bowler MW (2011) Inducing phase

- changes in crystals of macromolecules: status and perspectives for controlled crystal dehydration. *Journal of structural biology* **175**: 236-243
- Sakurada S, Okamoto H, Takuwa N, Sugimoto N, Takuwa Y (2001) Rho activation in excitatory agonist-stimulated vascular smooth muscle. *American journal of physiology Cell physiology* **281**: C571-578
- Sanchez-Weatherby J, Bowler MW, Huet J, Gobbo A, Felisaz F, Lavault B, Moya R, Kadlec J, Ravelli RB, Cipriani F (2009) Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta crystallographica Section D, Biological crystallography* **65**: 1237-1246
- Saxena M, Mustelin T (2000) Extracellular signals and scores of phosphatases: all roads lead to MAP kinase. *Seminars in immunology* **12**: 387-396
- Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, Wittinghofer A (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**: 333-338
- Schindelin H, Kisker C, Schlessman JL, Howard JB, Rees DC (1997) Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. *Nature* **387**: 370-376
- Schindler JF, Monahan JB, Smith WG (2007) p38 pathway kinases as anti-inflammatory drug targets. *Journal of dental research* **86**: 800-811
- Schlessinger K, Hall A, Tolwinski N (2009) Wnt signaling pathways meet Rho GTPases. *Genes & development* **23**: 265-277
- Schlichting I, Reinstein J (1997) Structures of active conformations of UMP kinase from *Dictyostelium discoideum* suggest phosphoryl transfer is associative. *Biochemistry* **36**: 9290-9296
- Schlichting I, Reinstein J (1999) pH influences fluoride coordination number of the AIF_x phosphoryl transfer transition state analog. *Nature structural biology* **6**: 721-723
- Schmitz AA, Govek EE, Bottner B, Van Aelst L (2000) Rho GTPases: signaling, migration, and invasion. *Experimental cell research* **261**: 1-12
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671-675
- Schroeder GK, Lad C, Wyman P, Williams NH, Wolfenden R (2006) The time required for water attack at the phosphorus atom of simple phosphodiester and of DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 4052-4055
- Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica Section D, Biological crystallography* **60**: 1355-1363
- Scott A, Haystead CM, Haystead TA (1995) Purification of a 12,020-dalton protein that enhances the activation of mitogen-activated protein (MAP) kinase by MAP kinase kinase. *The Journal of biological chemistry* **270**: 24540-24547
- Scrima A, Thomas C, Deaconescu D, Wittinghofer A (2008) The Rap-RapGAP complex: GTP hydrolysis without catalytic glutamine and arginine residues. *The EMBO journal* **27**: 1145-1153
- Sebolt-Leopold JS, Herrera R (2004) Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nature reviews Cancer* **4**: 937-947

