Synthetic Phosphorylation of Kinases for Functional Studies *in vitro*

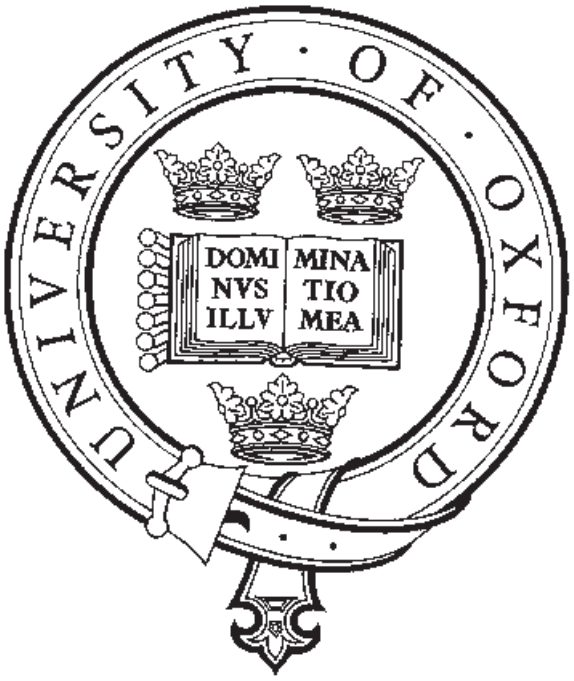
Kok Phin Chooi

October 2014

Supervisors: Prof. Benjamin Davis (Department of Chemistry, Oxford)
and Dr. Lyn Jones (Pfizer, Cambridge, MA, USA)

A thesis submitted to the Board of the Faulty of Physical Sciences in partial fulfilment for the degree of

Doctor of Philosophy (D.Phil)
at the
University of Oxford



Pembroke College, University of Oxford

Author's Declaration

I declare the work presented in this thesis was carried out at the Chemistry Research Laboratory and Inorganic Chemistry Laboratory, University of Oxford, UK under the guidance of Professor Ben Davis. All the work is my own, except where otherwise stated, and has not been submitted for any other degree at this or any other university.

Kok Phin Chooi

October 2014

Short Abstract

The activity of protein kinases is heavily dependent on the phosphorylation state of the protein. Kinase phosphorylation states have been prepared through biological or enzymatic means for biochemical evaluation, but the use of protein chemical modification as an investigative tool has not been addressed. By chemically reacting a genetically encoded cysteine, phosphocysteine was installed via dehydroalanine as a reactive intermediate. The installed phosphocysteine was intended as a surrogate to the naturally occurring phosphothreonine or phosphoserine of a phosphorylated protein kinase. Two model protein kinases were investigated on: MEK1 and p38α. The development of suitable protein variants and suitable reaction conditions on these two proteins is discussed in turn and in detail, resulting in p38α-pCys180 and MEK1-pCys222. Designed to be mimics of the naturally occurring p38α-pThr180 and MEK1-pSer222, these two chemically modified proteins were studied for their biological function.

The core biological studies entailed the determination of enzymatic activity of both modified proteins, and included the necessary controls against their active counterparts. In addition, the studies on p38α-pCys180 also included a more detailed quantification of enzymatic activity, and the behaviour of this modified protein against known inhibitors of p38α was also investigated. Both modified proteins were shown to be enzymatically active and behave similarly to corresponding active species.

The adaptation of mass spectrometry methods to handle the majority of project's analytical requirements, from monitoring chemical transformations to following enzyme kinetics was instrumental in making these studies feasible. The details of these technical developments are interwoven into the scientific discussion. Also included in this thesis is an introduction to the mechanism and function of protein kinases, and on the protein chemistry methods employed. The work is concluded with a projection of implications that this protein chemical modification technique has on kinase biomedical research.

Abbreviations

0.1 General Abbreviations

ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATF2	Activating Transcription Factor 2
ATP	adenosine triphosphate
β-mercaptoethanol	β-ME
CAM	*S*-carboxyamidomethylcysteine
CID	collision-induced dissociation
CMC	critical micelle concentration
CD	circular dichroism
CV	column volume
Dha	dehydroalanine
DMF	dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DSF	differential scanning fluorimetry
DTC	Doctoral Training Centre

DTT	DL-dithiothreitol
EGTA	ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetra-acetic acid
EM	electromagnetic
ERA	electrophoretic radioassay
ERK	extracellular-signal regulated kinase
ES(I)	electro-spray (ionisation)
ETD	electron-transfer dissociation
FPLC	Fast Protein Liquid Chromatography
GST	glutathione S-transferase
h	hour(s)
HCD	high-energy collision dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazine-1-ethanesulphonic acid
HePTP	hematopoietic protein tyrosine phosphatase
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
(k)bp	(kilo-)base pair
LB	lysogeny broth
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
MALDI	matrix-assisted laser desorption/ionisation
MAPK	mitogen-activated protein kinase
MEK	MAP or ERK kinase
MKK	MAP kinase kinase

MOPS	3-((N-morpholino)propanesulphonic acid
MS	mass spectrometry
MWCO	molecular weight cut-off
NAD	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
OD	optical density
PCR	polymerase chain reaction
pCys	phosphocysteine
PEP	phosphoenol pyruvate
ppm	parts per million
pSer	phosphoserine
pThr	phosphothreonine
pTyr	phosphotyrosine
rpm	revolutions per minute
RT	room temperature
sat.	saturated
SB	Super Broth
SDS	sodium dodecyl sulphate
SGC	Structural Genomics Consortium
SOC	Super Optimal broth with Catabolite repression
TB	Terrific Broth
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TIC	total ion current
UV	ultra-violet

WT	wild-type
YT	yeast and tryptone (nutrient medium)
2YT	2× yeast and tryptone (nutrient medium)

0.2 Amino Acids

Ala , A	alanine
Arg , R	arginine
Asn , N	asparagine
Asp , D	aspartic acid
Cys , C	cysteine
Gln , Q	glutamine
Glu , E	glutamic acid
Gly , G	glycine
His , H	histidine
Ile , I	isoleucine
Leu , L	leucine
Lys , K	lysine
Met , M	methionine
Phe , F	phenylalanine
Pro , P	proline
Ser , S	serine
Thr , T	threonine
Trp , W	tryptophan
Tyr , Y	tyrosine
Val , V	valine

0.3 Nucleic Acid Bases

A	adenine
C	cytosine
G	guanine
T	thymine

0.4 Definition of Terms

Chemical variant: A protein variant which has undergone chemical modification reaction(s). When described in a collective, this term can also refer to the unmodified protein variant species.

Acknowledgements

I would firstly like to thank Prof. Ben Davis and Dr. Lyn Jones for formulating the project and for giving me the opportunity to work on it, without which none of the following work described here would have happened.

There are a number of group members, past and present whom I would like to thank for their particular contributions to my time in the group: Dr. Seung Seo Lee and Keisuke Yamamoto for their expertise in molecular biology techniques, which gave me a firm start in the lab's biological setup; Smita Gunnoo and Justin Chalker as my predecessors in protein modification chemistry; Dr. Matt Schombs for being my Wingman on all things technical, and for being a great mentor and friend; Dr. Bruce Lichtenstein as a partner in discussion, always willing to listen to my musings and to discuss them with me critically; Sébastien Galan and Ritu Raj as the other Group Members of "Team Phin-possible" for the successful completion of the mission to write *Synthetic Phosphorylation of p38α Recapitulates Protein Kinase Activity*;[1] Marcel Scheepstra for being a great MSci student whom I have had the opportunity to train and work with; Amy Webb for always being a welcoming face to all of the lab's day-to-day necessities; Tom Taylor, Smita Gunnoo, Tom Wright, Chris Spicer and Sébastien Galan for helping me in proofing this thesis; and finally for everyone I will have missed, all past and present BGD Group Members, and more particularly, past and present lab members of LG8, Chemistry Research Laboratory, Department of Chemistry.

Outside the group, I would like to thank the Doctoral Training Centre for providing me with all the training I have felt that I ever wanted as a scientific researcher; Dr. Jon Elkins and Leila Alexander at the Structural Genomics Consortium, Oxford for providing me with useful knowledge, advice and (quite a few) plasmids for all things kinase; and Dr. Kathryn Scott, Dr. Nelly Dubarry and Gareth Blades for going out of their way to train me to use their radioisotope facility in Prof. Judy Armitage Lab, Department of Biochemistry.

This project was financially supported by the EPSRC and further supported by Pfizer Ltd.

Finally, special thanks goes to my brother, Kok Yean Chooi, for always being there for me, even in the toughest of times and similarly, I would like to thank my Mum and Dad for the moral and mental support they have given me over the years.

Abstract

Protein kinases play an essential role in cellular regulation and signalling. A keen interest in this class of proteins arises due to these cellular roles, as dysregulation can result in disease. The study of their function is therefore directly applicable to the development of new methods of therapy. In this context, the use of protein chemical modification as a tool for such functional studies is demonstrated here.

In Chapter 1, a brief outline of the purpose, function and mechanism of protein kinases as enzymes is given. The concept of protein chemical modification, particularly as applied to kinases is introduced, along with the rationale behind making these modifications. Finally, the two protein kinases investigated in this study are also introduced, namely p38α and MEK1.

The work on p38α is described first. In Chapter 2, the rationale behind the design of the variants used in the study is given, followed by the methods used to produce these protein variants, namely p38α-C119S/C162S/A172C (referred to as p38α-Cys172) and p38α-C119S/C162S/T180C (referred to as p38α-Cys180). The efforts in optimising the chemical modification on these p38α variants is then described, culminating in the production of chemically phosphorylated p38α (both p38α-pCys172 and p38α-pCys180) of high purity. This chapter therefore constitutes how the chemically modified p38α was produced, allowing for further functional studies *in vitro*.

Chapter 3 describes the functional studies that were carried out on the chemically modified p38α. The efforts to test the activity of all the chemically modified p38α variants is described first. Since it was found that only p38α-pCys180 out of all the chemically modified p38α variants produced was active, further efforts towards quantifying its activity are then detailed for this chemical variant only. Corresponding experiments on the necessary wild-type controls (active and inactive wild-type p38α) are also detailed. A further test of biological recognition, functional studies were further extended to look at the action of inhibitors on p38α-pCys180, again comparing the effects to those on the enzymatically activated wild-

type p38α. The development of the methods used to perform these studies is described during the course of discussion in this section, and is also a subject discussed in the course of the previous chapter (Chapter 2). Of these methods, the semi-automated mass spectrometry quantification methods used are highlighted. These descriptions of methods are also applicable to the subsequent two chapters (Chapters 4 and 5).

Chapters 4 and 5 have a similar outline to Chapters 2 and 3, but this time they describe the work on MEK1. More specifically, given that selective chemical modification of this protein is more demanding and that chemical modification at two native phosphorylation sites (Ser218 and Ser222) is considered, the descriptions of variant design and the detailing of the optimisation of chemical modification is more extensive than in Chapter 2. Chapter 4 builds up to the synthesis of MEK1-pCys222, after which the functional studies performed on all the chemical variants related to it, as well as the suitable controls, are discussed in Chapter 5.

A conclusion and scope of this work is then be discussed, particularly regarding to how this technology could be used as a tool in the elucidation of complex kinase pathways. Particular attention is given to the technology's potential use in studying complex kinase mechanisms. Regard is also given to how this technology may be applicable to investigation of kinases for further therapeutic manipulation.

Contents

Chapter 1

Introduction: Kinases and Chemical Modification

1.1 Cell Signalling and Phosphorylation

The action of cells in a multicellular organism, must be coordinated if the organism is to function as a single entity. Signalling between cells or between groups of cells is therefore required if they are to work together in a coordinated matter. The same is true at the cellular level, since the internal machinery of a cell is also complex and requires coordination in its own right. In a common scenario, the cell will be required to take an appropriate action upon receiving a signal at the cell surface receptors. This signal needs to get to the cell's nucleus for the cell to take any action, since cellular functions are achieved by the action of proteins, which are in turn synthesised from the genes encoded in the DNA contained in the nucleus. However, the nucleus and the cell surface are spatially separated, and both structures are relatively static. Mediators are therefore required to shuttle the signal from one region to the other.

One of the largest family of signal mediators in eukaryotes are the protein kinases.[2] Being one of the largest families, they are implicated in controlling a wide range of cellular functions, including cell growth, proliferation, differentiation, metabolism and apoptosis. Kinases more generally are phosphotransferases, which means that they catalyse the transfer of phosphate from a "high energy" phosphate co-factor onto their substrate. As this transfer is a biochemical process, kinases can be considered to be enzymes. Protein

kinases are a kinase subset and more specifically transfer the γ-phosphate from ATP onto their protein substrates. The phosphorylation of the substrate therefore constitutes the transfer of the signal.

In order to be a signal mediator, protein kinases also need to be able to receive a signal. They receive this signal also through phosphorylation, i.e. they themselves are phosphorylated. Therefore, a second protein kinase is required to phosphorylate the first. Considering the cell as a whole once again, in order to transmit a signal from receptor to a mediating protein kinase, a number of cell surface receptors also have protein kinase function,[3] but receptors do not account for the vast majority of protein kinases, of which there are 518 in the human genome.[2] Thus, for a signal to be transmitted from the cell surface receptor to the nucleus, it is not uncommon for there to be a number of intermediary protein kinases, giving rise to cascades of protein kinase signalling (Fig. 1.1). It should be noted that the term "kinase" is used interchangeably to refer to "protein kinase" for the remainder of this discussion, unless explicitly stated.

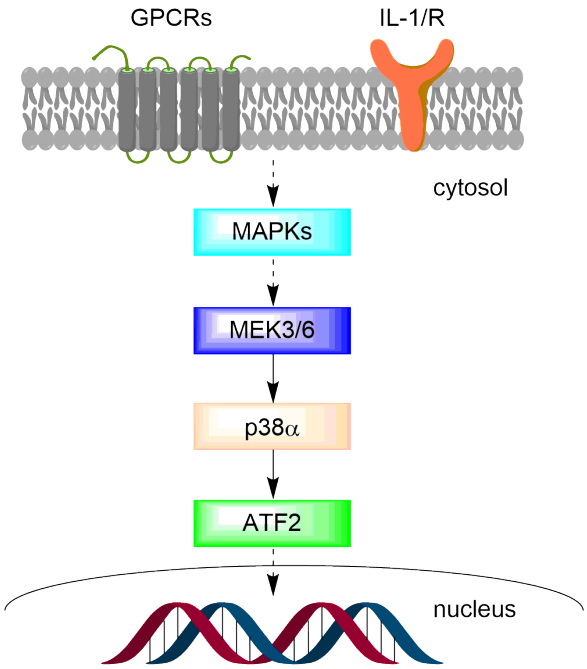


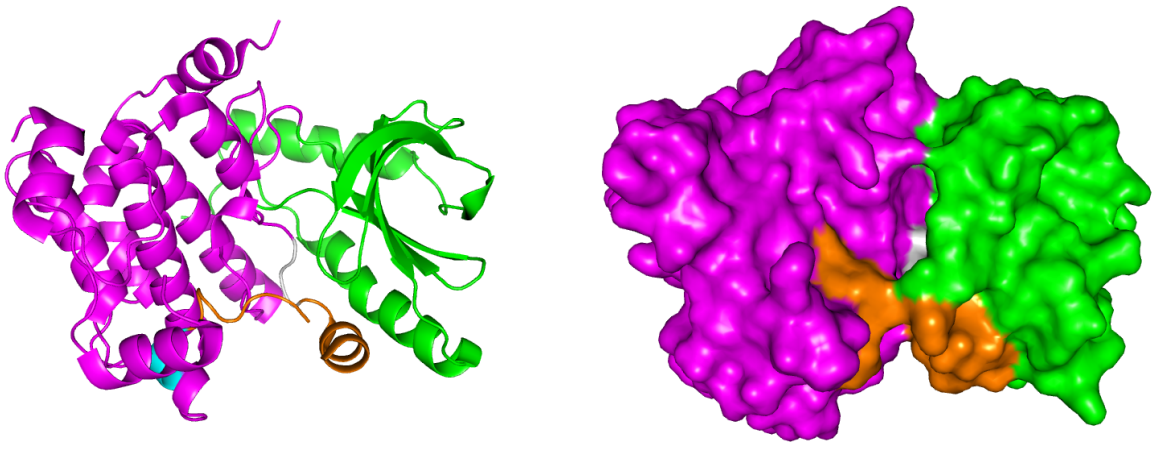
Fig. 1.1 A simple representation of a protein kinase signalling cascade showing the signalling pathway of p38α. Adapted from the literature.[1]

1.2 Anatomy of Protein Kinases and their Activation

The reception of a signal in the form of phosphorylation by a protein kinase results in a change of the kinase's activity from being a substrate of the upstream protein kinase, to the phosphorylation enzyme of

the downstream protein substrate. The event of phosphorylation is not merely a marker that a signalling event has taken place, but is in fact the cause of change in the kinase's function. However, to appreciate how phosphorylation is able to elicit such a change, the layout of a protein kinase must be explored.

Despite the wide range of protein kinases, a number of structural features of protein kinases are conserved.[4] Protein kinases have an overall two lobe structure, the lobes being designated *N*- and *C*-terminal (Fig. 1.2). These lobes are separated by the ATP binding "cleft" for the binding of the phosphate-containing co-factor, and are abridged by the so-called "activation loop". The activation loop, typically containing ∼20 residues, is flanked by the DFG and APE motifs. Named after the primary sequences of the three amino acids each of these motifs contain, these motifs are highly conserved throughout the protein kinases, and denotes the importance of the activation loop to the function of protein kinases.



(a) **(b)**

Fig. 1.2 **(a)** Cartoon and **(b)** surface renditions of the crystal structure of MEK1 protein kinase, with the different regions highlighted: N-terminal lobe (green), C-terminal lobe (magenta), DFG motif (grey, centre of each structure), APE motif (cyan) and activation loop (orange). The ATP binding cleft between the two lobes is more obvious in **(b)**. Drawn from 3EQC.pdb.[5]

It is the transfer of phosphate onto the activation loop when the signal is received by a protein kinase that causes the change in its function. When not phosphorylated, the protein kinase adopts a "DFG-out" conformation, which is enzymatically inactive. One of the reasons for its inactivity is that the activation loop then occupies the substrate binding site, thus preventing the substrate from binding to the kinase[6] (Fig. 1.3a). Phosphorylation of the activation loop adds negative charge to the vicinity, which calls for a re-equilibration of intramolecular forces. This re-equilibration causes the loop to change in conformation and in the process, unblocks access to the substrate binding site, giving rise to the kinase's active or "DFG-in" conformation (Fig. 1.3b). No longer denied access to its binding site, the substrate is thus able to bind

to the kinase. The substrate binding site lies adjacent to the ATP binding cleft where the close proximity of substrate with ATP facilitates substrate phosphorylation. Phosphorylation of the kinase in the activation loop therefore causes the kinase to become enzymatically active.

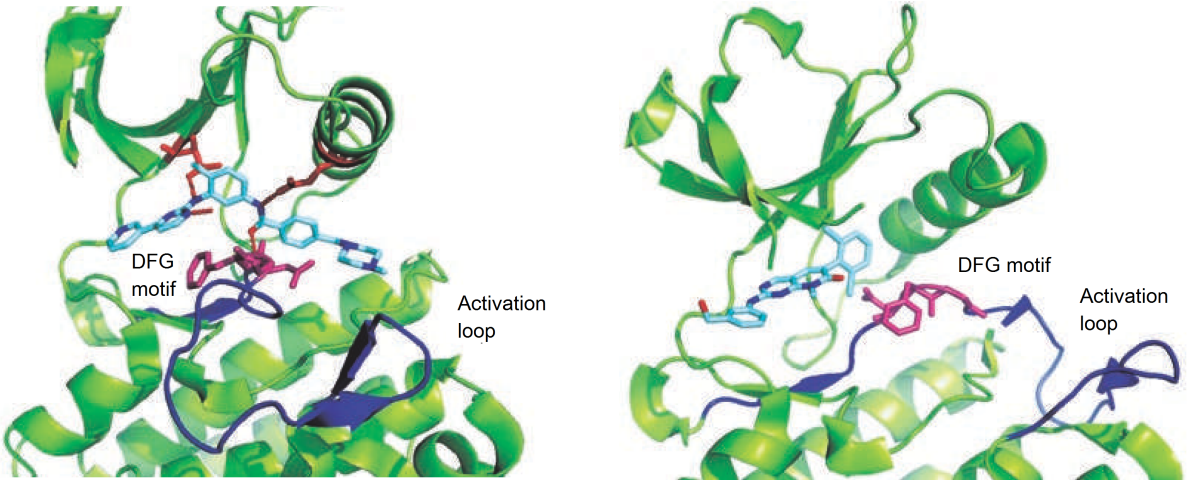


Fig. 1.3 Crystal structures of ABL protein kinase showing **(a)** the inactive "DFG out" conformation and **(b)** the active "DFG in" conformation. The change in conformation is associated with a change in activation loop position (blue) and DFG motif orientation (magenta). Adapted by permission from Macmillan Publishers Ltd: *Nature Reviews Cancer* 9(1): 28–39, Copyright © 2009.[6]

The active and inactive conformations are also distinguished respectively by their designations of "DFG-in" and "DFG-out". These designations refer to the position of the DFG motif relative to the ATP binding cleft in each of these conformations.[7] The function of the DFG motif is that it binds a magnesium ion that is required to coordinate to the phosphates of ATP.[8] The binding of magnesium is essential to the kinase's function,[9] and is why the DFG motif is so highly conserved across kinases. The DFG-in conformation allows the magnesium ion to associate with ATP in the ATP binding cleft, and is a further reason why DFG-in is the active conformation.

1.3 Basal Activity of Protein Kinases

The radically different conformations that a kinase is able to adopt is a testament to the flexibility of the activation loop. Due to this flexibility, the active and inactive conformations are not immutable states but merely the extremes of the conformational change, with the state of phosphorylation causing a strong bias towards one conformation over another. An observable result of this conformation flexibility arises when the enzymatic activity of the inactive conformation is investigated. It is often found that kinases have an intrinsic level of phosphorylating ability,[10–13] even when they themselves are not phosphorylated, and

therefore are in the "inactive" conformation. This low level of enzymatic activity demonstrates that even in the inactive conformation, some kinases are able to explore conformational space that still allows them to bring their substrate, magnesium and ATP together for phosphorylation. This level of activity is termed as basal activity.

1.4 Phosphorylation Sites on Protein Kinases

Protein kinases in eukaryotes are only able to catalyse phosphorylation to the side-chains of certain residues. While phosphorylation on histadine,[14] aspartate[14] and cysteine[15,16] have all been reported to occur in Nature, protein kinases in eukaryotes predominantly phosphorylate at serine, threonine and tyrosine.[17,18] Not all protein kinases are able to phosphorylate at all of these residues though, and the kinases are categorised accord to the residues they are able to phosphorylate. There are three main groups of protein kinases: serine/threonine kinases, tyrosine kinases and dual-specificity kinases, the last group being able to phosphorylate at both serine/threonine and tyrosine. Given the steric similarity of serine and threonine relative to tyrosine, this distinction between protein kinases is not surprising.

As protein kinases themselves are required for the phosphorylation of other protein kinase species, it follows that the activation of a protein kinase occurs by the specific phosphorylation of serine, threonine and tyrosine residues on the activation loops of these kinases. Due to the importance of phosphorylation to a kinase's activity, the identification of these specific phosphorylation sites has been the focus of much research into protein kinases.[19,20] While activation loop phosphorylation plays an important role in the functioning of protein kinases,[21] this is not the only region that phosphorylation has been identified on protein kinases; a range of studies have demonstrated phosphorylation at other sites.[18] As an example, human MEK1 (MAP or ERK kinase 1) is phosphorylated at two sites on the activation loop for kinase activation (Ser218 and Ser222),[22,23] but phosphorylation has also been detected at Thr23 and Ser298 among others,[24] outside the activation loop. These phosphorylation sites outside the activation loop are typically due to alternate regulatory pathways of controlling the behaviour of a protein kinase. Still, it is the active conformation that determines much of a protein kinase's behaviour, and most efforts have been devoted to the study of this form.

1.5 Biasing the Active Conformation

Study of an isolated protein's behaviour is a common method used to study the function of a protein. As proteins of interest often occur naturally in quantities that are difficult to observe or purify from source, recombinant protein expression is used to ensure a sufficient amount. For kinases, *Escherichia coli* and insect cells are common hosts for expression,[25] however, the expression of the active conformation of a kinase poses the challenge of how to convert the inactive form to the active one. Some kinases are able to autoactivate and if provided with ATP, their basal activity will allow them to autophosphorylate to the active conformation.[11] A rare few are intrinsically active, without requiring phosphorylation at all for their activity,[26] and are termed "constitutively active". For the majority of kinases however, a mechanism to bias the kinase towards the active conformation is required.

As already described, the native mechanism of biasing toward the active conformation is to phosphorylate the kinase on the activation loop, however, there are alternative methods of biasing without having to phosphorylate. One way is to introduce mutations that cause this conformational bias. With an emphasis of this discussion placed on biochemical transformations that occur in the kinase activation loop, these mutations fall into two categories: mutations that occur outside the activation loop (referred to here as "off-target mutations"), and those that occur within the activation loop.

In fact, as revealed through bioinformatic analysis, off-target mutations in kinases is a common mechanism to pathogenesis in cancer (Fig. 1.4).[27] In Nature, these mutations come to dominate through a natural selection process. A similar selection process can also be employed in the laboratory to find mutations of this category in a kinase of interest,[28,29] by using directed evolution.[30] However, from the perspective of designing a constitutively active mutant, mutations resulting in kinase gain-of-function, which in turn arise from directed evolution are difficult predict. This also means that the pattern of mutation(s) that arises from this process is rarely transferrable to another kinase. A further disadvantage to finding active mutants using this method is that it could produce protein variants that do have enzymatic activity, but bear little resemblance to the native active conformation. With directed evolution being an involved experimental process, overall, the use of directed evolution to find constitutively active mutants is not the a facile or general method.

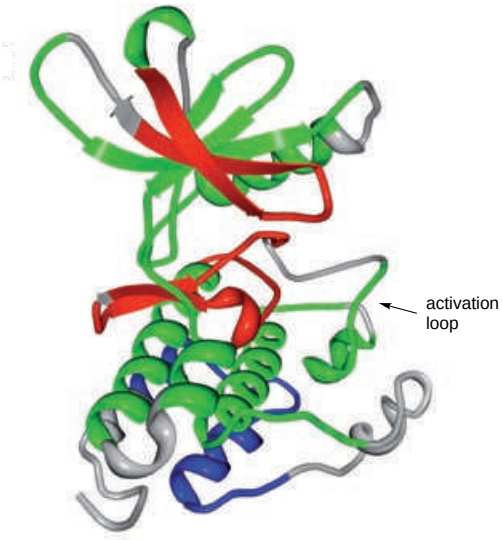
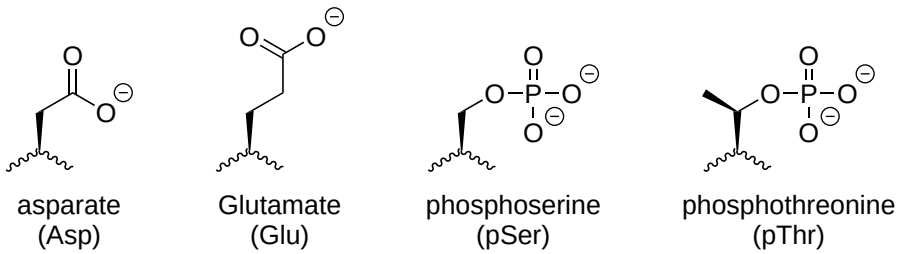


Fig. 1.4 A generic kinase structure showing regions rich (red) and poor (blue) of mutations associated with cancer. The activation loop is also marked. Adapted by permission from Macmillan Publishers Ltd: *Nature Reviews Genetics* 11(1): 60–74, Copyright © 2010.[27]

A more rational approach, while still using genetically encoded methods, is to substitute the negative charge that phosphorylation brings with aspartate or glutamate. Being acidic, aspartate and glutamate have a net negative charge at physiological pH and indeed, mutation of one of these in the place of the phosphorylation site residue is found to activate some kinases.[31] This strategy tends to be more successful when the phosphorylation site residue is either serine or threonine, since the side-chain length and sterics of phosphorylated serine and threonine are similar to that of aspartate and glutamate (Scheme 1.1). Tyrosine, on the other hand, has a larger side-chain, which conversely is not well represented by aspartate or glutamate. Since there are no other natural residues that are acidic, there are no suitable, genetically encoded substitutes for phosphotyrosine.



Scheme 1.1 Comparison of the acidic residues which the phosphorylated residues they are substitutes for.

This method of substitution has the obvious advantage that it is genetically encoded, which means that the method can be used not only to produce recombinant protein, but the mutated gene can be used in

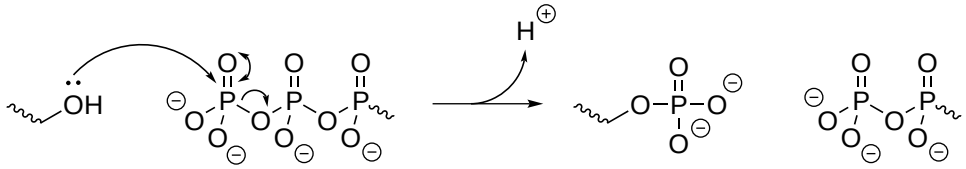
overexpression studies of the activated kinase *in cellulo*.[22] However, the substitution of the negative charge is only approximate, since physiological pH renders phosphoserine and phosphothreonine predominantly having -2 charge, while aspartate and glutamate have -1 charge. Furthermore, the group geometry is different with phosphate having tetrahedral geometry, while aspartate and glutamate have trigonal planar geometry. This usually results in a protein kinase variant of intermediate activity.[10]

However, even with partial activation using the method of acidic residue substitution, the scope of this method can be extended to preparing active forms of kinases downstream in a kinase cascade. In this case, the enzymatic action of the kinase that is activated by acidic residue substitution is used to phosphorylate its downstream substrate, which is another kinase. In the preparation of recombinant protein, there are two formats with which this method is commonly employed: either the phosphorylation of the downstream kinase occurs during protein expression in the expression host by co-expressing the genes of both kinases,[32–34] or it occurs after both kinases have been isolated and is done *in vitro*.[35,36] Both methods are of course equivalent by their mechanism, so the differences are mainly practical ones. The former method has the advantage of requiring fewer protein purification steps in isolating the phosphorylated, downstream kinase, while the latter method has the advantage that the conversion of the downstream kinase to the phosphorylated active form can be monitored. Preparation of recombinant protein *via* either of these formats often results in active kinase that is fully wild-type, but this is not always the case. These methods can easily lead to partial phosphorylation of the sites requiring phosphorylation,[36] or indeed over-phosphorylation with the additional phosphorylation occurring at a non-native position for phosphorylation.[34]

The directly aforementioned method of attaining the activated kinase highlights the key challenge in making a kinase active in the first place: that to obtain an active kinase requires the action of another active kinase, which may in turn require the action of another active kinase, and so on. The use of acidic residue substitutions allows this cascade arrangement to be bypassed, regardless of the shortfalls that this method has with activation efficiency, and the representativeness of the substitution for phosphoserine or phosphothreonine. This shortfall therefore leaves scope for alternative methods of bypassing the kinase cascade, all of which are chemical in nature.

1.6 Kinases in Single-Step Chemical Modification

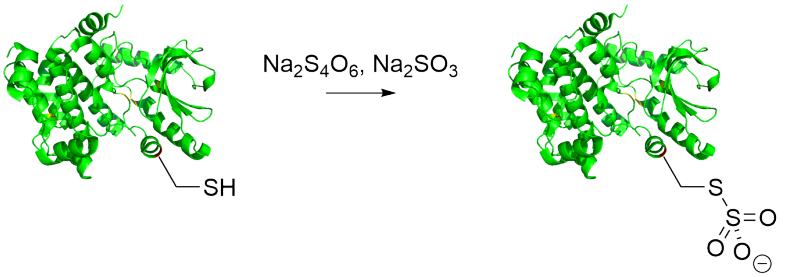
Fundamentally, the phosphorylation of a protein kinase is a chemical reaction where the hydroxyl oxygen of serine, threonine or tyrosine nucleophilically attacks the γ-phosphate of ATP, releasing ADP as the leaving group (Scheme 1.2).[37] It can therefore be envisaged that this reaction can be replicated solely by the use of chemical reagents.



Scheme 1.2 Mechanism of side-chain alcohol/phenol phosphorylation.

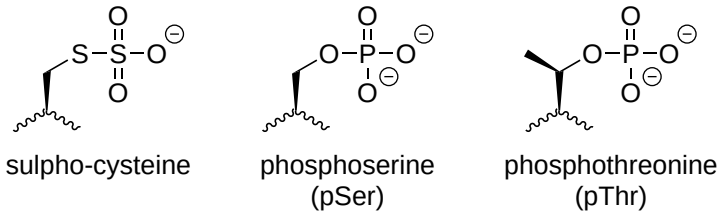
A major obstacle to the employment of a purely chemical approach is reaction specificity. As mentioned, activation of a kinase naturally requires the phosphorylation of a specific, alcoholic/phenolic residue(s) in the activation loop. The kind of specificity required can be easily achieved with another kinase due to the stringent steric demands of protein-protein interactions in biomolecular recognition, which confers the specificity in the natural system. The same specificity is difficult to achieve with chemicals, at least on alcoholic/phenolic residues since the natural occurrence of the hydroxyl functional group in proteins is quite high. Besides, alcohols are not the strongest nucleophiles out of all the natural amino acids, the strongest nucleophile being the thiol of cysteine.

The strong nucleophilic character of cysteine thiol has indeed been harnessed for the purpose of selective installation of negative charge on the kinase activation loop. Lamoureaux *et al.* reported a sulphonation reagent which was able to selectively form a disulphide with cysteine (Scheme 1.3).[38] Sulphonation in proteins had in fact been previously reported,[39] but was not applied to kinases. In using this reagent to give the active kinase, cysteine is substituted into the native phosphorylation site by genetic means. The purified kinase is then treated with the sodium tetrathionate and sodium sulphite which sulphonates the cysteine, thus installing the negative charge. Of course, the sulphonation reagent has the potential to react with any cysteine that is available on the kinase. However, due to the low natural occurrence of cysteine (cysteine is the second rarest amino acid in human proteins), site specific chemical modification is achievable.



Scheme 1.3 Sulphonation of MEK1 kinase.

Cysteine has similar steric properties to that of serine and threonine so this method, like the use of acidic residue substitution, is most suited to the substitution of phosphoserine and phosphothreonine. Compared to acidic residue substitution, this method has the advantage that the sulphonyl group has the same geometry as phosphate, both of them being tetrahedral (Scheme 1.4). However, the sulphonate still has the shortfall in net charge, being -1 charged. Despite this shortcoming, the sulphonated MEK1 kinase variant produced in the reported study showed similar activity to the enzymatically phosphorylated one.[38]

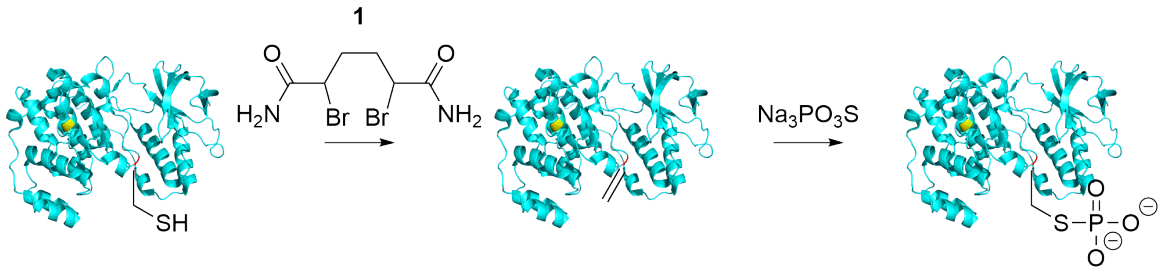


Scheme 1.4 Comparison of sulpho-cysteine with its natural counterparts.

1.7 Kinases in Multistep Chemical Modification

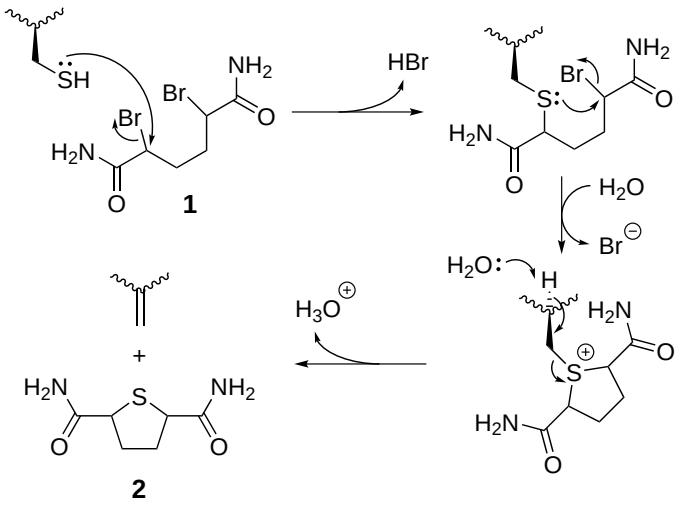
With the possibility of chemical modification at cysteine as the strategy for installing negative charge site-specifically, an obvious improvement on the -1 charge of the sulphonate mimic is to use phosphate itself. Direct phosphorylation of cysteine has been reported with the use of potassium phosphoramidate,[40] however, this reagent is more closely associated with studies on phosphohistidine,[41,42] which demonstrates that this reagent is not particularly chemoselective.[40] Instead, the strategy used to install phosphate on cysteine relies on a two-step, "tag-and-modify" process, where the cysteine is firstly further activated to a dehydroalanine (Dha) reactive intermediate, before the desired phosphocysteine is formed (Scheme 1.5).

A number of alternate strategies for obtaining the Dha intermediate have been devised,[43] but by far the most well-tolerated of these methods by proteins, as well as one of the highest yielding reactions is the bis-alkylation–elimination method, as can be achieved by using dibromide **1**.[44] Using this reagent, the thiol



Scheme 1.5 General reaction pathway of the "tag and modify" strategy used to install phosphocysteine on a protein, going *via* the Dha intermediate, as demonstrated on a p38α kinase variant.

of the cysteine being modified nucleophilically attacks the 2- and 5-centres of dibromide **1** sequentially, liberating bromide as the leaving group in each case (Scheme 1.6). From preliminary results, the product formed after the first alkylation is relatively long-lived and can easily be discerned. However, the second alkylation forms a cyclic sulphonium intermediate, which although has also been observed as a long-lived intermediate in some cases,[45] is more usually quickly eliminated as sulphonium **2** to give a Dha product. The elevated reactivity of the Dha over the original cysteine then allows the Dha to readily react with a wide range of nucleophiles,[46] which includes thiophosphate, thus resulting in the desired phosphocysteine.



Scheme 1.6 A proposed mechanism for the formation of dehydroalanine from cysteine, using the bis-alkylation–elimination reagent dibromide **1**.

Indeed, the installation of phosphate on cysteine has been reported in proteins,[46] including biologically relevant ones.[43,47] However, until now,[1,48] this method had not been applied to protein kinases for mimicking phosphoserine and phosphothreonine. It is the development of this chemical modification method, and its application onto kinases that is the focus of this project.

1.8 Project Motivations

This account on kinases and the signalling cascades they operate in has, up until this point, been a rather simplistic one. In reality, many kinases interact with a number of different partners, resulting in what is better described as a signalling network.[49] At the molecular level, this network behaviour manifests in kinases having multiple phosphorylation sites, as previously exemplified on MEK1. However, these alternative phosphorylation patterns may not only occur outside of the crucial activation loop, but also within the activation loop itself. It is not uncommon for a kinase to have more than one phosphorylation site in the activation loop,[48,50] and being in the activation loop means that each phosphorylation has a profound influence on the kinase's active function. Having multiple activation loop phosphorylation sites also presents the kinase with combinatorially alternate phosphorylation patterns, all of which could have subtle differences in behaviour, both in their enzymatic, as well as their regulatory functions.[51,52]

Ultimately, it is the importance that kinases have in cell signalling that brings them the most attention. Being responsible for many of the cell's functions, including cell growth, differentiation and apoptosis, any dysfunction of a kinase can lead to disease, the most notable being various cancers.[6] This has led to strong pharmacological interest in kinases, and the development of therapeutics for intervening in such cases of dysfunction. Although there are a number of therapeutics now available for the treatment of a range of diseases, mostly cancers, that implicate kinases,[53] the rapid expansion in this field is hindered by the sometimes limited understanding of a kinase's regulatory network, and the redundancy that a network system implicitly has.[49]

An illustration of this shortfall in the mechanistic understanding of disease lies in costs that the pharmaceutical industry has incurred due to the failures of proposed therapeutics in Phase II clinical trials (Fig. 1.5).[54] A large proportion of those costs are due to a lack in therapeutic efficacy, demonstrating the shortcomings in biological target validation. Given the large proportion of therapeutics that target kinases, there is a financial interest for a fuller understanding of kinase function and regulation, and by extension, in the molecular interactions they undertake.[55] The development of alternative investigative tools for the study of kinases and their interactions is therefore a contribution towards this need.

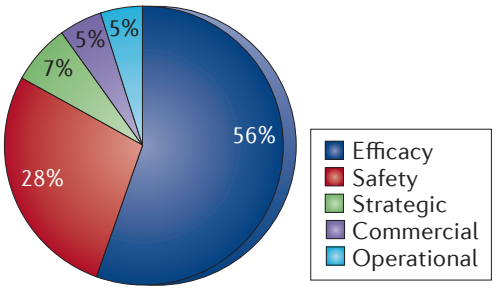


Fig. 1.5 Phase II clinical trial failures during 2011–2012. Of the 148 failures in that time period, 105 reported the reason for failure, the proportions of which are plotted in the chart. Reprintedbypermission-fromMacmillanPublishersLtd: *Nature Reviews Drug Discovery* 12(8): 569, Copyright © 2013.[54]

1.9 Advantages and Disadvantages of Chemical Modification

The use of phosphocysteine to mimic phosphoserine or phosphothreonine is a potential alternative investigative tool for kinase study, and has some distinct advantages over the more traditional, biologically-oriented methods in producing active forms, or phosphoforms of recombinant kinases. The key advantages over using acidic residue substitutes (aspartate or glutamate) have already been explained, but some of the advantages over enzymatic activation are less obvious.

As already mentioned, phosphoform homogeneity is a trait that can be difficult to achieve when enzymatic methods of phosphorylation are used. The issue of homogeneity becomes particularly apparent when a kinase of interest has two activation loop phosphorylation sites, and becomes yet more apparent if the interest lies in the intermediate phosphorylation states of a kinase. Having the same enzyme phosphorylating both phosphorylation sites further compounds this problem, since there would be no biological method in selectively phosphorylating a particular site. Despite these obstacles, interest in these intermediate phosphorylation states arises since they may be important in regulating a kinase's behaviour, so the understand of their purpose is central to pharmaceutical target validation.

The problem of a phosphoform mixture can however be overcome. Success is most likely if one of the phosphorylation sites is serine/threonine and the other tyrosine, since it may be possible to selectivity dephosphorylate one of the positions over the other. This can be achieved by the use of phosphatases. Phosphatases are the enzymatic counterparts to kinases so while kinases phosphorylate, phosphatases dephosphorylate. Like kinases, phosphatases which are involved in cell signalling are also generally subdivided according to the phosphorylated residues they act upon.[56] Thus, using the action of a phosphatase

with specific activity, selective dephosphorylation of one of the kinase's phosphorylation sites may be possible, and has been demonstrated in the literature.[33] If however the two phosphorylations occur at related phosphorylation sites (both serine/threonine, or both tyrosine), the separation of phosphoforms is likely to be difficult. The phosphoforms may be resolved by their difference in surface charge, but given that the charge difference between such closely related species will be small, physical separation is unlikely to yield satisfactory results. Regardless of success, as both of these methods require access to the bis-phosphorylated phosphoform during their preparation, any mono-phosphorylated phosphoform that results from either of these two processes cannot be guaranteed to be free from the bis-phosphorylated one.

A more common solution to obtaining the mono-phosphorylated states of a kinase where both phosphorylation sites are serine/threonine is again through mutation. In this regime, the phosphorylation site that is not of interest is mutated to an unreactive residue, which is usually one with aliphatic side-chain such as alanine, so that enzymatic phosphorylation at the other position can still continue.[12,22,52] This substitution means that a hydroxyl group is missing from the position where the mutation has been made, so no phosphorylation can occur at this position. The mutation also results in the lost opportunity for hydrogen bonding at the unphosphorylated position. This is turn can have deleterious consequences for the kinase's behaviour, which can lead to inconsistent or spurious conclusions.[12,22] With the chemical modification method however, the different destinies of the two phosphorylation sites are encoded in the difference in reactivity between the thiol of cysteine, and the alcohol or phenol of serine/threonine or tyrosine, so phosphorylation at one site can occur while in the presence of the native residue at the other phosphorylation site.

The installation of phosphocysteine by chemical means is not without its drawbacks though. Like the other methods of mimicry of acidic residue substitution and cysteine sulphonation described previously, cysteine phosphorylation is likely to be a suitable mimic only to phosphoserine and phosphothreonine. A more serious flaw arises from the synthetic strategy, which goes through the geometrically flat, Dha intermediate, a virtue of converting the α-carbon from an sp^3 to an sp^2 centre. It means that the L-amino acid stereochemistry is lost when the tetrahedral centre is reformed on attack of the thiophosphate nucleophile to form phosphocysteine. The key question of the project therefore is whether phosphocysteine produced by this synthetic strategy is a sufficient mimic of the phosphoserine or phosphothreonine that it is supposed to substitute, in the context of kinases.

1.10 MAPK Kinases and Choice of Model

The other, minor disadvantage of using cysteine for chemical modification becomes apparent on selecting suitable model kinases for the study at hand. Given that the chemical modification method relies on cysteine, an obvious selection criterion for a model kinase would be that it doesn't have any cysteines in its primary sequence, or at least has very few. Although cysteine is a rare amino acid, it is not natively absent in most proteins, where a bioinformatic analysis of all the human protein kinase primary sequences[2] showed that there are no kinases that have no cysteines in their primary sequence (Fig. 1.6). Instead, given the pharmacological motivations of this project, model kinases were selected from those of therapeutic interest, and only human kinases were used.

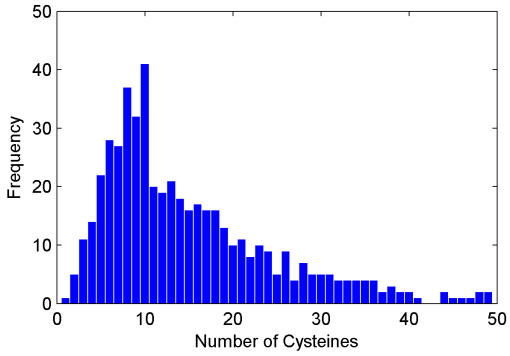


Fig. 1.6 Histogram of cysteine frequency in human kinases.

The MAP kinase (mitogen-activated protein kinases) family of kinases represent one of the most closely studied family of kinases. Interacting with many of the cell surface receptors,[57] MAP kinases are therefore implicated in many of the cell's functions, but they mainly respond to signals of cellular stress and to growth factors.[58] Being some of the most intensely studied kinases, MAP kinase cascades have been classified into tiers. Not including the surface receptor, a MAP kinase cascade has a minimum of three tiers and can have as many as five, with each tier of kinases interacting nearly exclusively with those in the tier below. Four, relatively distinct cascades have been identified (p38, JNK, ERK5 and ERK1/2),[59] usually referred to by the kinase in the MAPK tier (Fig. 1.7). Although there is some cross-talk between the cascades,[60] MAP kinases generally have few interaction partners, as exemplified by the Ras-Raf-MEK-ERK pathway. This results in the relatively linear cascade arrangement, particularly in the MAPKK (MAP kinase kinase) and MAPK tiers and makes them ideal candidates for use as model kinases, since their linear arrangement makes their biological interactions simpler to characterise. Kinases from the

MAPKK and MAPK tiers where thus selected as models, namely p38α and MEK1, both of which have also been studied for their disease implications.