- Seewald MJ, Korner C, Wittinghofer A, Vetter IR (2002) RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* **415**: 662-666
- Senior AE, Nadanaciva S, Weber J (2002) The molecular mechanism of ATP synthesis by F1F0-ATP synthase. *Biochimica et biophysica acta* **1553**: 188-211
- Sharrocks AD, Yang SH, Galanis A (2000) Docking domains and substrate-specificity determination for MAP kinases. *Trends in biochemical sciences* **25**: 448-453
- Shibata N, Sato H, Sakaki S, Sugita Y (2011) Theoretical study of magnesium fluoride in aqueous solution. *The journal of physical chemistry B* **115**: 10553-10559
- Shimokawa H (2002) Rho-kinase as a novel therapeutic target in treatment of cardiovascular diseases. *Journal of cardiovascular pharmacology* **39**: 319-327
- Shimokawa H, Hiramori K, Inuma H, Hosoda S, Kishida H, Osada H, Katagiri T, Yamauchi K, Yui Y, Minamino T, Nakashima M, Kato K (2002) Anti-anginal effect of fasudil, a Rho-kinase inhibitor, in patients with stable effort angina: a multicenter study. *Journal of cardiovascular pharmacology* **40**: 751-761
- Skoog MT, Jencks WP (1983) Phosphoryl transfer between pyridines. *J. Am. Chem. Soc.* 1983; 105:3356-57. *J Am Chem Soc* **105**: 3356-3357
- Skoog MT, Jencks WP (1984) Reactions of pyridines and primary amines with N-phosphorylated pyridines. *J Am Chem Soc* **106**
- Small JV, Stradal T, Vignal E, Rottner K (2002) The lamellipodium: where motility begins. *Trends in cell biology* **12**: 112-120
- Smart BE (2001) Fluorine substituent effects (on bioactivity). *J Fluorine Chem* **109**: 3-11
- Smith CA, Rayment I (1996) X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9 Å resolution. *Biochemistry* **35**: 5404-5417
- Smith JA, Poteet-Smith CE, Lannigan DA, Freed TA, Zoltoski AJ, Sturgill TW (2000) Creation of a stress-activated p90 ribosomal S6 kinase. The carboxyl-terminal tail of the MAPK-activated protein kinases dictates the signal transduction pathway in which they function. *The Journal of biological chemistry* **275**: 31588-31593
- Smith JA, Poteet-Smith CE, Malarkey K, Sturgill TW (1999) Identification of an extracellular signal-regulated kinase (ERK) docking site in ribosomal S6 kinase, a sequence critical for activation by ERK in vivo. *The Journal of biological chemistry* **274**: 2893-2898
- Snyder MA, Bishop JM, McGrath JP, Levinson AD (1985) A mutation at the ATP-binding site of pp60^{v-src} abolishes kinase activity, transformation, and tumorigenicity. *Molecular and cellular biology* **5**: 1772-1779
- Sondek J, Lambright DG, Noel JP, Hamm HE, Sigler PB (1994) GTPase mechanism of Gproteins from the 1.7-Å crystal structure of transducin alpha-GDP-AIF-4. *Nature* **372**: 276-279
- Stec B, Holtz KM, Kantrowitz ER (2000) A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *Journal of molecular biology* **299**: 1303-1311

- Sternweis PC, Gilman AG (1982) Aluminum: a requirement for activation of the regulatory component of adenylate cyclase by fluoride. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 4888-4891
- Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. *Annual review of biochemistry* **69**: 183-215
- Stronach BE, Perrimon N (1999) Stress signaling in Drosophila. *Oncogene* **18**: 6172-6182
- Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 2473-2478
- Tanoue T, Adachi M, Moriguchi T, Nishida E (2000) A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nature cell biology* **2**: 110-116
- Tanoue T, Maeda R, Adachi M, Nishida E (2001) Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *The EMBO journal* **20**: 466-479
- Tari LW, Matte A, Goldie H, Delbaere LT (1997) Mg(2+)-Mn2+ clusters in enzyme-catalyzed phosphoryl-transfer reactions. *Nature structural biology* **4**: 990-994
- Tari LW, Matte A, Pugazhenth U, Goldie H, Delbaere LT (1996) Snapshot of an enzyme reaction intermediate in the structure of the ATP-Mg2+-oxalate ternary complex of Escherichia coli PEP carboxykinase. *Nature structural biology* **3**: 355-363
- Ten Eyck LF (1985) Fast Fourier transform calculation of electron density maps. *Methods in enzymology* **115**: 324-337
- Teng T-Y (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J ppl Cryst* **23**: 387-391
- ter Haar E, Prabhakar P, Liu X, Lepre C (2007) Crystal structure of the p38 alpha-MAPKAP kinase 2 heterodimer. *The Journal of biological chemistry* **282**: 9733-9739
- Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica Section D, Biological crystallography* **65**: 582-601
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**: 4673-4680
- Tremblay LW, Zhang G, Dai J, Dunaway-Mariano D, Allen KN (2005) Chemical confirmation of a pentavalent phosphorane in complex with beta-phosphoglucomutase. *Journal of the American Chemical Society* **127**: 5298-5299
- Urzhumtseva L, Afonine PV, Adams PD, Urzhumtsev A (2009) Crystallographic model quality at a glance. *Acta crystallographica Section D, Biological crystallography* **65**: 297-300
- Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**: 22-25