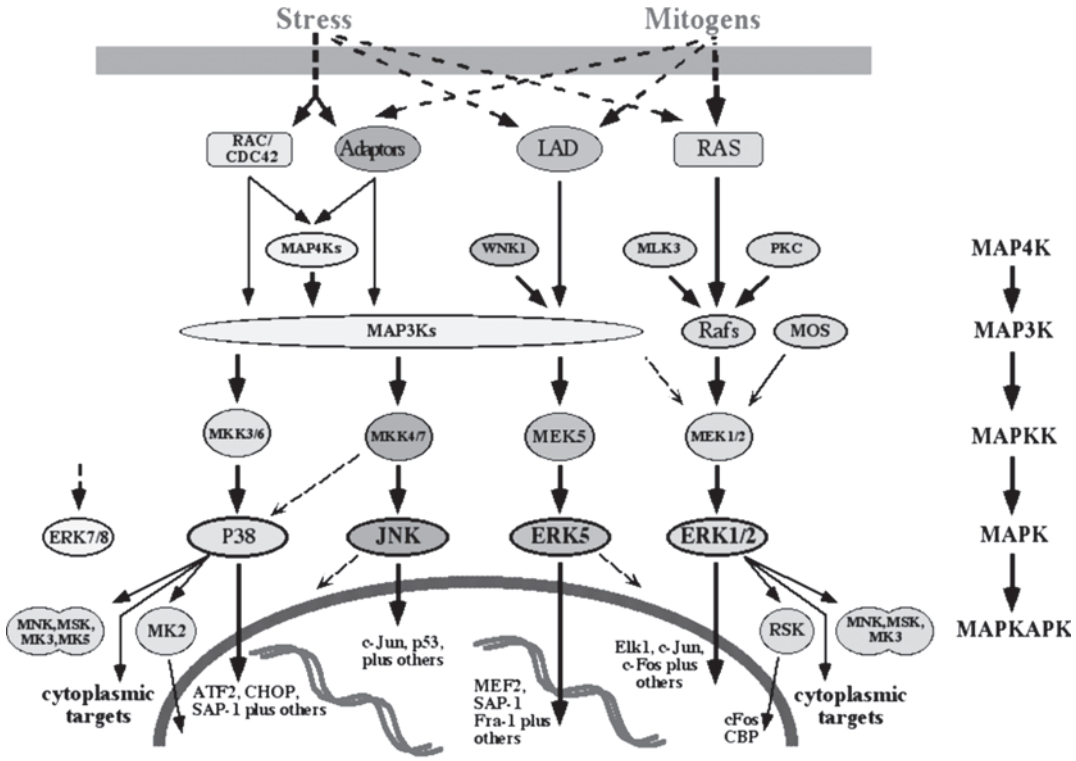


Fig. 1.7 Summary of the MAP kinase signalling network, showing the tiered structure according to their classification (far right). The Ras-Raf-MEK-ERK pathway shows a near linear cascade. Reprinted from Fig. 1 from H. Rubinfeld and R. Seger, The ERK cascade: a prototype of MAPK signaling, *Molecular Biotechnology* 31(2): 151–174, Copyright © 2005 Springer and Humana Press Inc.; with kind permission from Springer Science and Business Media.[58]

p38α comes from the MAPK tier of the p38 cascade. Being in the bottom tier of its cascade, p38α interacts with a host of transcription factors.[61] Having a wide range of downstream interactions results in p38α having a number of disease implications, where it has been studied mainly for its implications in inflammation,[62] and has received significant therapeutic interest in that field.[63,64] It also has implications in somatic,[65] cardiac[66] and neural[67] cellular regulation. p38α is phosphorylated at two sites in the activation loop for full activation, Thr180 and Tyr182.[33] Despite only the Thr180 phosphorylation site being suitable for study by cysteine chemical modification, mono-phosphorylated p38α-pThr180 is precedented and has been demonstrated to display different regulatory behaviours from the more common, bis-phosphorylated phosphoform.[51,52]

MEK1 on the other hand comes from the MAP2K tier of MAP kinases. Compared to p38α, MEK1 has few interaction partners, being only known to phosphorylate ERK1 and ERK2.[68] This specificity is

reflected in its disease implications, where it is implicated in cancer,[69] but the linear cascade arrangement downstream of MEK1 (and its homologue MEK2) has given it much pharmacological attention, since it is regarded as a bottleneck in what is otherwise a signalling network.[70] ERK1 and ERK2, like p38α and the other MAP kinases from the MAPK tier are phosphorylated at threonine and tyrosine for full activation. MEK1 is able to phosphorylate at both of these phosphorylation sites, making it a dual specificity kinase.[68] MEK1 is itself phosphorylated at two serine sites for full activation, Ser218 and Ser222.[12]

1.11 Project Aims

The aim of this project is to introduce phosphorylation onto kinases by chemical modification, resulting in their active conformation. With two model kinases, p38α and MEK1, the generality of the chemical modification method in activating kinases can be suitably challenged. Between these two model kinases, with p38α containing a threonine activation loop phosphorylation site and MEK1 containing two serine phosphorylation sites, the question of whether phosphocysteine makes good mimics of both phosphothreonine and phosphoserine can therefore be addressed.

Chapter 2

Chemical Modification on p38α Kinase

To realise the aim of chemically installing phosphocysteine into the activation loop of a protein kinase, p38α was chosen as one of the model kinases. The synthetic route used to do this involved reaction at cysteine, so the variants used in the study had to be carefully selected.

2.1 Construct Design

Plasmid constructs containing variants of p38α were donated by Richard Bazin, Pfizer, Sandwich. These variants, as received, had been designed for application in a fluorescence assay, in accordance with reported literature.[71,72] In these reports, fluorophores were attached to p38α *via* cysteine. However, p38α contains native cysteines, some of which are surface accessible. As control over the site of modification is crucial to installing the modification in the kinase activation loop, these natively occurring, surface accessible cysteines had been mutated out to prevent undesired chemical modification at these positions during the reactions (Fig. 2.1). Given the high accessibility of Cys119 and Cys162 in particular, both of these residues had been mutated to serine in all p38α variants. Cys39 however, being surface accessible but in a sterically crowded region of the protein and therefore sterically hindered, had been precautionarily mutated to alanine. An A172C mutation had been made as the site of fluorophore attachment, again according to the precedent, which was present in both variants. This gave rise to the two p38α variants p38α-C119S/C162S/A172C (**3**) and p38α-C39A/C119S/C162S/A172C (**4**).

As in the case of the fluorophore attachment, the method of installing dehydroalanine as described in Chapter 1 also relies on the reactivity of the cysteine thiol. These "surface cysteine-free" variants of

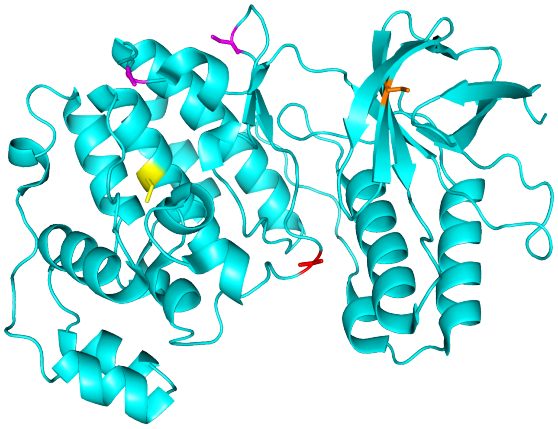


Fig. 2.1 Crystal structure of p38α kinase highlighting the surface accessibility of native cysteine residues. These cysteines are divided into three categories: highly solvent accessible Cys119 and Cys162 (magenta), hindered Cys39 (orange) and buried Cys211 (yellow). The site of chemical modification, A172C is also highlighted (red, centre of structure). Drawn from 1R3C.pdb.[73]

p38α were therefore directly relevant to the project aims, even though A172C is not a native site of phosphorylation. Given that different protein modification reagents can have different site selectivity profiles, it was decided to investigate both surface cysteine-free variants containing the A172C mutation in parallel against the range of chemical modification reagents used over the study.

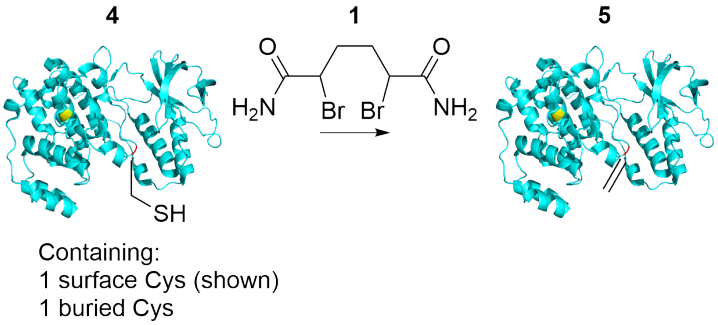
2.2 Variant Expression and Purification

Both p38α variants were expressed using the protocol established at Pfizer. Modifications were made to the Pfizer purification procedure though, with the resulting procedure used being a mixture of the Pfizer protocol and techniques used in the literature.[71] GST affinity chromatography was carried out according to the Pfizer protocol whilst the tag digestion was carried out using the techniques described in the literature. After tag rebinding, the purification protocols from both sources were the same. This purification method resulted in highly purified protein for both variants (Section 2.11.3).

2.3 Implementation and Optimisation of Reaction Conditions for Dehydroalanine Installation

The p38α variant **4** containing fewer cysteines was chosen as the first reaction testing model, since it was less likely to react multiple times with dibromide **1** (Scheme 2.1). Dibromide **1** is used to generate dehydroalanine (Dha) in place of cysteine as the first step towards the phosphocysteine installation on the

kinase. Dibromide **1** was synthesised according to literature procedures.[43] Along with the implementation of Dha installation on kinases, optimisation with a small range of buffers and varying buffer compositions was attempted.

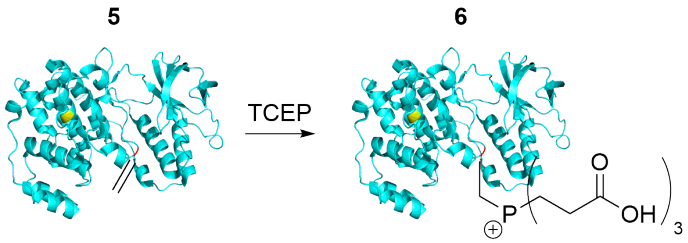


Scheme 2.1 Generation of Dha on p38α variant **4** with dibromide **1**. The remaining cysteines are also highlighted: buried cysteine (yellow) and cysteine reaction site (red).

The design premise on buffer composition was that the reaction buffer should be as similar as possible to the storage buffer, which in turn is important for protein stability and enzymatic competence. In previous work concerning the application of the bis-alkylation–elimination method for generating Dha in proteins, PBS was commonly used as a reaction buffer.[43] The use of a phosphate buffer with kinases was deemed to be undesirable since their enzymatic reaction involves phosphate. Chemical reaction conditions that allowed the kinase to be phosphate naïve were therefore sought. Co-solvents have also been employed in some procedures, most notably DMF, although other procedures had omitted this, therefore implying that the methodology would be applicable to proteins that may be unstable to organic solvent.[43] The use of organic co-solvent was considered potentially detrimental to the enzymatic activity[74] of the kinase and was therefore excluded from the reaction conditions. Finally, since the substrate of the chemical reaction is a protein, the use of protein stabilising agents in the buffer[75–77] were considered potentially beneficial to the enzymatic competence of the kinase. These stabilising agents needed to be tested to find out if they were tolerated by the chemical reaction. Thus, HEPES was tested as the buffering agent instead of phosphate, and the tolerance of the chemical reaction to HEPES, TCEP and glycerol was also tested.

The reaction of p38α variant **4** with dibromide **1** was run with 3 variations of p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP):

1. HEPES buffer (50 mM HEPES pH 7.5)
2. p38α storage buffer but at pH 7.5 and without TCEP



Scheme 2.2 Further reaction of Dha product **5** with TCEP resulted in the presence of this buffer component.

3. A 3:1 mixture of p38α storage buffer with HEPES buffer

In conditions 1. and 2., only the desired product containing Dha product **5** was observed (Fig. 2.2a and 2.2b) while with Condition 3., although **5** was the main product, a sideproduct corresponding to subsequent reaction of **5** with TCEP was also observed, resulting in the TCEP addition product **6** (Fig. 2.2c and Scheme 2.2). Confirmation of the adduct came with the addition of more TCEP, upon which **6** became the major product (Fig. 2.2d). It was thus concluded that the reaction could be run in the presence of glycerol, but not with even small quantities of TCEP. A further change to the pH (7.5 to 8.0) was made later on to be consistent in this parameter with the literature precedent for this reaction.[43]

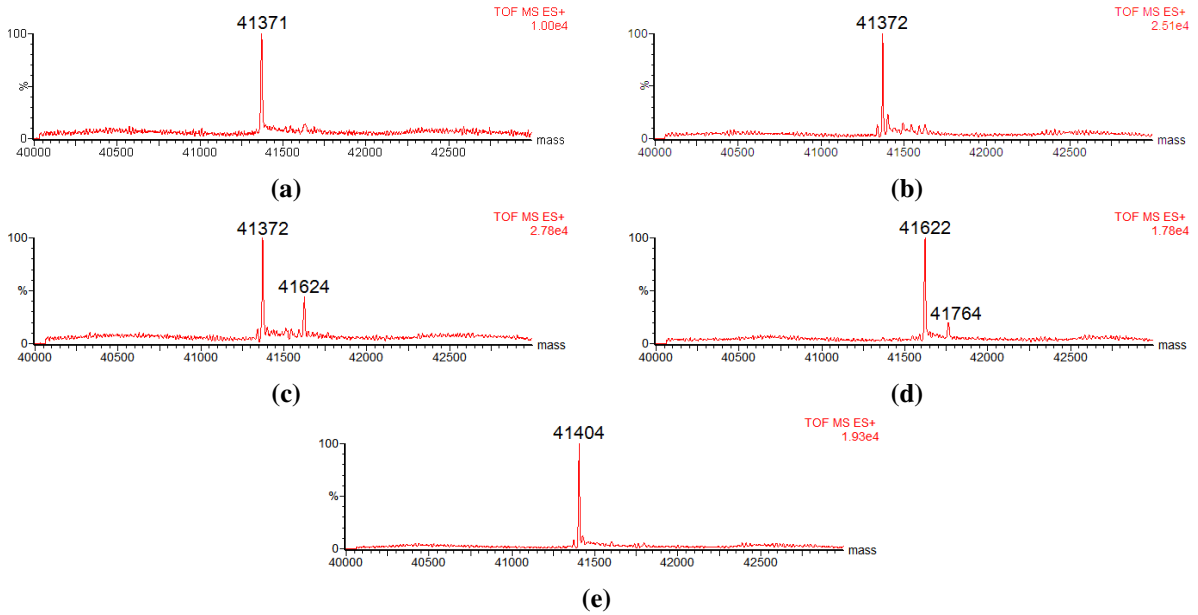
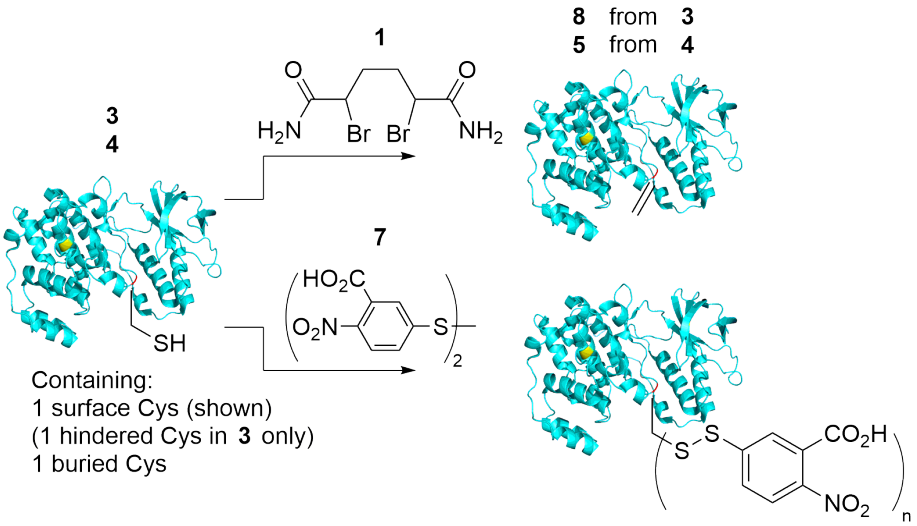


Fig. 2.2 Deconvoluted MS spectra of reactions of p38α variant **4** with dibromide **1** under different buffer conditions: **(a)** in HEPES buffer, **(b)** in p38α storage buffer at pH 7.5, without TCEP, **(c)** and **(d)** in 3:1 mixture of p38α storage buffer with HEPES buffer. Additional TCEP had been added to give the product in **(d)**. **(e)** MS of the unmodified **4** is shown for comparison. Expected mass change from unmodified **4** upon: Dha formation: -34 Da, TCEP addition product formation: $+285$ Da.

2.4 Chemical Reagent Regioselectivity

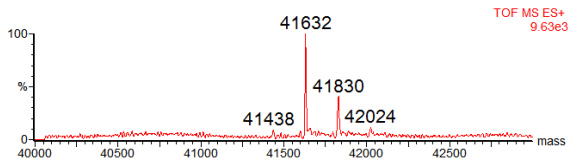
With suitable reaction conditions for Dha installation in hand, this chemical reaction could be tried against both the available p38α variants. Since the mutations of the p38α variants had been designed solely from a structural perspective, an additional purpose of this test was that the cysteine accessibility on these variants had to also be tested experimentally. In investigating reagent regioselectivity, the variants were not only tested against dibromide **1**, but also against Ellman's reagent (**7**) (Scheme 2.3). Ellman's reagent has strong precedence for being used as an indicator of cysteine surface accessibility.[78,79]



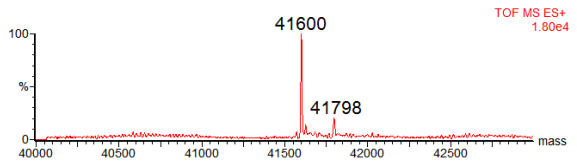
Scheme 2.3 Chemical reactions for testing cysteine surface accessibility.

Comparison of mass spectra (MS) taken during these reactions showed that while reaction with Ellman's reagent had the potential for multiple adduct formation, according to the number of overall cysteines in the variants (Fig. 2.3a and 2.3b), reaction with dibromide **1** resulted in the formation of a single product (Fig. 2.3c and 2.3d). This demonstrated firstly that different reagents have different reactivity profiles against the available cysteines of the variants, and secondly that although the more accessible cysteines are more likely to be the ones that react from the perspective of design, it is also possible for other, more hindered or buried cysteines to react with a modification reagent. In terms of using Ellman's reagent and structure aided design to predict the reactivity of cysteines, both of these methods therefore function merely as indictors.

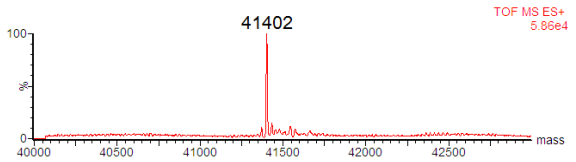
The rate of reaction at the different cysteines vary, as exemplified by the full conversion to the mono-Ellman's adduct, resulting in this adduct being the lowest modified state. This difference in rate according



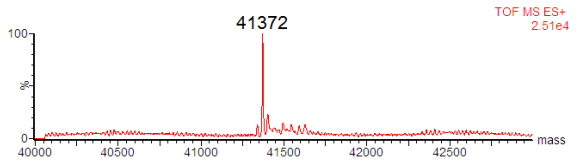
(a) p38α variant **3** with Ellman's reagent after 30 min at 4 °C.



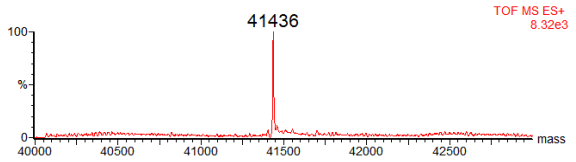
(b) p38α variant **4** with Ellman's reagent after 30 min at 4 °C.



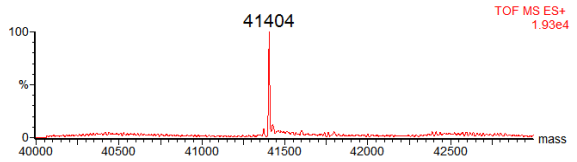
(c) p38α variant **3** with dibromide **1** after 4.5 h at 37 °C.



(d) p38α variant **4** with dibromide **1** after 4 h at 37 °C.



(e) Unmodified p38α variant **3**.



(f) Unmodified p38α variant **4**.

Fig. 2.3 Deconvoluted MS spectra demonstrating the different reactivity profiles, with respect to cysteine of **(a)** and **(b)** Ellman's reagent, and **(c)** and **(d)** dibromide **1** with p38α variants **3** and **4** (**(d)** reproduced from Fig. 2.2b for ease of comparison). Multiple peaks are observed in the MS from the reactions with Ellman's reagent, whereas a single peak is observed for the Dha formation with dibromide **1**. **(e)** and **(f)** MS of the unmodified p38α variants **3** and **4** are shown for comparison. Expected mass change from unmodified p38α variant upon: Dha formation: -34 Da, Ellman's adduct formation: $+197$ Da.

to a cysteine's stereoelectric environment then gives rise to the range of modification states observed (Fig. 2.3a and 2.3b). As for the reaction with dibromide **1**, given that only single site modification was observed, even at prolonged reaction times with both variants (Fig. 2.3c and 2.3d), it was assumed that the reaction with **1** occurred at the designed site at position 172. This was later confirmed by LC-MS/MS, as well as further observations of reaction rates (Sections 2.8.2 and 2.9). As the requirement of single site modification with dibromide **1** had been met by both variants, but only one variant containing the A172C mutation was required for further study, variant **3** (referred to exclusively as p38α-Cys172 later in the discussion) was chosen for the majority of further tests, particularly in the procedures described in Chapter 3, since its mutation profile is closer to the wild-type.

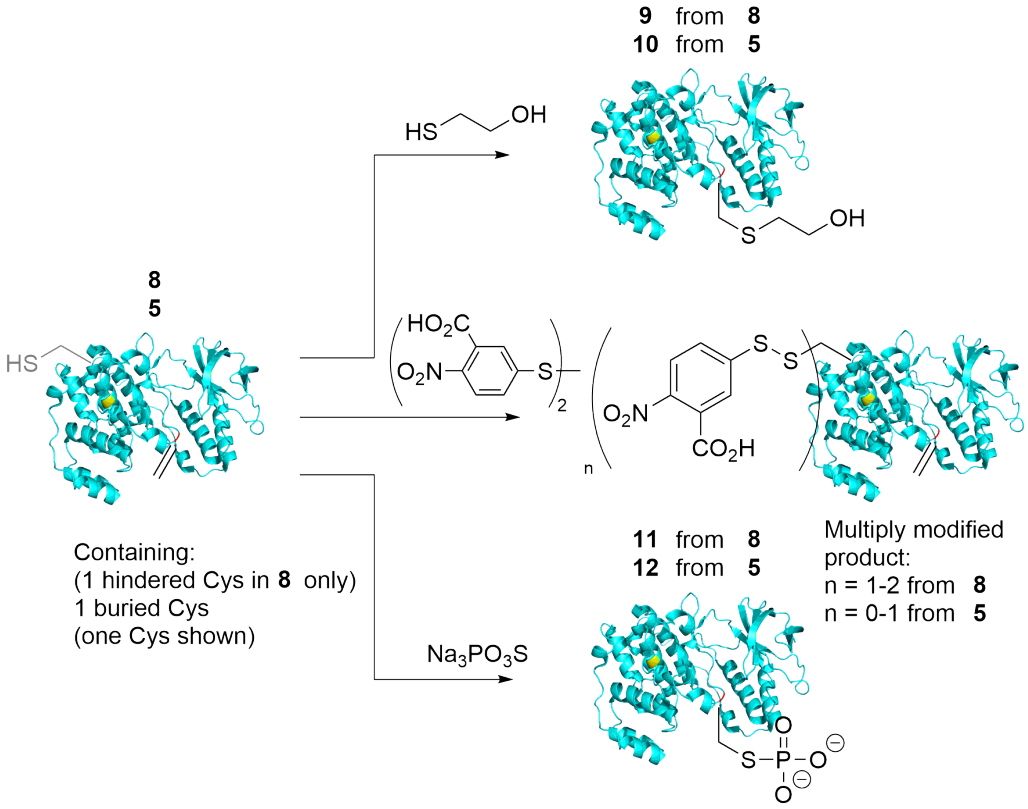
2.5 Chemical Tests against Installed Dehydroalanine

On top of the mass spectral evidence that Dha had been installed in both p38α variants, chemical tests were required to provide further evidence of this. A series of suitable control reactions as had been previously used[43] were applied to the Dha products (**8** and **5**) of both p38α variants (Scheme 2.4):

1. Reaction of the Dha product with β-mercaptoethanol (positive reaction control)

2. Treatment of the Dha product with Ellman's reagent (negative reaction control)

3. Reaction of the Dha product with sodium thiophosphosphate (desired modification reaction)



Scheme 2.4 Chemical reactions for determining the success of Dha installation in p38α variants.

The Dha products (**8** and **5** respectively) of both p38α variants **3** and **4** reacted completely with β-mercaptoethanol (β-ME) to give the β-ME adducts **9** and **10** (Fig. 2.4a and 2.5a). These adducts were stable products, and their presence was monitored over a few hours, unlike the products formed during disulphide bridge reduction.[80] The formation of **9** and **10** therefore demonstrated that a reactive moiety, namely Dha had been installed on the p38α variants. A similar observation was made for the reaction between sodium thiophosphate and the Dha products **8** and **5**, where masses corresponding to the expected phosphocysteine (pCys) derivatives were observed (Fig. 2.4c and 2.5c), giving the phosphocysteine products **11** and **12** respectively.

With free cysteine thiols still remaining in the Dha products, the results from the treatment of Dha products **8** and **5** with Ellman's reagent were less well defined. The number of Ellman's adducts formed were consistent with the remaining numbers of cysteines in both cases (Fig. 2.4b and 2.5b), and consistent with

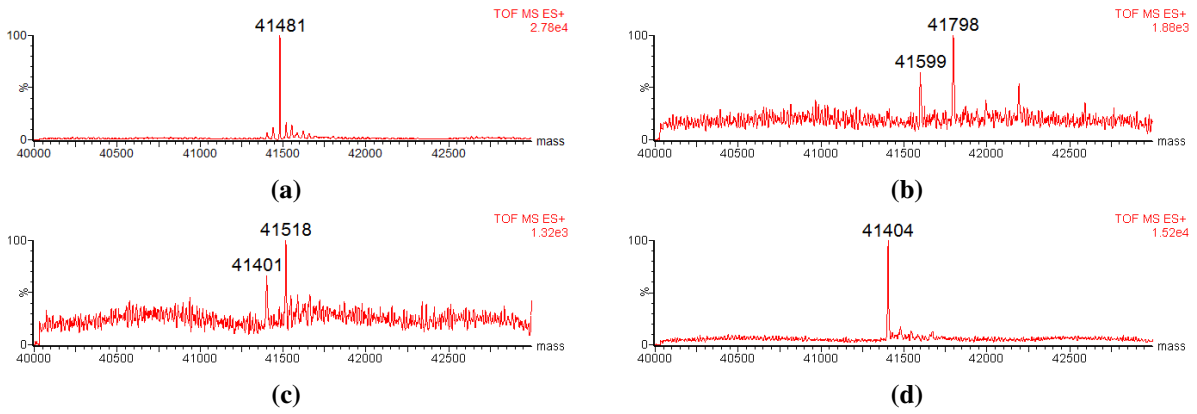


Fig. 2.4 Deconvoluted MS of products from control reactions on Dha product **8**: **(a)** from reaction with β-mercaptoethanol. Expected mass change from **8**: +78 Da, **(b)** from reaction with Ellman's reagent. Two cysteines are expected to remain, which is demonstrated from a series of two adduct products being formed. Expected mass change from **8**: +197 Da×n where n =1 or 2, **(c)** from reaction with sodium thiophosphate. Expected mass change from **8**: +114 Da. The MS shows incomplete conversion to the mass of the desired product. **(d)** Deconvoluted MS of Dha product **8**.

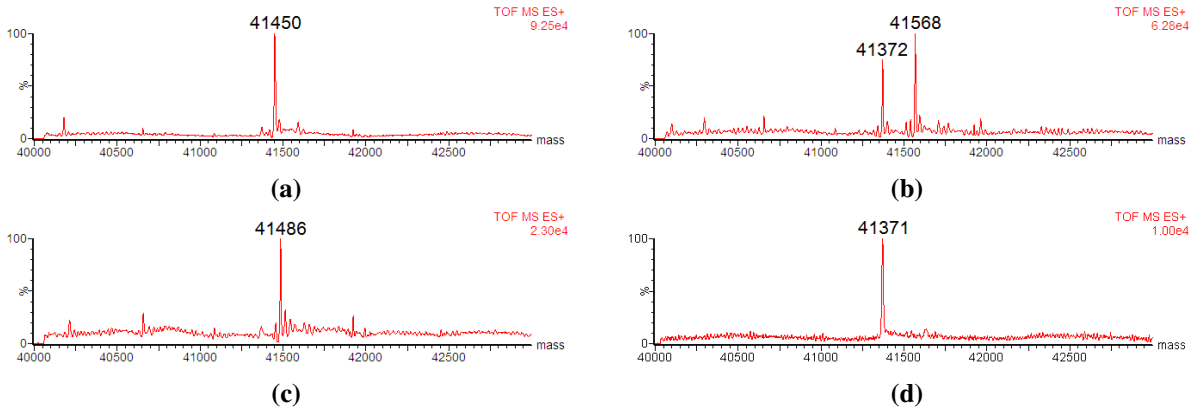


Fig. 2.5 Deconvoluted MS of products from control reactions on Dha product **5**: **(a)** from reaction with β-ME. Expected mass change from **5**: +78 Da, **(b)** from reaction with Ellman's reagent. One cysteine is expected to remain, which is demonstrated by further conversion to only one adduct product. Expected mass change from **5**: +197 Da, **(c)** from reaction with sodium thiophosphate. Expected mass change from **5**: +114 Da. **(d)** Deconvoluted MS of Dha product **5** (reproduced from Fig. 2.2a for ease of comparison).

the consumption of one cysteine being converted to Dha in each case. However, due to the presence of multiple products arising from the different reactivities of the remaining cysteine(s), this resulted in the low definition of this test. Since the results were not clear-cut, it was concluded that this test was of limited utility. For consistency, this test was used for all installations of Dha made on p38α, but was discontinued when working with MEK1.

2.6 Reaction Site Shift and Further Mutation

The successful installation of phosphocysteine at position 172 in p38α demonstrated that the scope of this chemical transformation could be expanded to kinases. Further, the studies on the A172C variants had given good grounding to the technical requirements and the reaction condition requirements of protein modification on p38α. However, position 172 is not a native phosphorylation site. Instead, p38α is phosphorylated natively at two sites, Thr180 and Tyr182 (Fig. 2.6). As the primary intension of the project is to use chemical modification as an additional tool for biological functional studies, and as phosphocysteine is unlikely to be a suitable surrogate for Tyr182, attention was turned only to phosphorylation at position 180.

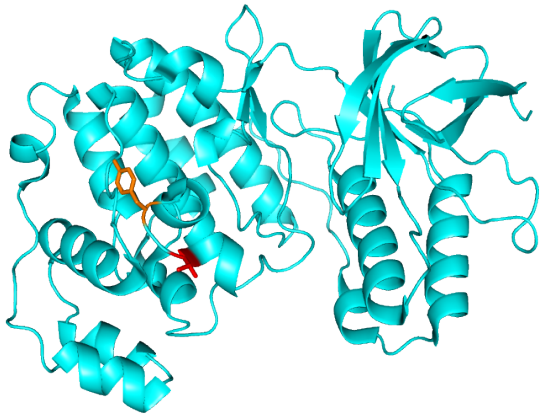
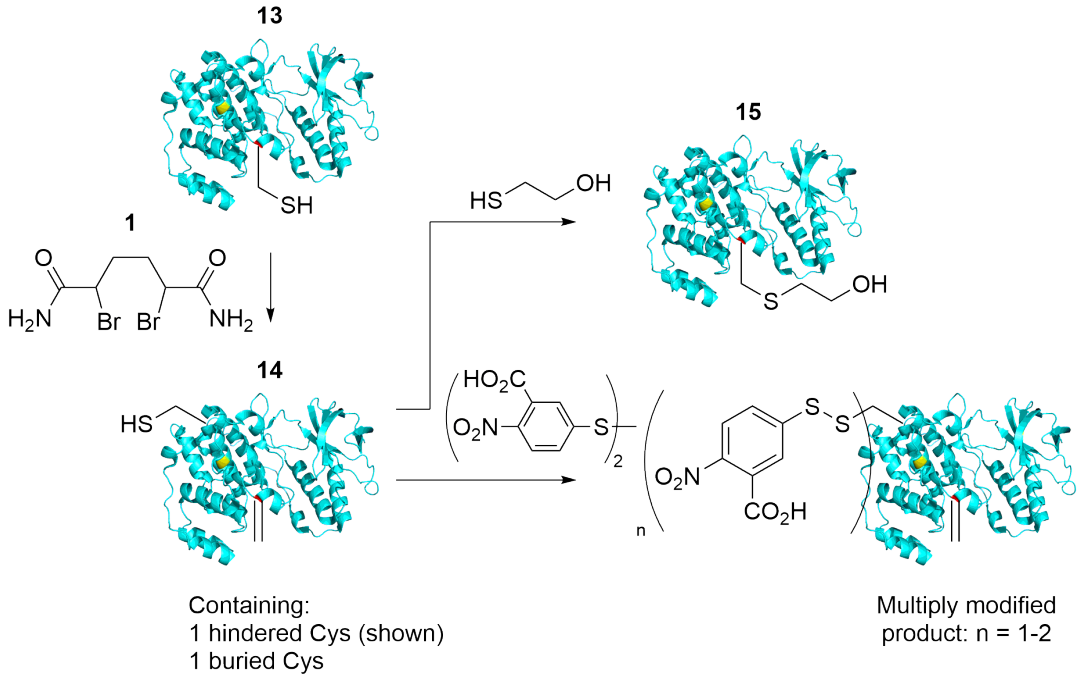


Fig. 2.6 Crystal structure of p38α kinase highlighting the two residues that are natively phosphorylated in active p38α: Thr180 (red) and Tyr182 (orange). Drawn from 1R3C.pdb.[73]

Appropriate site-directed mutagenesis was therefore done to reverse the A172C mutation and to install a new chemical modification site at 180 *via* T180C mutation. At the same time, mutation to obtain the wild-type sequence was carried out in parallel, which although was not used in the study of p38α activity, it was used in the verification of the biological recognition of MEK1 (Section 5.4). The plasmid construct encoding the new p38α variant p38α-C119S/C162S/T180C (**13**, also referred to as p38α-Cys180 later

on in the discussion) was suitably expressed and the protein purified in a similar manner to the other two p38α variants (Section 2.11.3).

2.7 Installation of Dehydroalanine on p38α-C119S/C162S/T180C



Scheme 2.5 Dha formation and Dha controls reactions on p38α variant **13**.

Dha was similarly installed onto p38α variant **13**, using the conditions established on p38α variants **3** and **4** (Fig. 2.7a and Scheme 2.5), with the results from the β-ME and Ellman's reagent control reactions giving the products as would be expected on a p38α variant with two native cysteines remaining (Fig. 2.7b and 2.7c). Addition of thiophosphate on the Dha product **14** however proved to be a little more challenging than for the Dha products already investigated, where more extensive efforts were taken for the reaction's optimisation.

2.8 Thiophosphate Reaction Optimisation on p38α-C119S/C162S/T180C

The major impetus for more detailed investigation into the thiophosphate addition reaction was the slower reaction kinetics of thiophosphate addition onto the Dha product **14** compared to Dha product **8** (Scheme 2.6 and Section 2.8.2). Since prolonged exposure to elevated temperatures (37 °C) could result in decreased enzymatic activity for assays further downstream (Section 3.5), and yet elevated temperatures were necessary for the reaction to be complete in a practical time, shorter reaction times were desired

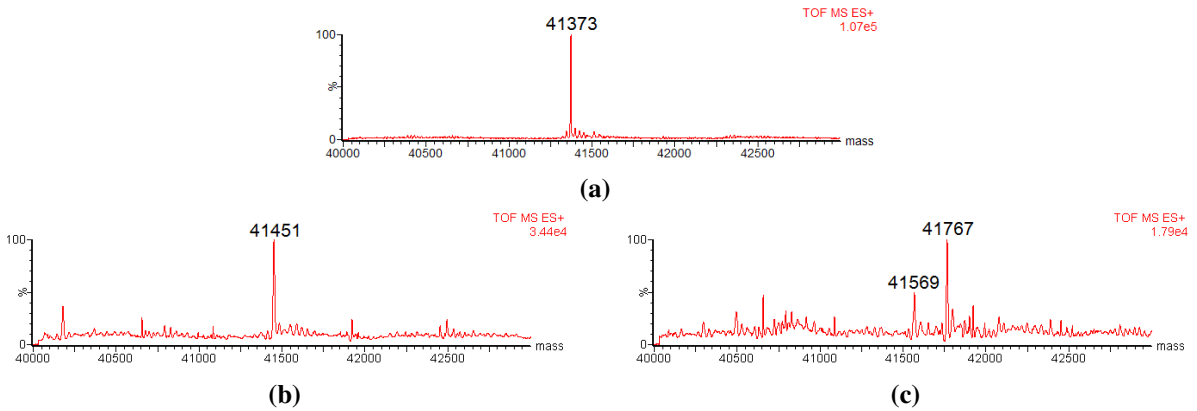
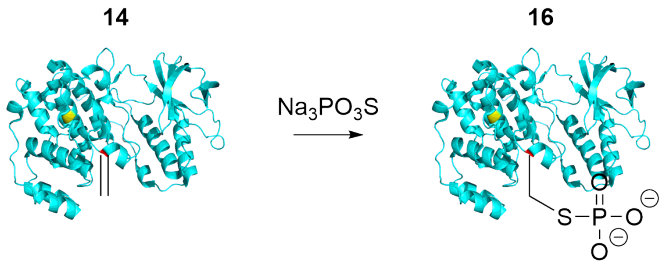


Fig. 2.7 Deconvoluted MS of the products from Dha formation on p38α variant **13** and from subsequent control reactions with the Dha product. **(a)** Dha product **14**. **(b)** From reaction of Dha product with β-ME. Expected mass change from **14**: +78 Da. **(c)** From reaction of Dha product with Ellman's reagent. Two cysteines are expected to remain, which is demonstrated from the series of two adduct products being formed. Expected mass change from **14**: +197 Da$\times n$ where $n = 1$ or 2.

in order to avoid problems to enzymatic activity. Thus, buffer conditions and pH were screened to see whether changes in these parameters could cause changes in reaction rate. However, before optimisation studies on the thiophosphate reaction could be done in earnest, improvement to the investigative methods were required.



Scheme 2.6 Thiophosphate addition onto Dha product **14**.

2.8.1 Practical Considerations and Quantitative Mass Spectrometry

Although MS is not widely considered a quantitative technique, it was hypothesised that this consideration would not apply to this study. The hypothesis was based on the postulate that since the chemical modifications made here were small compared to the size of the protein, the changes to the physiochemical properties of the protein would also be insignificant. This in turn would result in all chemically modified protein species ionising similarly, therefore giving MS signals that are proportional to the quantity of a given protein species. Since quantitative analysis of the collected data would improve the richness of the data, efforts were made to make this type of analysis of the MS data routine.

While developing the workflow for routine quantitative analysis, the opportunity was also taken to streamline the existing workflow. Major limitations of the LC-MS method previously used were the sample size (~40 µL of 1 mg/mL protein) and the time (35 min) required to run each sample, both of which were excessive. The LC-MS system used was miniaturised using a shorter column (predominantly ProSwift® RP-4H monolith, 1×50 mm column from Dionex) so that both of these parameters would be $^1/_3$–$^1/_2$ of their original values. This reduction in time and sample requirements allowed more samples to be screened per unit time.

With the capacity for more samples to be run in a given time, there is more data that needs processing and analysing. These tasks were also streamlined and much of the data handling workflow was automated. The increased LC-MS and data handling capacity then allowed for 4–5 samples per reaction to be executed and analysed quantitatively over the reaction timescale (4–8 h). Since one of the aims in optimisation was to get shorter reaction times, the quantitative data was used for kinetic analysis of the reactions, and this analysis was used as a further guide in reaction optimisation.

The quantitative nature of the MS technique used was tested after the conversion to thiophosphate adduct **16** had been optimised. In an experiment where the thiophosphate adduct **16** and the Dha product **14** were combined in known ratios and the samples analysed, it was found that peak height in the MS was suitably proportional to species quantity (Fig. 2.8). The streamlining of the LC-MS and data analysis workflow was instrumental to the collecting of large biological kinetic assay datasets further downstream (Sections 3.5 and 5.4), where the same techniques were used.

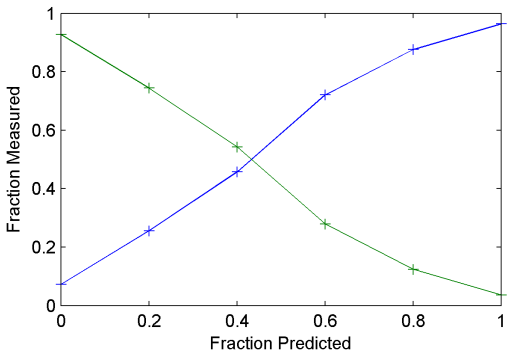
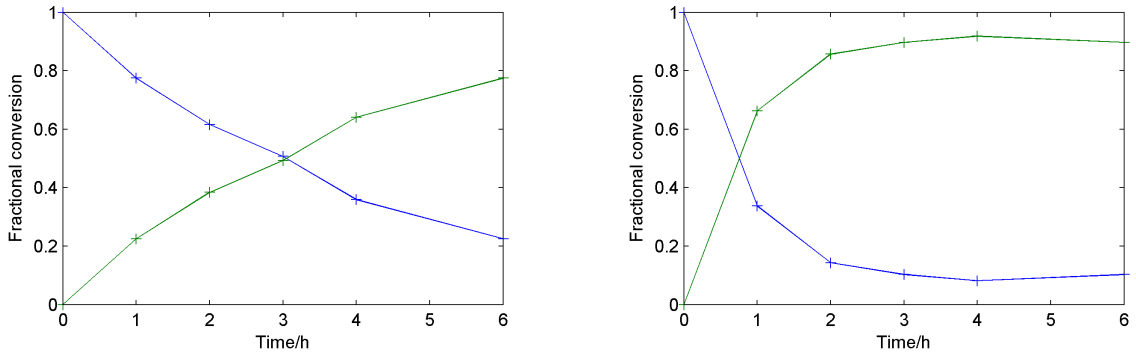


Fig. 2.8 Calibration curve made with mixtures of known proportions of thiophosphate adduct **16** (green) and Dha product **14** (blue) and the resulting read-out from LC-MS analysis.

2.8.2 Comparison of Reactions on p38α at Positions 172 and 180

The slower reaction rate of thiophosphate addition to Dha product **14** compared to Dha product **8** (referred to at this point as p38α-Dha180 and p38α-Dha172 respectively to emphasise the difference in reactive cysteine position, with all other cysteine mutations being equal) could easily be determined using the quantitative analysis workflow (Fig. 2.9). In making this conclusion, the assumption was also made that despite the difference in the position of the reactive cysteine, both p38α variants p38α-Cys180 (**13**) and p38α-Cys172 (**3**) had folded in the same way. This assumption was considered reasonable for two reasons. Firstly, the difference between the two variants lies in the highly flexible activation loop, so the small changes in polarity which occur from these single residue changes are unlikely to radically affect the structure of the kinase as a whole. Secondly, the literature precedent on making mutations in the activation loop was that it still resulted in the active kinase,[71] which must therefore have folded correctly.



(a) **(b)**

Fig. 2.9 Kinetic analysis of conversion of Dha product (blue) to thiophosphate adduct (green) for Dha products: **(a)** p38α-Dha180 (**14**) and **(b)** p38α-Dha172 (**8**).

With the folds of the two variants likely to be the same, this difference in chemical reaction rate must therefore arise predominantly from the difference in reaction site position, and more specifically in the differences of the steric and electronic nature of the two positions. To support this notion, the surface accessibility and electrostatic surfaces for p38α were calculated (Fig. 2.10). It was observed that while there was little difference in the sterics between positions 180 and 172 (Fig. 2.10a), this was not the case with the net charge. Position 180 was found to have net negative charge, while position 172 was net positive (Fig. 2.10b). These calculations as made on the p38α crystal structure were therefore consistent with the observation in thiophosphate addition reaction rate, since attacking a negatively charged region (as in the case on p38α-Dha180) with a highly negatively charged nucleophile is likely to be more difficult,

therefore causing a reduction in chemical reaction rate and making this reaction slower. Further, it was concluded that the difference in electronics between the two positions had a greater effect than the sterics on reaction rate. The difference in reaction rate also constitutes evidence that the cysteine designed for chemical modification (either at 180 or 172) was indeed the one being chemically modified, since it would not be expected to see such difference in chemical reaction rate at corresponding cysteines in the same position (as mentioned previously in Section 2.4).

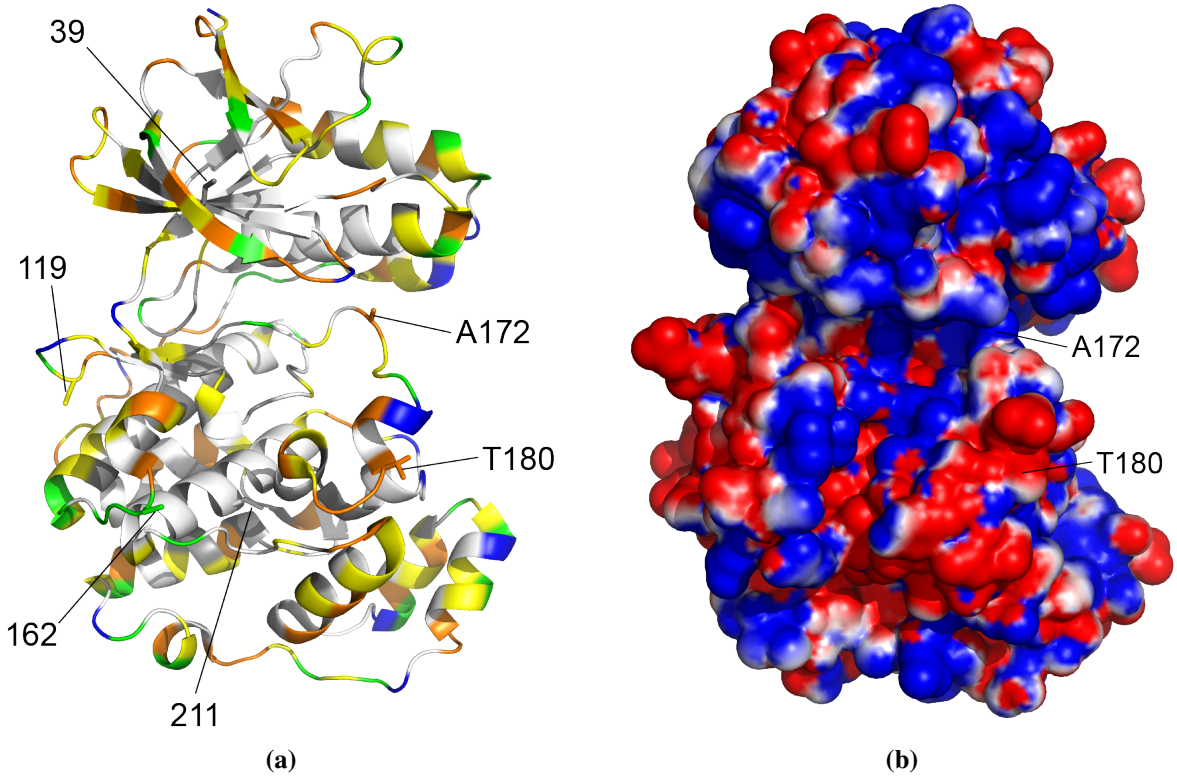


Fig. 2.10 Electrostatic and surface accessibility calculations for p38α. **(a)** Surface accessibility calculations show that the two reaction sites have similar solvent accessibility. The position of the native cysteines are also highlighted. Levels of accessibility: >100% (blue), 100–75% (green), 75–50% (orange), 50–25% (yellow), 25–0% (white), 0% (grey). **(b)** The region around Thr180 has net negative charge (red) while the region around Ala172 is positive (blue).[1] Drawn from 1R3C.pdb,[73] a structure with p38α in its DFG-in conformation.

The converse was found to be true for the reaction rate of Dha incorporation with dibromide **1**, although the effect was less pronounced (Fig. 2.11).

2.8.3 Effect of Buffer Composition

With the electronics of the reaction site being the dominant factor in the rate of thiophosphate addition onto Dha, strategies to mask the charge were sort after. It was thus perceived that a change in buffer composition could result in this masking to give the increase in reaction rate. The p38α reaction buffer

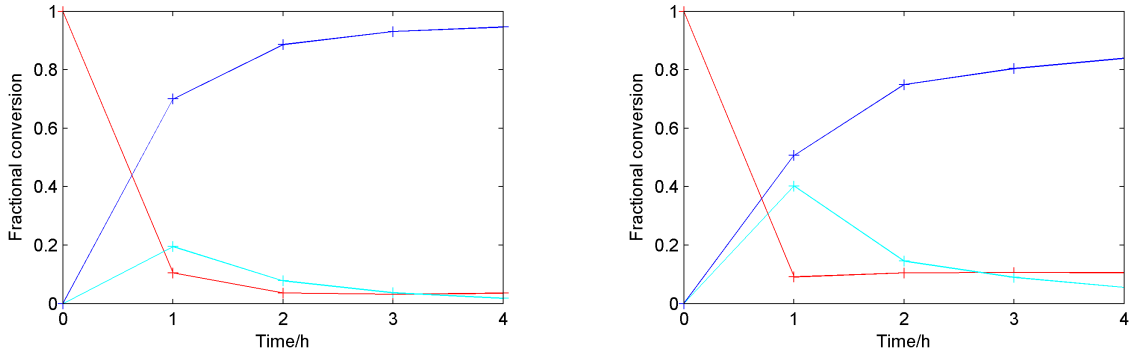


Fig. 2.11 Kinetic analysis of conversion of unmodified p38α variant (red) to Dha product (blue) for p38α variants: **(a)** p38α-Cys180 (**13**) and **(b)** p38α-Cys172 (**3**). The reaction goes *via* a mono-alkylated bromide adduct (cyan), which is a long-lasting intermediate.