- Van Heek M, France CF, Compton DS, McLeod RL, Yumibe NP, Alton KB, Sybertz EJ, Davis HR, Jr. (1997) In vivo metabolism-based discovery of a potent cholesterol absorption inhibitor, SCH58235, in the rat and rhesus monkey through the identification of the active metabolites of SCH48461. *The Journal of pharmacology and experimental therapeutics* **283**: 157-163
- Vasioukhin V, Bauer C, Yin M, Fuchs E (2000) Directed actin polymerization is the driving force for epithelial cell-cell adhesion. *Cell* **100**: 209-219
- Vega FM, Ridley AJ (2008) Rho GTPases in cancer cell biology. *FEBS letters* **582**: 2093-2101
- Vetter IR, Wittinghofer A (2001) Signal transduction - The guanine nucleotide-binding switch in three dimensions. *Science* **294**: 1299-1304
- Vulpetti A, Dalvit C (2012) Fluorine local environment: from screening to drug design. *Drug discovery today* **17**: 890-897
- Waas WF, Lo HH, Dalby KN (2001) The kinetic mechanism of the dual phosphorylation of the ATF2 transcription factor by p38 mitogen-activated protein (MAP) kinase alpha. Implications for signal/response profiles of MAP kinase pathways. *The Journal of biological chemistry* **276**: 5676-5684
- Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**: 945-951
- Wang Z, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, Abdel-Meguid S, Adams JL, Goldsmith EJ (1998) Structural basis of inhibitor selectivity in MAP kinases. *Structure* **6**: 1117-1128
- Wang ZL, Harkins PC, Ulevitch RJ, Han JH, Cobb MH, Goldsmith EJ (1997) The structure of mitogen-activated protein kinase p38 at 2.1-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 2327-2332
- Watanabe N, Kato T, Fujita A, Ishizaki T, Narumiya S (1999) Cooperation between mDia1 and ROCK in Rho-induced actin reorganization. *Nature cell biology* **1**: 136-143
- Watanabe N, Madaule P, Reid T, Ishizaki T, Watanabe G, Kakizuka A, Saito Y, Nakao K, Jockusch BM, Narumiya S (1997) p140mDia, a mammalian homolog of Drosophila diaphanous, is a target protein for Rho small GTPase and is a ligand for profilin. *The EMBO journal* **16**: 3044-3056
- Webster CE (2004) High-energy intermediate or stable transition state analogue: theoretical perspective of the active site and mechanism of beta-phosphoglucomutase. *Journal of the American Chemical Society* **126**: 6840-6841
- Wei Y, Zhang Y, Derewenda U, Liu X, Minor W, Nakamoto RK, Somlyo AV, Somlyo AP, Derewenda ZS (1997) Crystal structure of RhoA-GDP and its functional implications. *Nature structural biology* **4**: 699-703
- Weiss MS, Sicker T, Djinovic-Carugo K, Hilgenfeld R (2001) On the routine use of soft X-rays in macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* **57**: 689-695
- Welsh CF, Roovers K, Villanueva J, Liu Y, Schwartz MA, Assoian RK (2001) Timing of cyclin D1 expression within G1 phase is controlled by Rho. *Nature cell biology* **3**: 950-957