(50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) used until now was therefore compared against buffer with a higher concentration of sodium chloride (150 mM) in the hope that the increase in ionic strength of the buffer could cause such masking. As a further comparison, HEPES and phosphate buffer (50 mM pH 8.0) were used as standards.

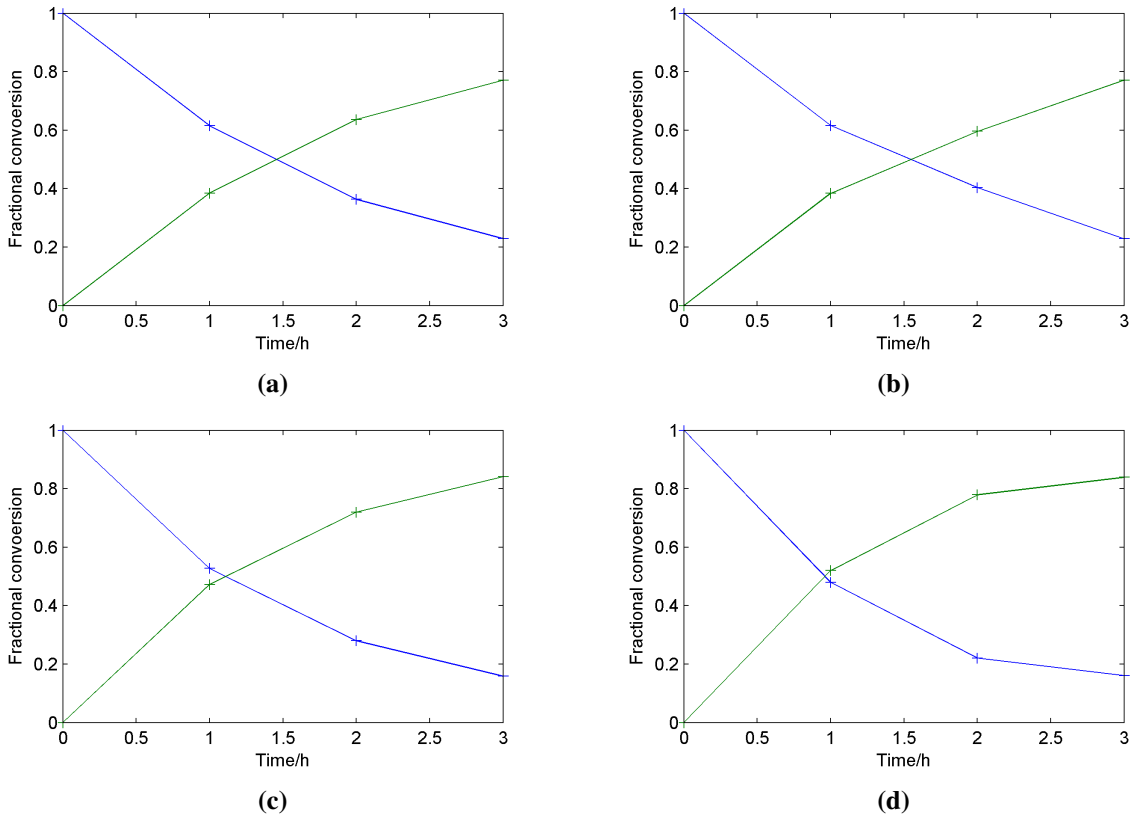


Fig. 2.12 Kinetic plots of thiophosphate addition reactions on p38α-Dha180 (blue) under various buffer conditions: **(a)** in p38α reaction buffer, **(b)** in p38α reaction buffer with elevated NaCl concentration, **(c)** in HEPES buffer, **(d)** in phosphate buffer.

It was found that changing salt concentration (comparing Fig. 2.12a and 2.12b) had little effect on the rate whilst omission of glycerol increased the rate (comparing Fig. 2.12c and 2.12d), but led to an increase in protein precipititation. There was little difference in the rate when different buffering agents were used (comparing Fig. 2.12c and 2.12d). However, given the buffer design considerations and criteria described in Section 2.3, none of the modified buffer conditions were adopted.

2.8.4 Effect of pH

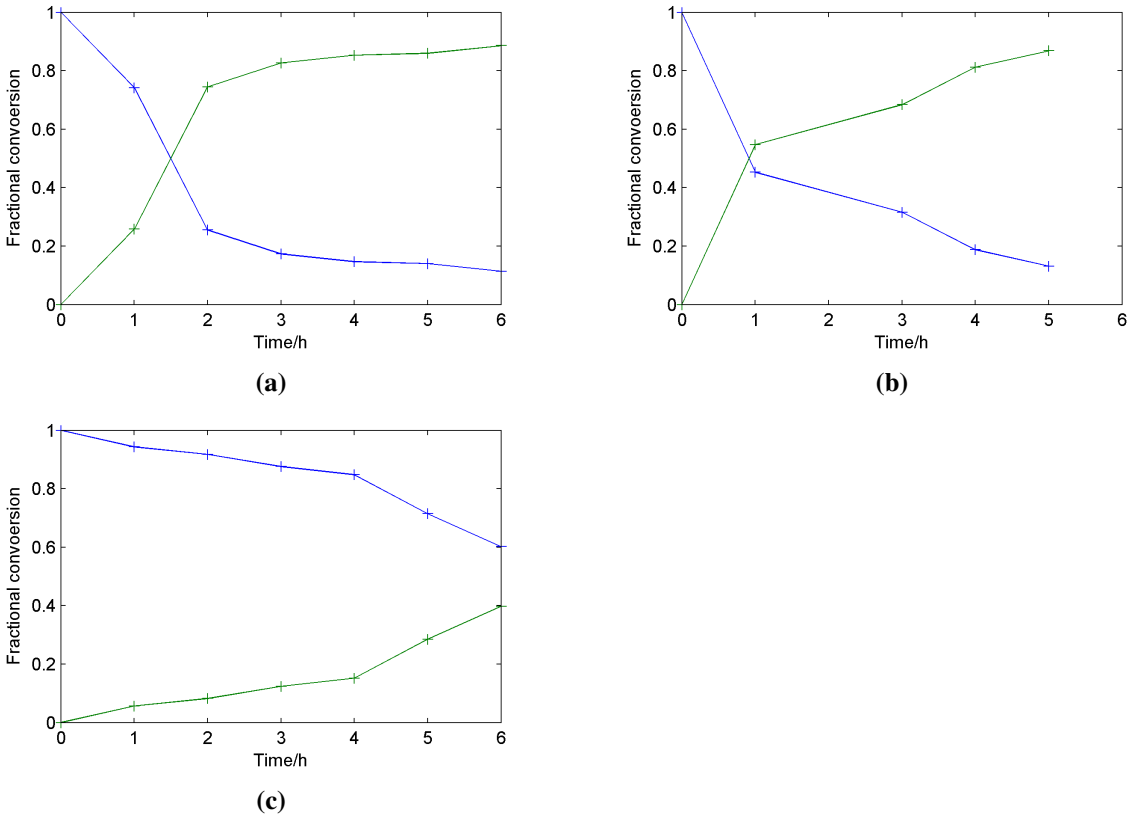


Fig. 2.13 Kinetic plots of thiophosphate addition reactions on p38α-Dha180 (blue) under varying buffer pH: **(a)** pH 10.0, **(b)** pH 8.0, **(c)** pH 6.0.

pH in the range of 6–10 were tried. Upon comparison, higher pH was found to give a faster reaction rate (Fig. 2.13). This is unsurprising, given that higher pH will result in an increased level of deprotonation of the thiol sulphur, thus making it a stronger nucleophile. The pH optimal for the chemical reaction however was very likely to cause the protein to be structurally unstable (Fig. 2.14),[81] which in turn would be detrimental to the activity of an enzyme. A change in reaction pH was therefore not adopted for further synthetic use.

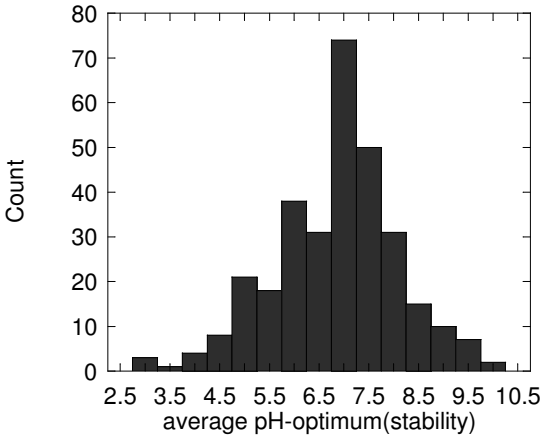
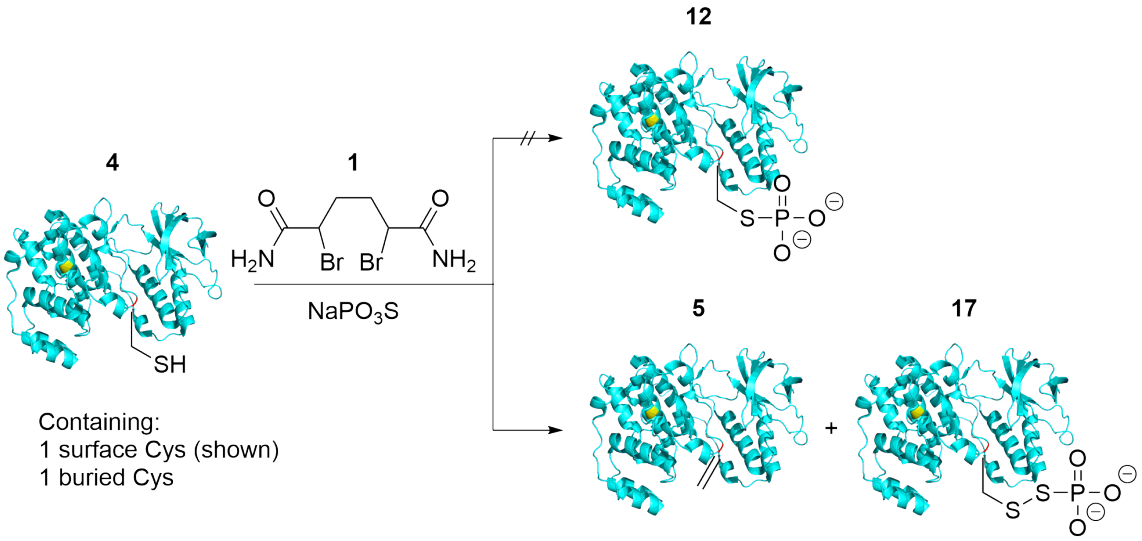


Fig. 2.14 Distribution of optimum pH for protein stability for a set of 380 proteins. Reprinted, by permission, from: Proteins 78: 2699–2706, Copyright © 2010 Wiley-Liss, Inc.[81]

2.8.5 One-pot Strategy for Thiophosphate Addition



Scheme 2.7 Simultaneous treatment of p38α variant **4** with dibromide **1** and sodium thiophosphate did not produce the desired installation of phosphocysteine, but resulted in a mixture of products.

A one-pot strategy for converting the unmodified p38α variant to the phosphocysteine product was also attempted, still with the aim of shortening total reaction time (Scheme 2.7). Using p38α variant **4** as the model, the p38α variant was treated simultaneously with dibromide **1** and sodium thiophosphate. However, rather than finding the thiophosphate exclusively reacting with any Dha product that was formed, the

thiophosphate also reacted directly with the unmodified p38α variant. The result was a mixture of Dha product and thiophosphate disulphide adduct **17** and in fact, no desired phosphocysteine product was observed (Fig. 2.15a). The formation of the thiophosphate disulphide adduct was confirmed upon treatment of unmodified p38α variant **4** with sodium thiophosphate (Fig. 2.15b). Thiophosphate addition onto Dha therefore cannot occur exclusively in the presence of an available cysteine thiol.

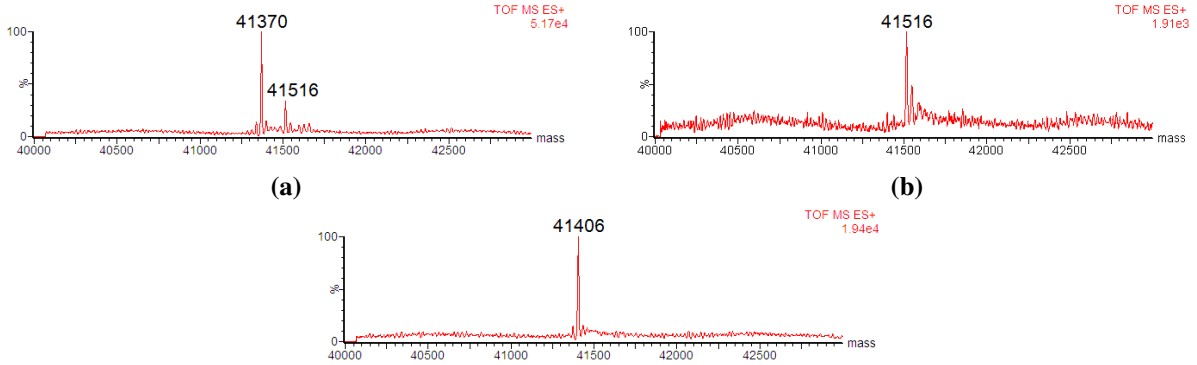


Fig. 2.15 Deconvoluted MS from the treatment of p38α variant **4**: **(a)** with dibromide **1** and sodium thiophosphate (one-pot reaction). Both the masses for the Dha product and thiophosphate disulphide adduct were observed. **(b)** with sodium thiophosphate only. **(c)** Deconvoluted MS of the unmodified **4** is shown for comparison. Expected mass change from unmodified **4** upon: Dha formation: -34 Da, thiophosphate addition onto Dha: $+80$ Da, disulphide formation with thiophosphate: $+112$ Da.

2.8.6 Summary of Phosphocysteine Formation Trails on p38α-C119S/C162S/T180C

Reaction trials of thiophosphate addition on Dha product **14** did not yield a more efficient way of adding this functionality to the protein, but it did more generally give scope to the reaction's tolerance to buffer and pH conditions. Instead, simply a longer reaction time (8 h) had to be adopted to get high conversion to the desired phosphocysteine product **16** (Fig. 2.16), which subsequent characterisation and testing showed that this longer reaction time was not associated with a loss in kinase activity or structural stability.

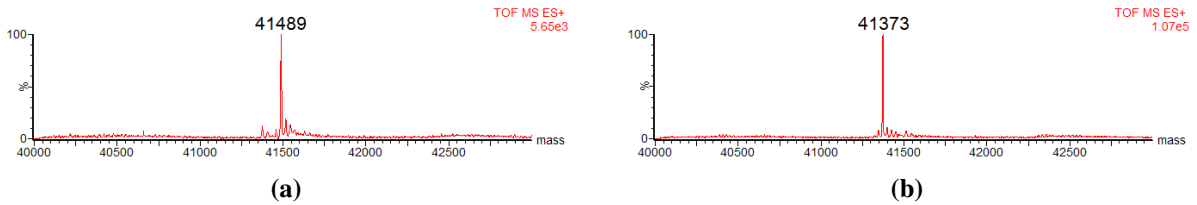


Fig. 2.16 **(a)** Deconvoluted MS spectrum showing high conversion to phosphocysteine product **16**, **(b)** MS of the starting Dha product **14** is shown for comparison (reproduced from Fig. 2.7a for convenience). Expected mass change from Dha product **14** upon thiophosphate addition: $+114$ Da.

2.9 Structural and Functional Characterisation of Modified p38α Variants

MS could only be used to determine that a single chemical modification had occurred on each of the p38α variants. Although indirect, there was also reasonable evidence that this modification had occurred at the desired site (Section 2.8.2). However, the structural and exact functionalised state of the chemically modified variants remained unprobed. Some of the questions that needed to be answered in this area were:

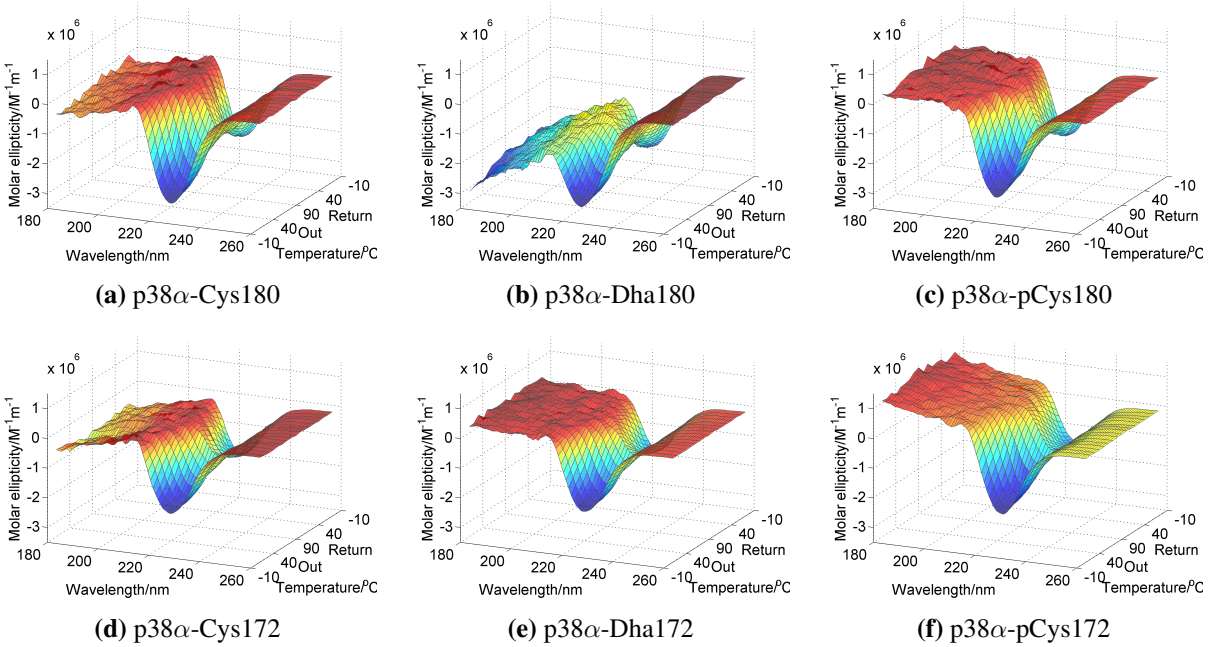
1. Whether the secondary structure of the p38α variants had been retained upon chemical modification.
2. Regarding their secondary structure, whether the different variants had responded similarly to their treatment to chemical modification, given that different variants were subjected to different reaction conditions, the alternative being that the prolonged reaction times had deleterious effects on secondary structure.
3. Whether direct evidence of modification at the site designed for chemical modification could be obtained.

Circular dichroism (CD) spectra were taken to address the first two points, while MS/MS was used to determine the third. Of the three p38α variants mentioned in the work until now, only two of them, namely p38α-Cys172 (**3**) and p38α-Cys180 (**13**) were taken for further study using the techniques mentioned here, as well as for the methods described in the next Chapter.

2.9.1 Circular Dichroism

Comparison of spectral profiles revealed no significant differences between the spectra of chemically modified derivatives (termed "chemical variants", a term which also includes the unmodified variant) of both p38α variants p38α-Cys172 and p38α-Cys180 (Fig. 2.17), suggesting that chemical modification had not significantly altered the proteins' secondary structure. In order to study the behaviour to thermal changes of the derivatives, the spectra were also collected over a range of temperatures, first ascending and then descending. This allowed the data to be analysed for the melting and subsequent annealing characteristics of all the p38α chemical variants. Knowing the protein melting temperature would then verify whether the reaction temperatures used were suitable for the protein, in that it would give an indication

of the temperature range which could have adverse effects on protein secondary structure and enzymatic function.



(a) p38α-Cys180 **(b)** p38α-Dha180 **(c)** p38α-pCys180

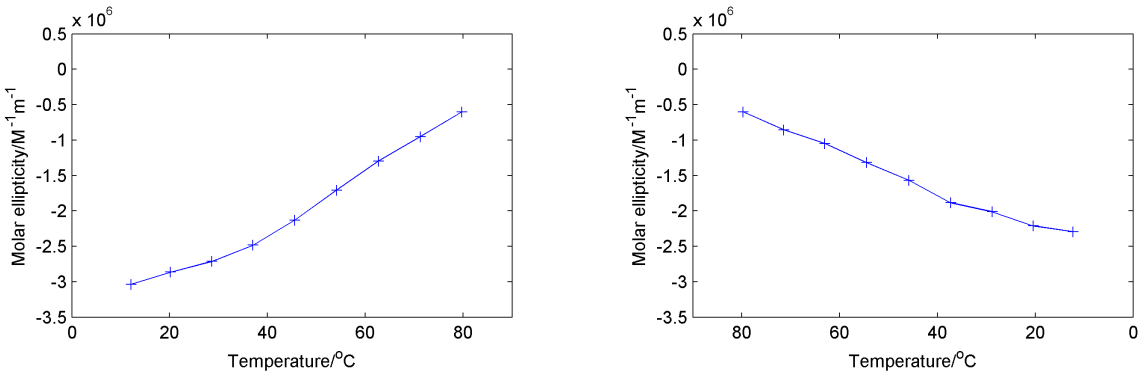
(d) p38α-Cys172 **(e)** p38α-Dha172 **(f)** p38α-pCys172

Fig. 2.17 CD spectra of all chemical variants of p38α-X180 and p38α-X172 (X = Cys, Dha or pCys), overall showing no significant changes between the unmodified variants (Cys variants, left) and phosphorylated variants (pCys variants, right).

Choosing the wavelength of maximum negative ellipticity at 10 °C as the wavelength for comparison, the temperature melt plots revealed a linear trend in both melting and annealing stages (Fig. 2.18). This is atypical of protein melt curves, which are usually sigmoidal.[82] Protein melting temperatures therefore could not be determined from this experiment. However, more general observations taken from the two curves showed that although there was a regain of ellipticity on protein cooling in the annealing curve (Fig. 2.18b), this regain was not complete when compared to the start of the melting curve (Fig. 2.18a), suggesting that the refolding of the protein ended in misfolding.

2.9.2 Melting Point Determination

Since melting temperature could not be determined from CD, differential scanning fluorimetry (DSF) with SYPRO orange dye was used to do this.[83,84] In this technique, the increase in temperature causes the melting protein to start exposing hydrophobic regions, which the dye binds to. The binding in turn causes the dye to fluoresce. Melting temperatures of 45–46 °C were found for the chemical variants derived from both p38α-Cys172 and p38α-Cys180, with the melting portion of the curve starting at 35–37 °C



(a) Melting curve **(b)** Annealing curve

Fig. 2.18 Representative melting and annealing curves of p38α-Cys180 taken using CD.

(Fig. 2.19). Since the reaction temperature is somewhat below the melting temperature, and is actually around the threshold temperature where melting starts, it was concluded that the different reaction times at 37 °C used between the two p38α variants were unlikely to cause significant differences on protein secondary structure degradation.

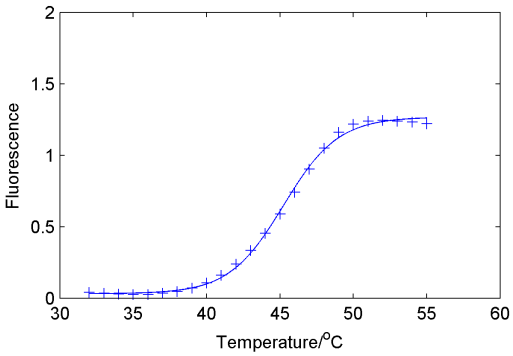


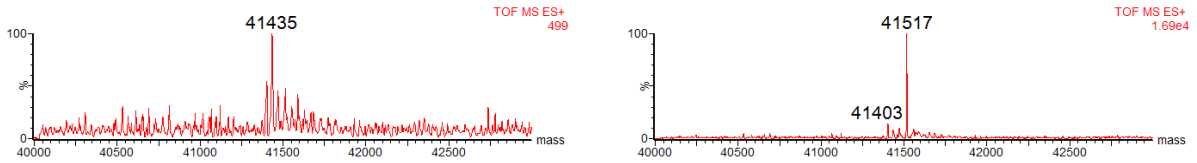
Fig. 2.19 Representative melting curve of p38α-Cys180 taken using DSF.

2.9.3 Establishing a Suitable Digest Protocol for MS/MS Analysis

To answer the question of whether the chemical modification had been made at the desired site, MS/MS detection of tryptically digested proteins was carried out. Using LC-MS/MS to detect the peptide containing the position of modification (position 172 or 180) would then give direct evidence of modification at this site.

In establishing a suitable digestion protocol, the protocol needed to be compatible with all the chemical derivatives of the p38α variants (Cys, Dha and pCys), so that the same protocol could be used on all these derivatives. Reported in-solution tryptic digestion protocols commonly involve the initial treatment of the protein of interest with DTT at 80–95 °C.[85,86] The purpose of these conditions are for the reduction of

disulphide bridges and the denaturation of the protein respectively. However, the use of a reducing agent such as DTT is not compatible with the Dha products of the p38α variants. These reducing agents can also act as nucleophiles and therefore add to the Dha. Furthermore, the use of such elevated temperatures for denaturation is not compatible with phosphocysteine, where the phosphate is labile. The labile nature of the phosphate on phosphocysteine was demonstrated on heating samples at 95 and 60 °C, where it was found that hydrolysis of the phosphate occurred at 95 °C (Fig. 2.20a), but not at 60 °C (Fig. 2.20b). Modifications to the standard protocol were therefore required.



(a) after heating at 95 °C **(b)** after heating at 60 °C

Fig. 2.20 Deconvoluted MS obtained on testing p38α-pCys172 for hydrolysis by heating. Expected mass change from p38α-pCys172 upon hydrolysis: -80 Da.

The digestion conditions were changed to ones that could be tolerated by all groups introduced through chemical modification. The temperature used for protein denaturation was therefore reduced. As the melting curves from DSF showed that the protein is fully melted beyond 55 °C (Fig. 2.19), 60 °C was used as the temperature for denaturation. Along similar lines, DTT was omitted from the digestion protocol. Observation of the crystal structure of p38α showed that the cysteines were distributed sparsely around the structure (Fig. 2.1), suggesting that no disulphide bonds were formed in the folding of this protein. Omission of DTT from the standard protocol was therefore deemed unlikely to have any detrimental effect. Upon testing variations on the digest protocol with and without DTT and using sequence coverage as the marker for protocol success, it was found that the omission of DTT from the protocol did not adversely impact sequence coverage (Table 2.1). Hence, the modifications in denaturation temperature and method were adopted for general use in digesting samples for MS/MS analysis on p38α. All the chemical variants derived from p38α-Cys172 and p38α-Cys180 were digested using this modified protocol.

2.9.4 LC-MS/MS Data Analysis

The digested samples were analysed by LC-MS/MS using collision-induced dissociation (CID) as the peptide fragmentation method, and the datasets subsequently analysed using MASCOT. Due to the number of datasets requiring analysis and the number of peptides of interest in each dataset ([3 cysteines in

Table 2.1 Sequence coverage obtained after LC-MS/MS analysis of samples prepared. Sections of protein sequence in black represent sections which have no peptide match from the MS/MS data. Top: with DTT for disulphide bridge reduction (56% coverage), Bottom: without DTT (69% coverage). p38α-pCys180 was used as the protein for the tests. Note: The numbering of the residues is off-set by 2 (i.e. 182 here corresponds to 180 used throughout the rest of this discussion) since the numbering has not been corrected for the uncleavable leader sequence from the *N*-terminal tag.

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES

  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

each p38α variant] $\times$ [1–3 possible modification states at each cysteine position]), the use of MASCOT facilitated the data analysis process in both being a tool able to handle the volume of data at hand, and in ensuring consistency of analysis between datasets.

Three peptides of interest per sample arise since not only is the interest in showing that the chemical modification has been incorporated at the correct position (either position 172 or 180, the positive results), but that there was no unwanted modification at the remaining cysteines (at Cys39 or Cys211 respectively, the negative results). In the negative results, analysis of data revealed that the native Cys39 was detected in all samples. This cysteine like all cysteine-containing peptides was detected as the carbamidomethyl derivative, since the free thiol had been capped in the digestion protocol to prevent disulphide formation. This practise is also common in many digestion protocols and for the purposes of this discussion, the detection of cysteine in MS/MS analysis refers to this capped state.[85,86] No significant levels of other chemical modifications were detected at this position and in most samples, no other modified peptides were detected at all. As for the positive results and peptides containing the site designed for chemical modification (both 172 and 180), MASCOT analysis detected Cys172 and Cys180 in the chemically unmodified variants

p38α-Cys172 and p38α-Cys180, while it found Dha172 and Dha180 respectively in the Dha derivatives. However, the presence of peptides containing pCys172 or pCys180 in the phosphocysteine derivatives were not detected, as was also the case with Cys211 in all datasets (Table 2.2).

Table 2.2 Detection status for all cysteine-containing peptides in MS/MS analysis, using CID as the fragmentation method. Key: ✓ = expected and detected, ✗ = expected but undetected, ✗ = not expected and not detected.

	Cys39			Cys172/180			Cys211		
Chemical Mutant	**CAM**	**Dha**	**pCys**	**CAM**	**Dha**	**pCys**	**CAM**	**Dha**	**pCys**
p38α-Cys172	✓	✗	✗	✓	✗	✗	✗	✗	✗
p38α-Dha172	✓	✗	✗	✗	✓	✗	✗	✗	✗
p38α-pCys172	✓	✗	✗	✗	✗	✗	✗	✗	✗
p38α-Cys180	✓	✗	✗	✓	✗	✗	✗	✗	✗
p38α-Dha180	✓	✗	✗	✗	✓	✗	✗	✗	✗
p38α-pCys180	✓	✗	✗	✗	✗	✗	✗	✗	✗

It should be noted that MASCOT detects peptides of interest by considering whether the peaks in MS/MS corresponding to a peptide could occur at random. Having more peaks in the MS/MS that correspond to a predicted peptide decreases the probably that the spectrum is a random observation. This probabilistic nature of peptide detection means that the presence of any peptide can never be completely ruled out. Thus, by stating that "no significant levels of other chemical modifications were detected", it is meant that there were MS/MS spectra that contained peaks corresponding to peptides chemically modified at Cys39, but these peaks could not be significantly differentiated from a random background.

2.9.5 MS/MS Detection of Phosphopeptides

MS/MS analysis so far had detected Dha at the site designed for chemical modification, but had not detected phosphocysteine. The labile nature of the phosphate on phosphocysteine during sample preparation or under the MS conditions was suspected to be the problem. Since a number of precautions had already been taken to prevent the phosphate of phosphocysteine from being hydrolysed during sample preparation, the fragmentation method used in MS/MS analysis was subsequently suspected at having prevented phosphocysteine detection. Loss of phosphate is precedented when certain MS fragmentation methods are used,[87,88] making some fragmentation methods more suited to phosphopeptide analysis than others. Other fragmentation methods (CID, ETD and HCD) were therefore tried (this work done in collaboration with Ritu Raj and Dr. Shabaz Mohammed). Analysis of the data collected using data-dependent decision tree (DDDT) allowed for a suitable, automated choice of fragmentation method (between CID and ETD)

for the MS2 analysis, for peptides already found in the MS. Analysis of the data collected using this method was then able to identify both the phosphocysteine-containing phosphopeptides at positions 172 and 180, corresponding to p38α-pCys172 and p38α-pCys180 respectively (Section 2.11.9). The detection of phosphocysteine-containing phosphopeptides therefore constitutes the full package of direct evidence that the chemical modification was installed at the desired position.

2.9.6 Searching for Cys211-containing peptide

Attempts were also made to detect peptides containing Cys211, either with the cysteine itself or if there were chemicals modification thereof (Dha or phosphocysteine). Prediction of tryptic digestion showed that the peptide containing Cys211 was 31 residues long (Table 2.3). As previous analyses had not identified a peptide containing Cys211, it was thus hypothesised that the length of this peptide could prevent it from being readily ionised in LC-MS/MS analysis. Using p38α-Cys180 as the test subject for digestion, the sample preparation protocol was modified to vary the digestion protease used (Asp-N, Glu-C and pepsin were tried). MASCOT analysis of LC-MS/MS datasets from these samples yielded reasonable sequence coverage in all cases, but none of the datasets revealed a Cys211-containing peptide (Table 2.4).

Table 2.3 Peptides containing Cys211 after digestion of p38α using various proteases.

Enzyme	Cys211-Containing Peptide	Peptide Length
Trypsin	APEIMLNWMHYNQTVDIWSVGCIMAELLTGR	31
Pepsin (at pH 1.3)	SVGCIMAE	8
Asp-N	DIWSVGCIMAELLTGRTLFPGT	22
Glu-C	IMLNWMHYNQTVDIWSVGCIMAE	23

Changing the MS/MS fragmentation method also failed to detect Cys211-containing peptide. Finally, the last thing that could be changed in the protocol was the use of liquid chromatography (LC). As this peptide of interest is not only long but lipophilic, it was postulated that the peptide could adhere strongly onto the chromatography column of the LC system and therefore never reach the MS for detection. The peptide's lipophilic nature is illustrated in the crystal structure, where the peptide comes from a buried α-helix in the helix bundle (Fig. 2.21). To overcome the burden of LC, MALDI (done in collaboration with Inga Pfeffer) was used, thus avoiding any use of LC, but again, this method was unable to identify even a parent ion corresponding to the $[M+H]^+$ mass (Fig. 2.22).

In conclusion, the peptide containing Cys211 could not be detected. The main purpose of this peptide's detection had been to directly show that unwanted chemical modification had not occurred at this position.

Table 2.4 Sequence coverage obtained after LC-MS/ MS analysis of samples prepared by digestion with different proteases. Cys211 is underlined. Section of protein sequence in black represent sections which have no peptide match from the MS/MS data. From top to bottom: **(a)** with trypsin, **(b)** with pepsin, **(c)** with Asp-N, **(d)** with Glu-C. The coverage shows that Cys211-containing peptide was identified in the trypsin digested sample, however, the data reveals that this is insignificant. Note: the numbering of the residues is off-set by 2 (i.e. 182 here corresponds to 180 used throughout the rest of this discussion) since the numbering has not been corrected for the uncleavable leader sequence from the *N*-terminal tag.

(a)

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

(b)

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLCRHTDDE MTGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

(c)

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

(d)

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

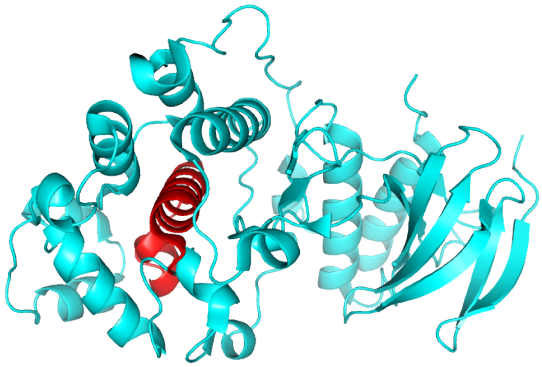


Fig. 2.21 Crystal structure of p38α kinase highlighting the Cys211-containing peptide (red), which is completely embedded. Drawn from 1R3C.pdb.[73]

Without the identification of the peptide, there is no direct evidence that this position was not chemically modified. There is however much indirect evidence:

1. Peptides from the other native cysteine (Cys39) were detected, where there was no unwanted chemical modification.

2. Whole protein MS showed that only a single modification event had taken place over the protein.

3. In the (hypothetical) situation where there is modification at Cys211, there would have to be a sub-population of protein that is only modified at position 211. This sub-population would be in addition to the sub-population of protein modified at the desired position (172 or 180), for which there is direct MS/MS evidence. Only then would the previous point be satisfied. This situation would be statistically unlikely.

It was thus concluded that the chemical modifications had been installed only in the desired positions of 172 and 180 on the p38α variants p38α-Cys172 and p38α-Cys180, without any off-target modification.

2.10 Summary

The design, execution and evidence of chemical modification on p38α kinase have been described in this chapter. The initial focus was on expanding the substrate scope of the "tag-and-modify" chemical modification methodology. This modification strategy uses Dha as a reactive intermediate, which can then be further functionalised. The expansion in scope was to apply this chemistry to protein kinases. As the modification method relies on cysteine, the emphasis in designing p38α variants was on making them "cysteine-free". Two distinct cysteine-free variants were therefore studied, p38α variants **3** and **4**.

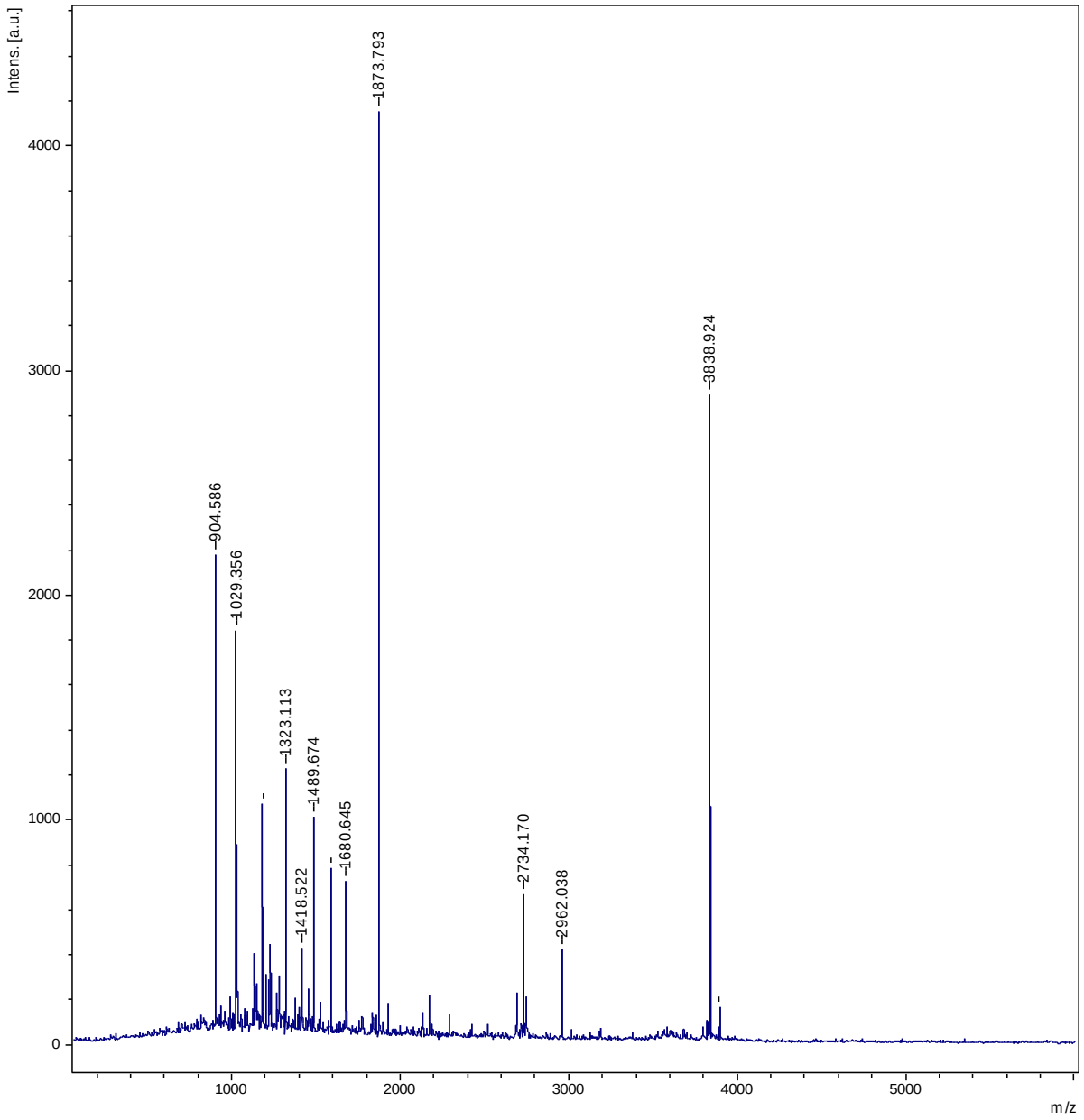


Fig. 2.22 MALDI MS spectrum of a tryptically digested sample of p38α-Cys180. No peak corresponding to the Cys211-containing peptide (monoisotopic mass: 3648.746) was observed.

Testing of these variants revealed that stringent exclusion of native cysteine was not required, leading to the retention of two of the four cysteines native to p38α. The variant p38α-C119S/C162S/A172C (variant **3**, designated as p38α-Cys172) was hence adopted as the first major variant for further study.

Turning attention to creating chemically modified variants of more biological relevance, attention then turned to a second major p38α variant, p38α-Cys180, derived from p38α-Cys172. The change in reaction site was accompanied by a change in cysteine reactivity, prompting a more in-depth investigation into the tolerances to reaction conditions of the two main chemical reactions used, namely Dha formation by bis-addition–elimination with dibromide **1**, and thiophosphate addition onto Dha. While choosing the optimal reaction conditions, efforts were also taken to safeguard the enzymatic integrity of the p38α variant being chemically modified. Optimisation of the chemistry was therefore two-fold: both for the high conversion of the chemical reaction, as well as for the stability and structural integrity of the protein. Efforts were also taken to better understand the chemical modification reactions and the scope over which they could operate in. This process eventually led to the high conversion to products where phosphocysteine had been installed once onto each of the p38α variants.

Further evidence was required to show that this single-site installation of thiophosphate was at the intended position, and that the proteins' structures had indeed been retained. The required evidence was obtained through MS/MS analysis, and CD and DSF studies respectively. It could therefore be concluded that p38α variants have been chemically modified as their design had intended, and that the desired products p38α-pCys172 and p38α-pCys180 had been obtained: the two phosphocysteine derivatives.

In the next chapter, the biological function and activity of p38α-pCys172 and p38α-pCys180 will be explored, and the optimisation of the biology and the assays used to demonstrate the activity of p38α will be explained and detailed.

2.11 Experimental

2.11.1 General Measures

Synthetic Chemistry

Infrared spectra were recorded using a Bruker TENSOR 27 spectrometer. Nuclear magnetic resonance (NMR) spectra were recorded using a Bruker AV-400 spectrometer, proton NMR (^{1}H NMR) at 400 MHz

and carbon NMR (^{13}C NMR) at 101 MHz. Chemical shifts (δ) are reported in parts per million (ppm) and are referenced relative to the residual proton-containing solvent (^{1}H NMR: 2.52 ppm for DMSO; ^{13}C NMR: 41.2 ppm for DMSO). Accurate mass spectra (ESI) were recorded using a Micromass Bruker MicroTOF instrument. TLC was performed using aluminium plates pre-coated with Merck Kieselgel 60 F254. The plates were visualised using either ultraviolet (UV) light or potassium permanganate staining, as appropriate. All other chemicals were bought from common suppliers and used as purchased.

Molecular Biology

Media and Bacterial Strains: All bacterial handling was done in a sterile environment, either within close proximity to a Bunsen burner flame or inside a biological safety cabinet. Samples of p38α mutant plasmid constructs p38α-C119S/C162S/A172C and p38α-C39A/C119S/C162S/A172C were kindly sent by Dr. Richard Bazin and co-workers, Pfizer, Sandwich, Kent. LB was bought as pre-formulated dry granules (Melford) and diluted according to the manufacturer's specifications. SOC and NZY^{+} media were prepared from biological laboratory reagents. 2YT medium was bought as pre-formulated dry granules (Sigma Aldrich) and diluted according to the manufacturer's specifications. All other biological laboratory buffer and media reagents were bought from common suppliers and used as purchased. BL21(DE3) and NovaBlue *E. coli* chemically competent cells (Novagen) were handled according to the manufacturer's instructions, thawing on ice before incubation with DNA. NovaBlue was used for cloning, while BL21(DE3) was used for protein expression. XL1-Blue *E. coli* chemically competent cells (Agilent) likewise were handled according to the manufacturer's instructions. XL1-Blue was used for cloning.

Mutagenesis and Bioinformatics: Site-directed mutagenesis was done using *Pfu*Ultra DNA polymerase AD (Agilent), dNTPs (Sigma Aldrich) and DpnI (NEB). Primers for mutagenesis were designed using the MutaPrimer design function in SimVector 4 from Premier Biosoft, which follows the guidelines given by Stratagene. The sequences were then sent to Life Technologies (Invitrogen) for custom oligonucleotide synthesis. Plasmid purification was done by using the QIAprep® Miniprep Kit (Qiagen) according to the manufacturer's recommendations. LyseBlue reagent had been added to buffer P1 from the Miniprep kit to aid identification of cell lysis. DNA and protein concentration determination was done using a NanoDrop 1000 spectrophotometer (Thermo Scientific), which allowed the calculation of yield. The wavelengths used were a ratio of 260/230 nm and 280/260 nm respectively. The extinction coefficients for proteins

were calculated using the ProtParam tool (ExPASy).[89,90] Other bioinformatics tools from the ExPASy website for various applications were also used.

Protein Expression and Purification: Fast protein liquid chromatography (FPLC) was done using an ÄKTA Purifier™ system (Amersham Biosciences, now part of GE Healthcare). Pre-packed columns for FPLC and G-25 desalting columns (SpinTrap™, MiniTrap™ and MidiTrap™) were purchased from GE Healthcare. Unless otherwise stated, the analytes were filtered before loading onto the column for each chromatographic step. Dialysis was performed using Slide-A-lyser® MWCO 10 000 dialysis cassettes (Thermo Scientific). Concentration of protein was performed using Vivaspin 500, 15 or 6, MWCO 10 000 spin columns (Sartorius Stedim Biotech) as appropriate. Autoclaving was performed at 121 °C for 10–15 min. Solutions which were used as media or buffers and which could not be autoclaved were filtered through either 0.2 µm cellulose acetate filters, 0.2 µm polyamide filters, or 0.2 µm Minisart® filter cartridges (Sartorius Stedim Biotech), as appropriate. In particular, all filtrations of buffers, solutions and lysates involved in protein purification or chemical reactions involving proteins were done through 0.2 µm filters. Filtration of volumes of <2 mL (for HPLC) was done using 0.2 µm spin filters (Corning). HPLC samples were otherwise clarified by centrifugation (13 200 rpm, 10–20 min, 4 °C). Microcentifugation was performed using an Eppendorf Centrifuge 5415R. Shaking incubation was performed using either a New Brunswick Scientific Innova® incubator shaker or a New Brunswick Scientific Model G25 incubator shaker. Plate incubation was performed using a Heraeus B6030 incubator. Low speed centrifugation of large volumes (>2 mL) was performed using a Beckman Coulter Allegra™ X-12R centrifuge. High speed centrifugation of large volumes (>2 mL) was performed using a Beckman Coulter Avanti™ J-25 centrifuge. Sonication was performed using a Sanyo Soniprep 150 sonicator, equipped with a microtip.

SDS-PAGE protein analysis was done using NuPAGE® gels (Invitrogen). The samples (5 µL) were prepared by mixing with equal volumes of 2× SDS gel-loading buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 0.2% bromophenol blue, 20% glycerol), with β-mercaptoethanol (200 mM) being added to the buffer just before use. The samples were mixed by vortexing and heated at 95 °C for 10 min. The samples were run on 10% bis-tris gels as appropriate. The gels were run in MOPS running buffer (50 mM MOPS pH 7.7, 50 mM Tris base, 0.1% SDS, 1 mM EDTA) at 200 V. Samples (5–9 µL) were loaded as appropriate. Staining of the gels was done with Instant Blue™ Coomassie® stain for 12–18 h and destained with MilliQ water for a minimum of 18 h.

Solid LB agar medium was prepared by suspending LB granules (6.25 g) and agar (3.75 g) in MilliQ water (250 mL). The suspension was autoclaved as previously described to give a solution which once sufficiently cooled, was supplemented with the appropriate antibiotic (ampicillin: final concentration: 100 µg/mL) and poured into petri dishes (10–20 mL per plate).

Glycerol Stocks: For cloning strains (NovaBlue XL1-Blue), glycerol stocks were made by adding bacterial broth (750 µL) to glycerol solution (250 µL of a 60% stock solution in water) and stored directly at −80 °C. For expression strains (BL21(DE3)), glucose (28 µL of a 2 M stock solution in water) was added as a supplement before adding the bacterial broth (722 µL) and stored as above.

Gene Sequence Analysis

DNA sequencing was done by Geneservice, Oxford using the primers pGEX forward (5′−GGGCTGGCA AGCCACGTTTGGTG−3′) and pGEX reverse (5′−CCGGGAGCTGCATGTGTCAGAGG−3′) for genes contained in pGEX vectors. The forward and reverse complimentary sequences returned from the gene sequencing service were analysed by sequence alignment of the two returned sequences, followed by base-by-base string comparison to the desired reference sequence. Sequence data manipulation and comparison was done using scripts written in MATLAB. In the case of serious mismatching between observed and desired sequences, the raw data were visualised and evaluated using FinchTV version 1.4.0. Plasmid sequences displayed are consensus sequences spliced together from the results of a pGEX forward/pGEX reverse pair read. Examples of the scripts and functions used are in Appendix A.