- Wennerberg K, Der CJ (2004) Rho-family GTPases: it's not only Rac and Rho (and I like it). *Journal of cell science* **117**: 1301-1312
- Wennerberg K, Rossman KL, Der CJ (2005) The Ras superfamily at a glance. *Journal of cell science* **118**: 843-846
- Westheimer FH (1987) Why nature chose phosphates. *Science* **235**: 1173-1178
- Williams NH (2004) Models for biological phosphoryl transfer. *Biochimica et biophysica acta* **1697**: 279-287
- Wilson KP, Fitzgibbon MJ, Caron PN, Griffith JP, Chen WY, McCaffrey PG, Chambers SP, Su MSS (1996a) Crystal structure of p38 mitogen-activated protein kinase. *Journal of Biological Chemistry* **271**: 27696-27700
- Wilson KP, Fitzgibbon MJ, Caron PR, Griffith JP, Chen W, McCaffrey PG, Chambers SP, Su MS (1996b) Crystal structure of p38 mitogen-activated protein kinase. *J Biol Chem* **271**: 27696-27700
- Wittinghofer A (1997) Signaling mechanistics: aluminum fluoride for molecule of the year. *Current biology : CB* **7**: R682-685
- Wittinghofer A, Pai, E. & Goody, R. S. (1993) Structural and mechanistic aspects of the GTPase reaction of H-ras p21. *Handbk Exp Pharmacol* **108**: 195-212
- Wojciak-Stothard B, Potempa S, Eichholtz T, Ridley AJ (2001) Rho and Rac but not Cdc42 regulate endothelial cell permeability. *Journal of cell science* **114**: 1343-1355
- Wolfenden R (1969) Transition state analogues for enzyme catalysis. *Nature* **223**: 704-705
- Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Accounts of chemical research* **34**: 938-945
- Wong WT, Faulkner-Jones BE, Sanes JR, Wong RO (2000) Rapid dendritic remodeling in the developing retina: dependence on neurotransmission and reciprocal regulation by Rac and Rho. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20**: 5024-5036
- Worby CA, Mattoo S, Kruger RP, Corbeil LB, Koller A, Mendez JC, Zekarias B, Lazar C, Dixon JE (2009) The fic domain: regulation of cell signaling by adenylylation. *Molecular cell* **34**: 93-103
- Worthylake RA, Lemoine S, Watson JM, Burrige K (2001) RhoA is required for monocyte tail retraction during transendothelial migration. *The Journal of cell biology* **154**: 147-160
- Wroblewski ST, Doweyko AM (2005) Structural comparison of p38 inhibitor-protein complexes: a review of recent p38 inhibitors having unique binding interactions. *Current topics in medicinal chemistry* **5**: 1005-1016
- Xiaoxia L, Marston JP, Baxter NJ, Hounslow AM, Yufen Z, Blackburn GM, Cliff MJ, Waltho JP (2011) Prioritization of charge over geometry in transition state analogues of a dual specificity protein kinase. *Journal of the American Chemical Society* **133**: 3989-3994
- Xu B, Wilsbacher JL, Collisson T, Cobb MH (1999) The N-terminal ERK-binding site of MEK1 is required for efficient feedback phosphorylation by ERK2 in vitro and ERK activation in vivo. *The Journal of biological chemistry* **274**: 34029-34035

Xu Y, Lee SA, Kutateladze TG, Sbrissa D, Shisheva A, Prestwich GD (2006) Chemical synthesis and molecular recognition of phosphatase-resistant analogues of phosphatidylinositol-3-phosphate. *Journal of the American Chemical Society* **128**: 885-897

Xue W, Krasnitz A, Lucito R, Sordella R, Vanaelst L, Cordon-Cardo C, Singer S, Kuehnel F, Wigler M, Powers S, Zender L, Lowe SW (2008) DLC1 is a chromosome 8p tumor suppressor whose loss promotes hepatocellular carcinoma. *Genes & development* **22**: 1439-1444

Yarbrough ML, Li Y, Kinch LN, Grishin NV, Ball HL, Orth K (2009) AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* **323**: 269-272

Yuan BZ, Durkin ME, Popescu NC (2003) Promoter hypermethylation of DLC-1, a candidate tumor suppressor gene, in several common human cancers. *Cancer genetics and cytogenetics* **140**: 113-117

Zerial M, McBride H (2001) Rab proteins as membrane organizers. *Nature reviews Molecular cell biology* **2**: 107-117

Zhang G, Dai J, Wang L, Dunaway-Mariano D, Tremblay LW, Allen KN (2005) Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry* **44**: 9404-9416

Zhang YY, Wu JW, Wang ZX (2011) A distinct interaction mode revealed by the crystal structure of the kinase p38alpha with the MAPK binding domain of the phosphatase MKP5. *Science signaling* **4**: ra88

Zhou P, Zou J, Tian F, Shang Z (2009) Fluorine bonding--how does it work in protein-ligand interactions? *Journal of chemical information and modeling* **49**: 2344-2355

Zhou Q, Gensch C, Liao JK (2011) Rho-associated coiled-coil-forming kinases (ROCKs): potential targets for the treatment of atherosclerosis and vascular disease. *Trends in pharmacological sciences* **32**: 167-173

Zhou T, Sun L, Humphreys J, Goldsmith EJ (2006) Docking interactions induce exposure of activation loop in the MAP kinase ERK2. *Structure* **14**: 1011-1019

Zuniga A, Torres J, Ubeda J, Pulido R (1999) Interaction of mitogen-activated protein kinases with the kinase interaction motif of the tyrosine phosphatase PTP-SL provides substrate specificity and retains ERK2 in the cytoplasm. *Journal of Biological Chemistry* **274**: 21900-21907