Protein Modification Reactions

Sodium thiophosphate was bought from Sigma Aldrich and used as purchased. Shaking incubation of the reactions was performed using PHMT-PSC-18 from Grant Bio. Incubation without shaking was done in a waterbath at the appropriate temperature. LC-MS analysis was done on a Micromass LCT Premier instrument using a ProSwift® RP-4H monolith, 1×50 mm column (Dionex). A linear gradient was run from 5–95% of solvent mixture B (1% formic acid in acetonitrile) into solvent mixture A (1% formic acid in water). The gradient was run with a flow rate of 0.4 mL/min over 4 min for the ProSwift column. Modified proteins used in subsequent reactions had their yield determined by volume measurement using a Gilson-type pipette, and concentration measurement by UV absorbance as described previously, making the assumption that the functional group incorporated in the chemical reaction did not have a large effect

on the extinction coefficient as predicted for the protein's unmodified sequence. In determining the protein concentration in reaction mixtures during reaction monitoring, the samples of the reaction mixtures were first filtered and compared against a blank "reaction" (the reaction buffer with the other reagents, but no protein), which was also filtered. Analysis of the MS data collected was done using MassLynx version 4.1. Deconvolution of the protein ion series was done using the in-built maximum entropy algorithm with calibration of the MS done with equine myoglobin from heart.

Quantification of MS Data

LC-MS datasets were processed using MassLynx as described above. Data from the deconvoluted spectra were output from MassLynx as a spectrum list for quantitative analysis using MATLAB R2012b.

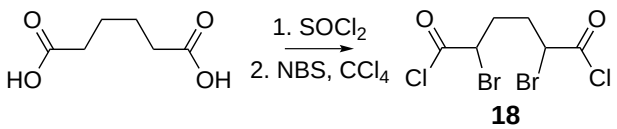
In MATLAB, quantities of the species expected to be present in the sample were estimated from relative peak intensity. The spectrum baseline was set at the median of all the values in the spectrum list. The peaks were picked by taking the maximum intensity values in the ranges where the peaks corresponding to the species of interest were expected. The proportion of each species was then calculated for each timepoint and plotted accordingly. Examples of related scripts for MS kinetic data analysis are given in Appendix B.

Enzymatic Reactions

All proteolytic enzymes were bought from common suppliers and used as per manufacturer's instructions. All other chemicals and buffer components were bought from common suppliers and used as purchased.

2.11.2 Synthesis of Protein Modification Reagents

Preparation of 2,5-dibromoadipoyl dichloride

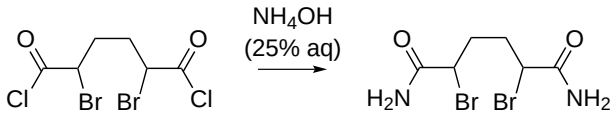


18

Adipic acid (10.00 g, 68.4 mmol, 1.0 eq.) was suspended in thionyl chloride (30 mL, 414 mmol, 6.0 eq.) and the mixture heated under reflux at 80 °C under a drying tube of $CaCl_2$ in air. After 30 min heating, all the adipic acid had dissolved. Heating was continued for a further 1.5 h, after which the mixture

was returned to room temperature and diluted with carbon tetrachloride (40 mL). *N*-bromosuccinimide (24.27 g, 136 mmol, 2.0 eq.) and conc. HBr (4 drops of a 36% solution in water) were subsequently added and the mixture returned to reflux at 80 °C for 2 h. As the reaction progressed, the reaction mixture first turned red and red vapour was observed and after 1 h, transiently lost its colour before turning black. The mixture was then cooled to 0 °C and allowed to stand for 30 min before being filtered and washed with Et_2O (40 mL). The filtrate was concentrated *in vacuo* to give *2,5-dibromoadipoyl dichloride* (**18**) as a dark red oil. The product was not further purified or characterised but directly subjected to the next reaction.

Preparation of 2,5-dibromo-adipic acid diamide



18 **1**

To a stirred solution of NH_4OH (80 mL of a 25% solution in water, 1180 mmol, 17.2 eq.) at 0 °C was added dropwise the crude acid chloride **18** over 25 min. The mixture was allowed to stir at 0 °C for 1 h before being filtered to give a dark green precipitate and a dark filtrate. The precipitate was titurated by suspending in 1:1 methanol:H_2O (80 mL) and heating the mixture at 60 °C for 30 min before further filtration. The precipitate was further washed with methanol (80 mL) and dried to give *2,5-dibromo-adipic acid diamide* (**1**, 11.281 g, 55%) as a white powder: ν_{max} (film) 3196, 1666 (C=O), 1606, 1418, 1321, 1220 cm^{-1}; ^{1}H NMR (400 MHz, DMSO) δ ppm 7.71 (2 H, s, NH_2), 7.30 (2 H, br s, NH_2), 4.29–4.36 (2 H, m, 2×CHBr), 1.76–2.10 (4 H, m, CH_2CH_2); ^{13}C NMR (101 MHz, DMSO) δ ppm (both diastereomers reported) 170.84, 170.78 (C=O), 49.2, 48.9 (CH_2*C*HBr), 33.4, 33.3 (*C*H_2CHBr); *m/z* (LRMS: ESI+) 327, 325 ($[M+Na]^+$); in agreement with previous data.[43]

2.11.3 Protein Expression and Purification

Cloning of pGEX-2T-p38α-Cys172 and pGEX-2T-p38α-C39A/C119S/C162S/A172C Plasmids

Transformation: Plasmid DNA of pGEX-2T-p38α-C119S/C162S/A172C (referred to as p38α-Cys172) (1 µL) and pGEX-2T-p38α-C39A/C119S/C162S/A172C was transformed into NovaBlue *E. coli* chemical competent cells (25 µL) by heat shock. The DNA was incubated with the thawed cells on ice for 5 min before heat shock was performed at 42 °C for 30 s, returning immediately on ice. After incubating on ice

for a further 2 min, the cells were fed with SOC medium (125 µL) before incubation at 37 °C, 250 rpm for 15 min.

Bacterial Culturing: The cells were plated onto LB agar medium supplemented with ampicillin (100 µg/mL) at different cell concentrations (80 µL, 50 µL and 20 µL + 20 µL SOC medium) and incubated at 37 °C for 14 h. All plates had bacterial lawns growing on them with small numbers of distinct colonies growing on plates of lowest cell concentration. 2 colonies were picked and grown up further in liquid LB medium (2×10 mL) supplemented again with ampicillin (100 µg/mL) in a shaking incubator at 37 °C, 250 rpm for 17 h.

DNA Purification and Storage: After glycerol stocks were made, the remaining cells were then pelleted by centrifugation (3000 rpm, 20 min, 4 °C). The resulting cell pellet was resuspended in P1 buffer and the plasmid DNA purified out using the QIAprep® Miniprep kit. The plasmid DNA was eluted in EB buffer and stored at −20 °C. DNA sequencing confirmed that the sequences of the plasmid samples received were as expected.

Sequence of pGEX-2T- p38α-C119S/C162S/A172C (pGEX-2T-p38α-Cys172):

```
   1 ATG AGC CAG GAA CGT CCG ACC TTT TAT CGT CAG GAA CTG AAT AAA
  46 ACC ATT TGG GAA GTG CCG GAA CGT TAT CAG AAT CTG TCT CCG GTT
  91 GGT AGC GGT GCA TAT GGT AGC GTT TGT GCA GCA TTT GAT ACC AAA
 136 ACC GGT CTG CGT GTT GCA GTT AAA AAA CTG AGC CGT CCG TTT CAG
 181 AGC ATT ATT CAT GCC AAA CGT ACC TAT CGT GAA CTG CGT CTG CTG
 226 AAG CAT ATG AAA CAT GAA AAT GTG ATT GGT CTG CTG GAT GTT TTT
 271 ACA CCG GCA CGT AGC CTG GAA GAG TTT AAT GAT GTG TAT CTG GTG
 316 ACA CAT CTG ATG GGT GCA GAT CTG AAT AAT ATT GTG AAA AGC CAG
 361 AAA CTG ACC GAT GAT CAT GTG CAG TTC TTA ATC TAT CAG ATT CTG
 406 CGT GGC CTG AAA TAT ATT CAT AGC GCA GAT ATT ATT CAT CGT GAT
 451 CTG AAA CCG AGC AAT CTG GCA GTT AAT GAA GAT AGC GAA CTG AAA
 496 ATT CTG GAT TTT GGT CTG TGT CGT CAT ACC GAT GAT GAA ATG ACC
 541 GGT TAT GTT GCA ACC CGT TGG TAT CGT GCA CCG GAA ATT ATG CTG
 586 AAT TGG ATG CAT TAT AAT CAG ACC GTG GAT ATT TGG AGC GTT GGT
 631 TGT ATT ATG GCA GAA CTG CTG ACC GGT CGT ACC CTG TTT CCG GGT
 676 ACA GAT CAT ATT GAT CAG CTG AAA CTG ATT CTG CGT CTG GTT GGT
 721 ACA CCG GGT GCC GAA CTG CTG AAA AAA ATT AGC AGC GAA AGC GCA
 766 CGC AAT TAT ATT CAG AGC CTG ACC CAG ATG CCG AAA ATG AAT TTT
 811 GCC AAT GTG TTT ATT GGT GCA AAT CCG CTG GCA GTT GAT CTG CTG
 856 GAA AAA ATG CTG GTT CTG GAT AGC GAT AAA CGT ATT ACC GCA GCA
 901 CAG GCA CTG GCA CAT GCA TAT TTT GCC CAG TAT CAT GAT CCG GAT
 946 GAT GAA CCG GTT GCA GAT CCG TAT GAT CAG AGC TTT GAA AGC CGT
 991 GAT CTG CTG ATT GAT GAA TGG AAA AGC CTG ACC TAT GAT GAA GTG
1036 ATT AGC TTT GTT CCG CCT CCA CTG GAT CAA GAA GAA ATG GAA AGC
1081 TAA
```

Sequence of pGEX-2T- p38α-C39A/C119S/C162S/A172C:

```
   1 ATG AGC CAG GAA CGT CCG ACC TTT TAT CGT CAG GAA CTG AAT AAA
  46 ACC ATT TGG GAA GTG CCG GAA CGT TAT CAG AAT CTG TCT CCG GTT
  91 GGT AGC GGT GCA TAT GGT AGC GTT GCA GCA GCA TTT GAT ACC AAA
 136 ACC GGT CTG CGT GTT GCA GTT AAA AAA CTG AGC CGT CCG TTT CAG
 181 AGC ATT ATT CAT GCC AAA CGT ACC TAT CGT GAA CTG CGT CTG CTG
 226 AAG CAT ATG AAA CAT GAA AAT GTG ATT GGT CTG CTG GAT GTT TTT
 271 ACA CCG GCA CGT AGC CTG GAA GAG TTT AAT GAT GTG TAT CTG GTG
 316 ACA CAT CTG ATG GGT GCA GAT CTG AAT AAT ATT GTG AAA AGC CAG
 361 AAA CTG ACC GAT GAT CAT GTG CAG TTC TTA ATC TAT CAG ATT CTG
 406 CGT GGC CTG AAA TAT ATT CAT AGC GCA GAT ATT ATT CAT CGT GAT
 451 CTG AAA CCG AGC AAT CTG GCA GTT AAT GAA GAT AGC GAA CTG AAA
 496 ATT CTG GAT TTT GGT CTG TGT CGT CAT ACC GAT GAT GAA ATG ACC
 541 GGT TAT GTT GCA ACC CGT TGG TAT CGT GCA CCG GAA ATT ATG CTG
 586 AAT TGG ATG CAT TAT AAT CAG ACC GTG GAT ATT TGG AGC GTT GGT
 631 TGT ATT ATG GCA GAA CTG CTG ACC GGT CGT ACC CTG TTT CCG GGT
 676 ACA GAT CAT ATT GAT CAG CTG AAA CTG ATT CTG CGT CTG GTT GGT
 721 ACA CCG GGT GCC GAA CTG CTG AAA AAA ATT AGC AGC GAA AGC GCA
 766 CGC AAT TAT ATT CAG AGC CTG ACC CAG ATG CCG AAA ATG AAT TTT
 811 GCC AAT GTG TTT ATT GGT GCA AAT CCG CTG GCA GTT GAT CTG CTG
 856 GAA AAA ATG CTG GTT CTG GAT AGC GAT AAA CGT ATT ACC GCA GCA
 901 CAG GCA CTG GCA CAT GCA TAT TTT GCC CAG TAT CAT GAT CCG GAT
 946 GAT GAA CCG GTT GCA GAT CCG TAT GAT CAG AGC TTT GAA AGC CGT
 991 GAT CTG CTG ATT GAT GAA TGG AAA AGC CTG ACC TAT GAT GAA GTG
1036 ATT AGC TTT GTT CCG CCT CCA CTG GAT CAA GAA GAA ATG GAA AGC
1081 TAA
```

General Procedure for Mutagenesis of pGEX-2T-p38α-Cys172 Plasmid

Mutagenesis Primers: The following primers were used to further generate p38α-C119S/C162S/T180C (referred to as p38α-Cys180):

- C172A forward: 5′−GGATTTTGGTCTGGCGCGTCATACCGATGATGAAATGACCG−3′
- C172A reverse: 5′−CGGTCATTTCATCATCGGTATGACGCGCCAGACCAAAATCC−3′
- T180C forward: 5′−CCGATGATGAAATGTGCGGTTATGTTGCAACCCGTTGG−3′
- T180C reverse: 5′−CCAACGGGTTGCAACATAACCGCACATTTCATCATCGG−3′

Mutagenesis Reaction: The mutagenesis reactions were made up by adding the following components in the given order:

Substance	Concentration	Quantity/µL
dNTPs	10 mM each	0.25
forward primer	12.5 µM	0.25
reverse primer	12.5 µM	0.25
MilliQ water	–	9.625 (10.25 if no DMSO was used)
Pfu DNA polymerase	2.5 U/µL	0.25
10× reaction buffer	–	1.25
DMSO (optional)	–	0.63
plasmid	10 or 50 ng/µL	0.25
final volume		12.5

In this way, 2 reactions with different DNA template concentrations were run for each mutation made. The reactions were then heated in a thermal cycler using the programme:

Segment	No. of Cycles	Temp./°C		Time
1	1	95		30 s
2	16	95	denaturation	30 s
		55	annealing	1 min
		68	extension	15 min
3	1	4		∞

After this time, DpnI (0.25 µL of a 10 U/µL solution) was added and the mixture incubated for 1–2 h at 37 °C. The extent of DNA synthesis was determined by running samples (2.5 µL) of the reaction mixtures on a 0.8% agarose gel supplemented with ethidium bromide (5.3 µL of a 1% stock solution) in TAE buffer (50 mL) with the gel run at 150 V in TAE buffer. The samples were loaded with 6× DNA loading dye (0.5 µL). The gel was visualised under a UV lamp and only those reactions where plasmid DNA was observed were transformed into *E. coli*. Where there were 2 reactions containing the same template/primer combination, only the reaction with lower template concentration was subsequently used.

Transformation: The selected reaction mixtures were transformed into XL1-Blue *E. coli* supercompetent cells by heat shock. Reaction mixture (1 µL) was added to aliquots of freshly thawed cells (25 µL) and incubated on ice for 5 min. Heat shock was then performed at 42 °C for 40 s before returning the cells onto ice and incubating for a further 2 min. The cells were fed with NZY$^+$ broth (250 µL) and incubated in a shaker incubator at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The cells were plated onto LB agar supplemented with ampicillin (100 µg/mL). The plates were incubated at 37 °C for 21–24 h. 3 colonies of each construct were then selected from the

plates and cultured further in LB medium (3×10 mL), again supplemented with ampicillin (100 µg/mL) and the cultures incubated for a further 13–22 h in a shaker incubator at 37 °C, 250 rpm.

DNA Purification and Storage: Glycerol stocks of the cultures were made before cells were harvested by centrifugation (3750 rpm, 10 min, 4 °C) of the cultures and the supernatant decanted off. The cells were resuspended in P1 buffer for plasmid purification by Miniprep, where the manufacturer's protocol was used. DNA sequencing confirmed the colonies containing the desired constructs, which were selected for either expression or further rounds of mutagenesis.

Sequence of pGEX-2T- p38α-C119S/C162S/T180C:

```
   1 ATG AGC CAG GAA CGT CCG ACC TTT TAT CGT CAG GAA CTG AAT AAA
  46 ACC ATT TGG GAA GTG CCG GAA CGT TAT CAG AAT CTG TCT CCG GTT
  91 GGT AGC GGT GCA TAT GGT AGC GTT TGT GCA GCA TTT GAT ACC AAA
 136 ACC GGT CTG CGT GTT GCA GTT AAA AAA CTG AGC CGT CCG TTT CAG
 181 AGC ATT ATT CAT GCC AAA CGT ACC TAT CGT GAA CTG CGT CTG CTG
 226 AAG CAT ATG AAA CAT GAA AAT GTG ATT GGT CTG CTG GAT GTT TTT
 271 ACA CCG GCA CGT AGC CTG GAA GAG TTT AAT GAT GTG TAT CTG GTG
 316 ACA CAT CTG ATG GGT GCA GAT CTG AAT AAT ATT GTG AAA AGC CAG
 361 AAA CTG ACC GAT GAT CAT GTG CAG TTC TTA ATC TAT CAG ATT CTG
 406 CGT GGC CTG AAA TAT ATT CAT AGC GCA GAT ATT ATT CAT CGT GAT
 451 CTG AAA CCG AGC AAT CTG GCA GTT AAT GAA GAT AGC GAA CTG AAA
 496 ATT CTG GAT TTT GGT CTG GCG CGT CAT ACC GAT GAT GAA ATG TGC
 541 GGT TAT GTT GCA ACC CGT TGG TAT CGT GCA CCG GAA ATT ATG CTG
 586 AAT TGG ATG CAT TAT AAT CAG ACC GTG GAT ATT TGG AGC GTT GGT
 631 TGT ATT ATG GCA GAA CTG CTG ACC GGT CGT ACC CTG TTT CCG GGT
 676 ACA GAT CAT ATT GAT CAG CTG AAA CTG ATT CTG CGT CTG GTT GGT
 721 ACA CCG GGT GCC GAA CTG CTG AAA AAA ATT AGC AGC GAA AGC GCA
 766 CGC AAT TAT ATT CAG AGC CTG ACC CAG ATG CCG AAA ATG AAT TTT
 811 GCC AAT GTG TTT ATT GGT GCA AAT CCG CTG GCA GTT GAT CTG CTG
 856 GAA AAA ATG CTG GTT CTG GAT AGC GAT AAA CGT ATT ACC GCA GCA
 901 CAG GCA CTG GCA CAT GCA TAT TTT GCC CAG TAT CAT GAT CCG GAT
 946 GAT GAA CCG GTT GCA GAT CCG TAT GAT CAG AGC TTT GAA AGC CGT
 991 GAT CTG CTG ATT GAT GAA TGG AAA AGC CTG ACC TAT GAT GAA GTG
1036 ATT AGC TTT GTT CCG CCT CCA CTG GAT CAA GAA GAA ATG GAA AGC
1081 TAA
```

General Procedure for Protein Expression of p38α Variants

Transformation: Plasmid DNA of pGEX-2T-p38α variants (1 µL each) were transformed into BL21(DE3) *E. coli* chemical competent cells (20–25 µL) by heat shock. The heat shock conditions used were the same as that described for cloning. After feeding with SOC medium (100–125 µL), the cells were incubated in a shaker incubator at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The cells were plated on LB agar plates supplemented with ampicillin (100 µg/mL) at different cell concentrations (5 µL and 20 µL + 40 µL SOC medium, 50 µL) and incubated at 37 °C for 14–16 h. 2–3 colonies were selected and cultured further in liquid LB medium (2–3×10 mL), supplemented again with ampicillin (100 µg/mL) in a shaking incubator at 37 °C, 250 rpm for 15–15.5 h.

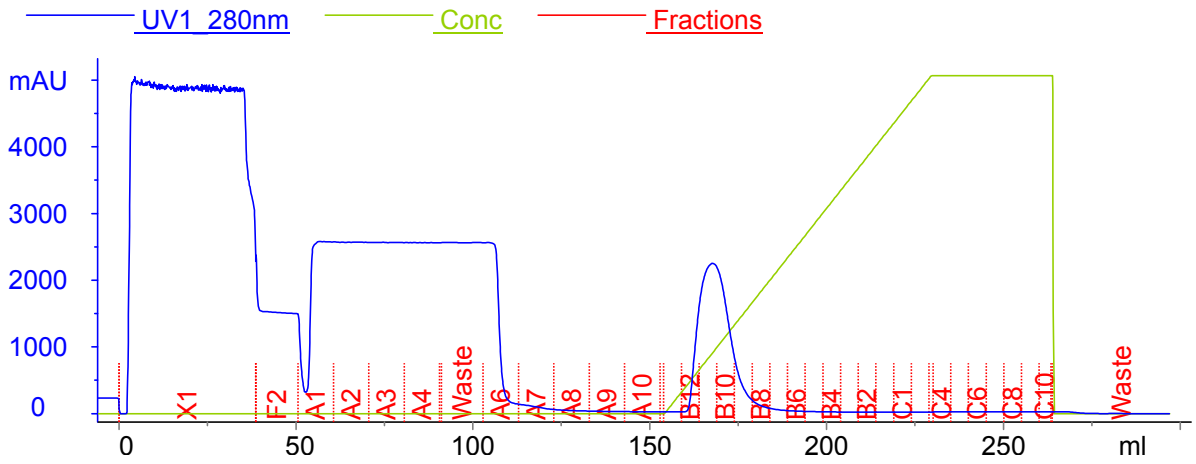
Protein Expression: After glycerol stocks of the cell starter cultures were made, one of the cultures (2×1–4 mL) was used a inoculate a larger culture in 2YT medium (2×800 mL) supplemented with ampicillin (100 µg/mL). These larger cultures were grown to $OD_{600} = 0.4$ at 37 °C, 180 rpm. The temperature was reduced to 20 °C and the cultures grown further to $OD_{600} = 0.65$–0.85 before protein expression was induced with IPTG (800 µL of a 1 M solution in water, final concentration: 1.0 mM). Expression was allowed to continue at 20 °C, 180 rpm for 14.5–16 h. The cells were harvested by pelleting with centrifugation (8000 rpm, JA10 rotor, 3×10 min, 4 °C). The pellets were flash frozen in liquid nitrogen and stored at −80 °C.

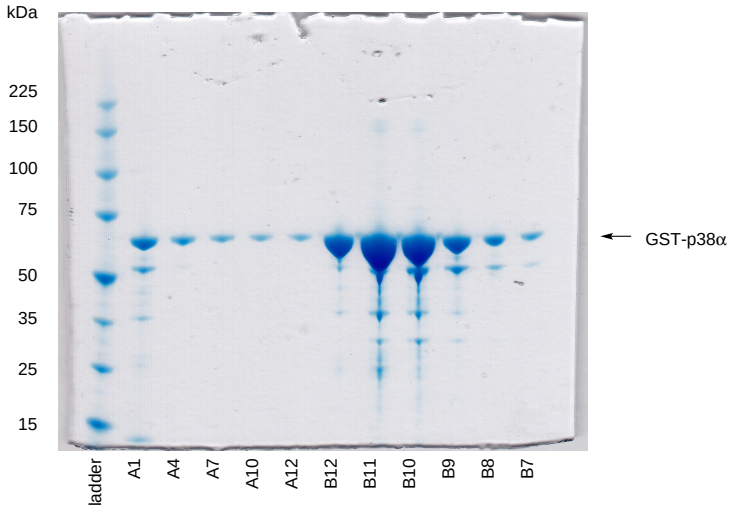
Protein Purification of p38α-Cys172 and p38α-C39A/C119S/C162S/A172C

Cell Lysis: The frozen pellets (10 g) were combined and thawed on ice. Ice cold GST lysis buffer (PBS pH 7.3, 0.5 mM TCEP) (40 mL) was added and the mixture stirred until the pellets were completely resuspended. Lysozyme (50 mg) was added and the mixture stirred on ice for a further 1.5 h, after which the cells were sonicated using a sonicator equipped with a microtip (15×2 s blasts). DNase (1 mg) was then added and the mixture stirred on ice for 20 min before centrifugation (22 000 rpm, JA25.50 rotor, 20 min, 4 °C) and the supernatant taken.

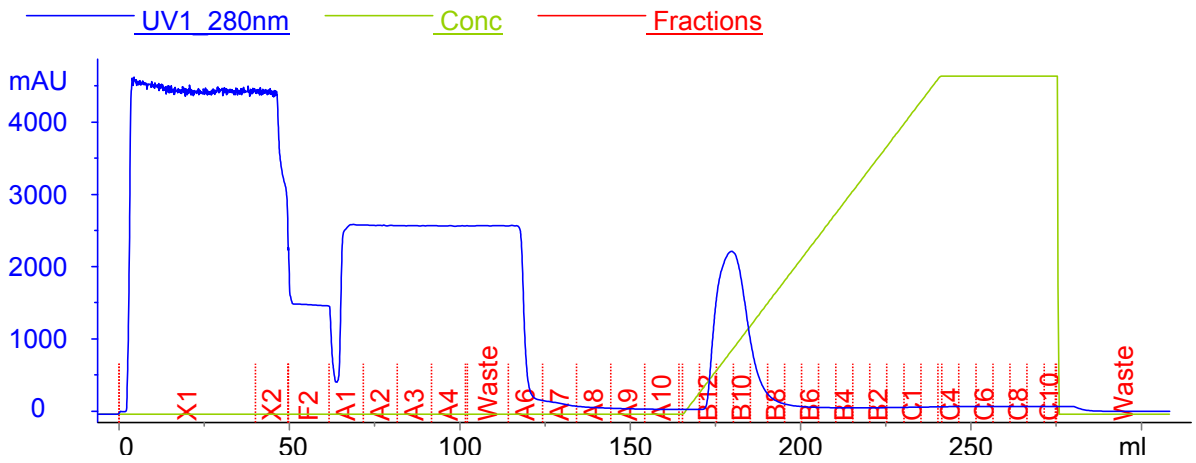
GST Affinity Chromatography: The supernatant was further clarified by sequential filtration through 0.8 µm, 0.45 µm and 0.2 µm filters before loading onto a pre-packed 5 mL GSTrap™ HP GST affinity column using an ÄTKA Purifier® system, chasing through (5 mL, 1 CV) with GST lysis buffer. The column was washed (40 mL, 8 CV) with GST wash buffer (PBS pH 7.3, 1% Triton X-100) and further washed (50 mL, 10 CV) with GST lysis buffer. Elution was with a linear gradient (75 mL, 15 CV, 5 mL fractions) from lysis buffer to GST elution buffer (50 mM Tris pH 8.0, 20 mM L-glutathione). The fractions containing the p38α mutant fusion protein, as determined from the FPLC report file and by SDS-PAGE analysis, were combined.

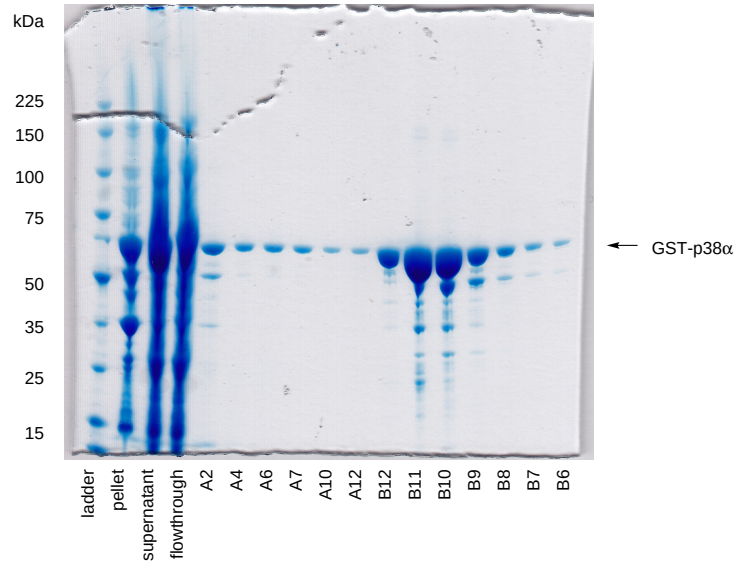
p38α-Cys172 (3):





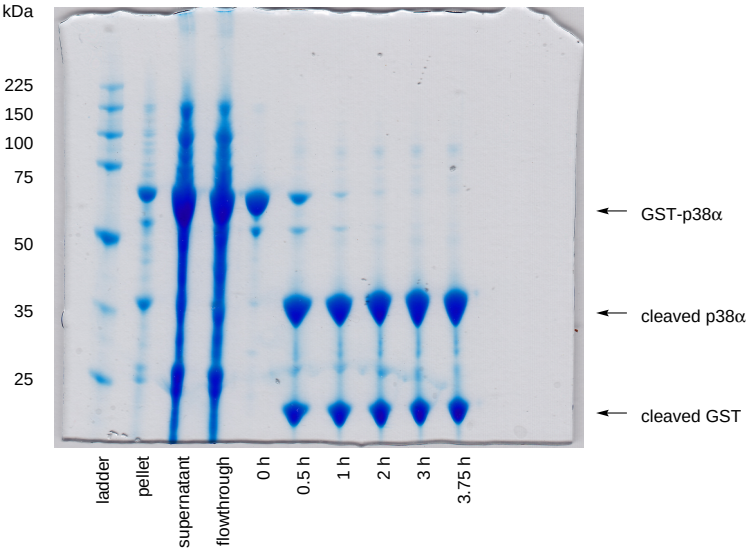
p38α-C39A/C119S/C162S/A172C (4):



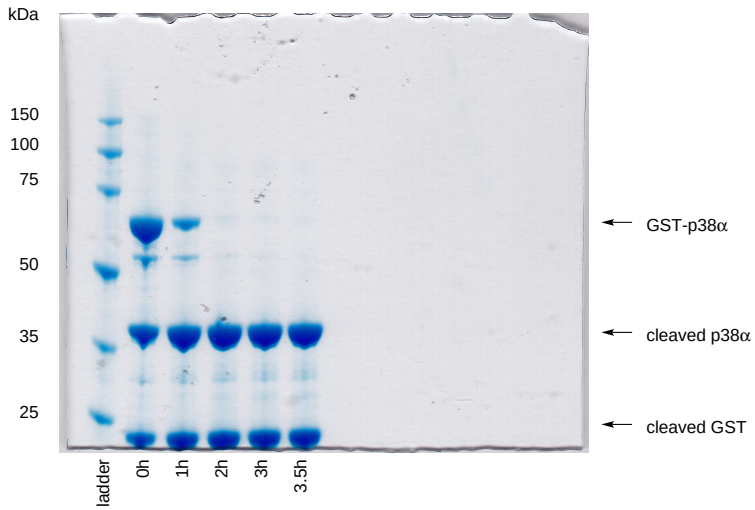


GST Tag Cleavage: The buffer of the combined fractions was exchanged by dialysis (MWCO 10 000) overnight against PBS pH 7.3. The protein solution was warmed to room temperature and thrombin (350 U) added. Incubation was continued at room temperature with gentle rocking for 3.5 h with samples being taken at regular intervals for timecourse monitoring. After this time, PMSF (50–110 µL of a 10 mg/mL solution in isopropanol) was added to a final concentration of 0.3 mM. Some filamentous precipitation was also observed after around 2.5 h.

p38α Variant 3:

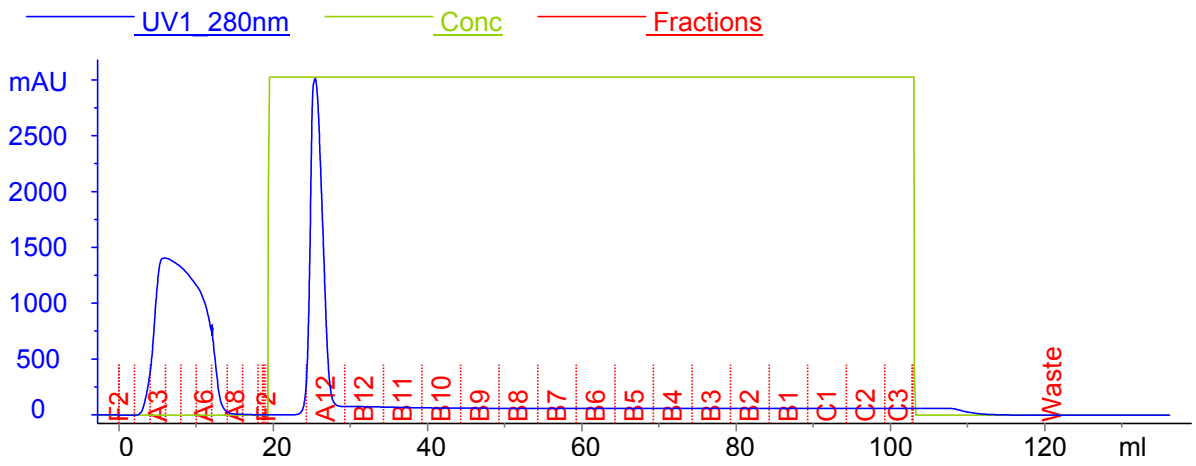


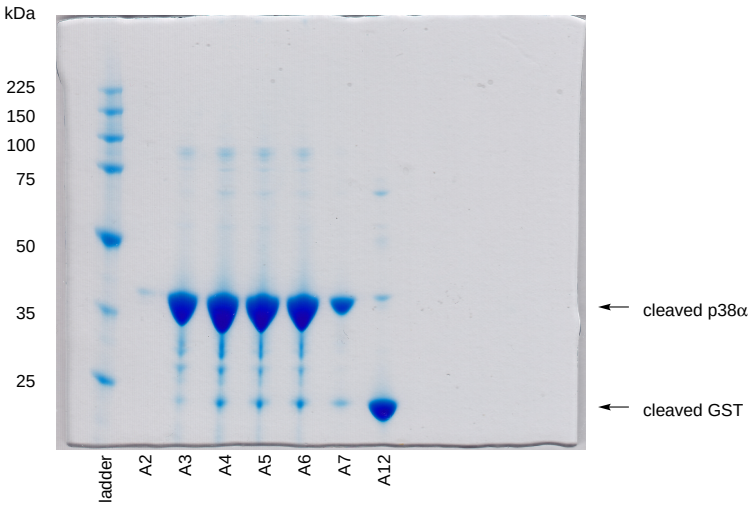
p38α Variant 4:



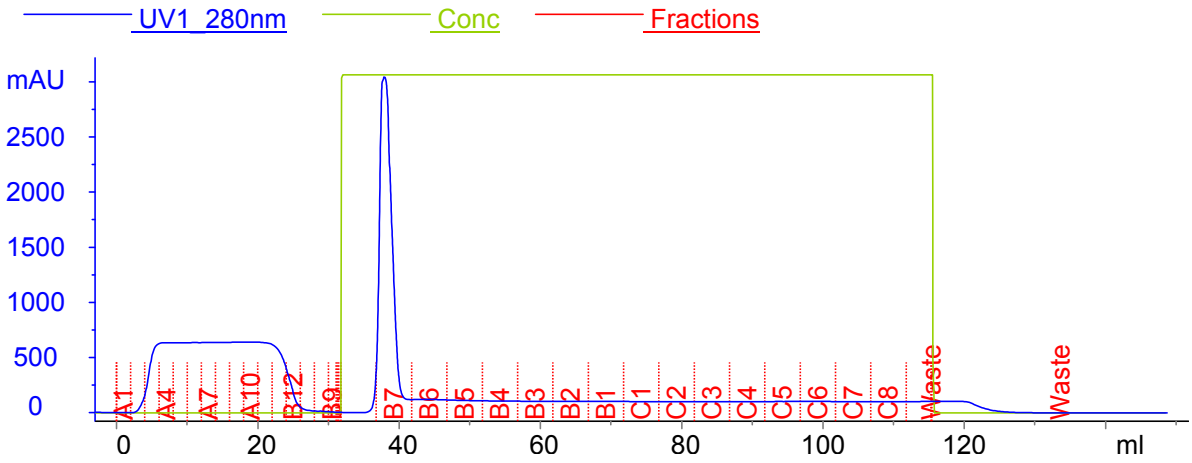
GST Tag Rebinding: The digested protein was filtered before passing through a pre-packed 5 mL GSTrap™ HP GST affinity column. The fractions in the flow-through containing protein, as determined from the UV trace and from SDS-PAGE analysis were combined, while the column was regenerated by elution (75 mL, 15 CV) with GST elution buffer.

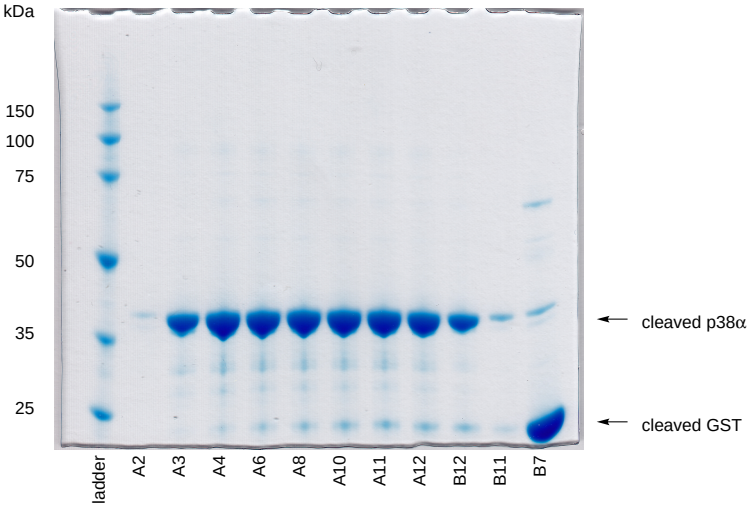
p38α Variant 3:





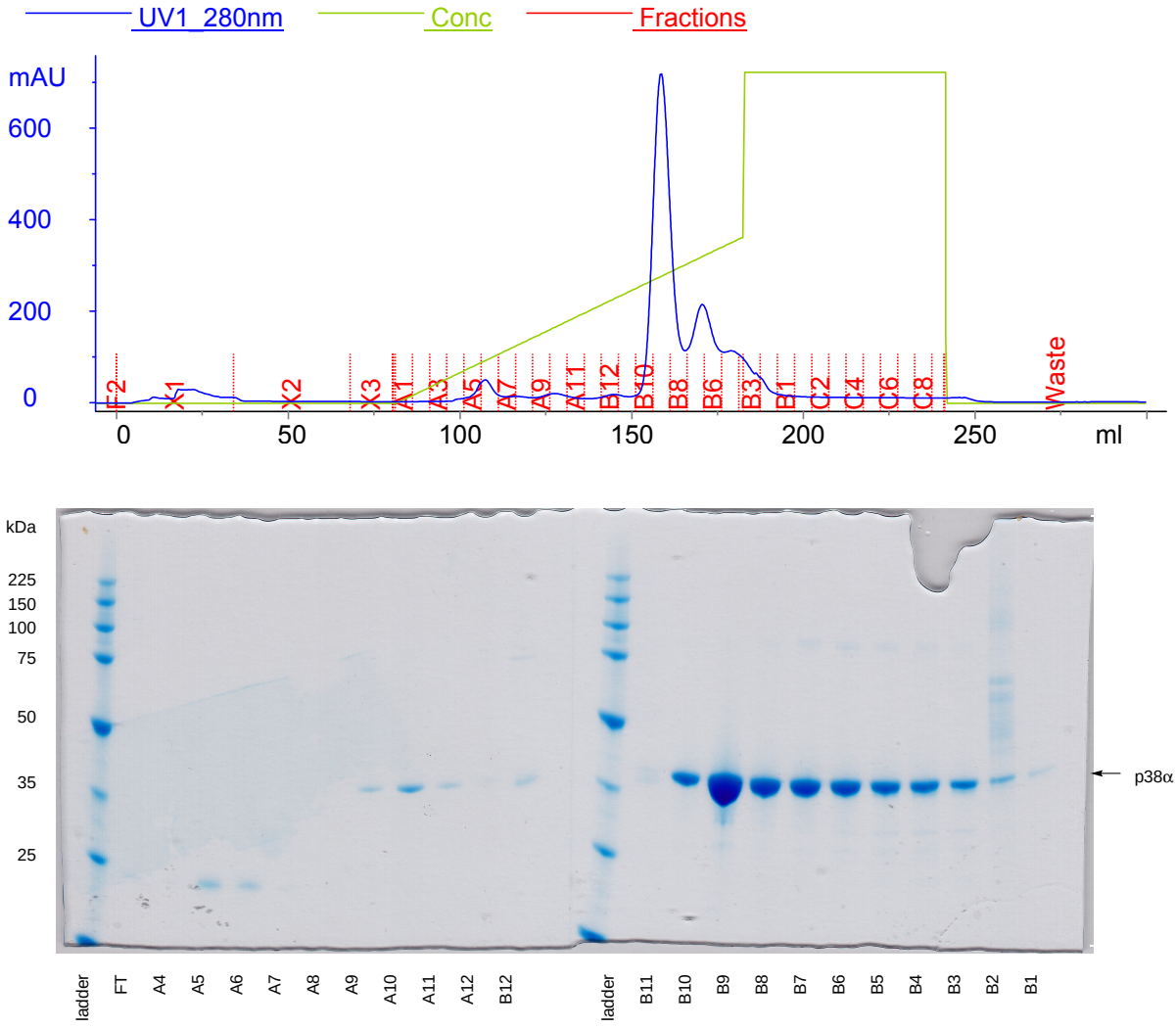
p38α Variant 4:



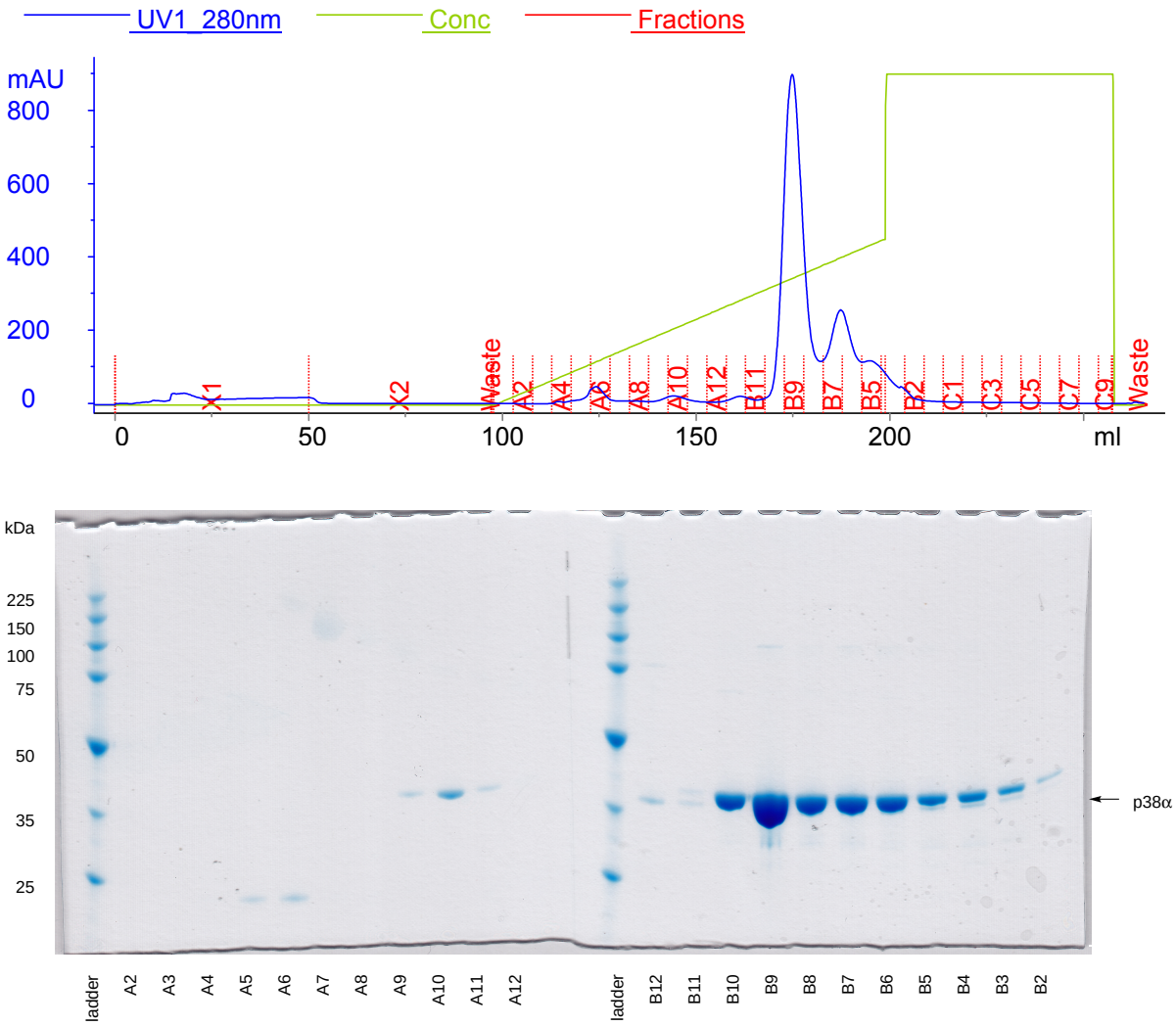


Anion Exchange Chromatography: The protein was diluted 2× with anion exchange start buffer (25 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP) before being re-filtered and loaded onto a 5 mL HiTrap™ Q HP anion exchange column. The column was washed (50 mL, 10 CV) with anion exchange start buffer before the protein was eluted using a linear gradient (100 mL, 20 CV, 5 mL fractions) from start buffer to 50% anion exchange elution buffer (same as start buffer but with 1.0 M NaCl). The protein corresponding to the major peak in the UV trace from the FPLC report file and with the correct mass as determined from SDS-PAGE analysis was taken and these fractions were combined.

p38α Variant 3:

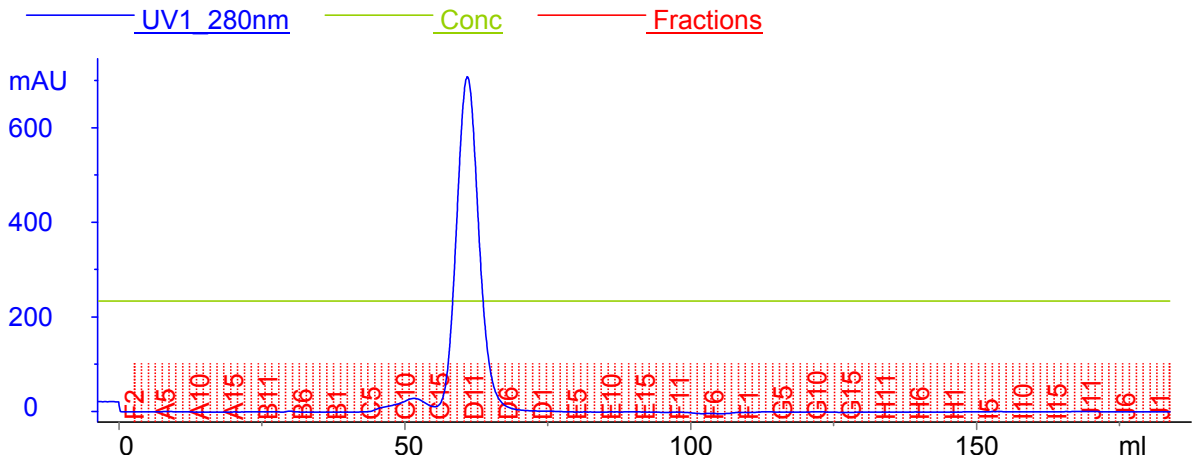


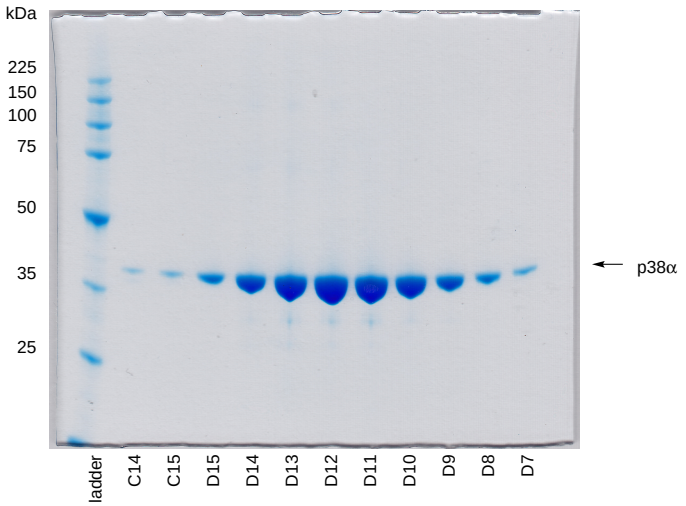
p38α Variant 4:

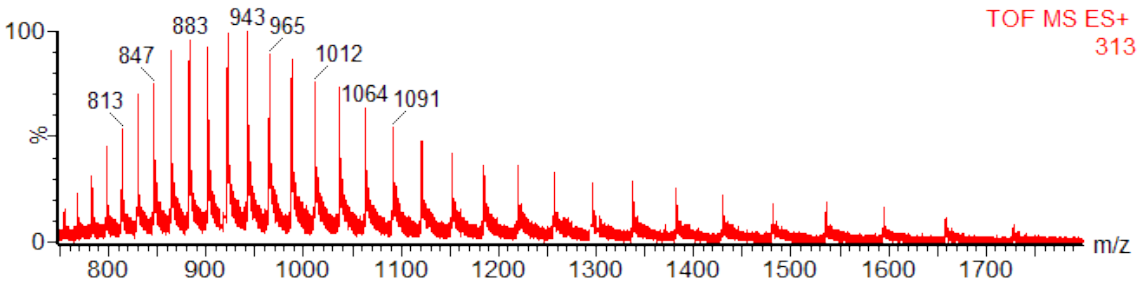


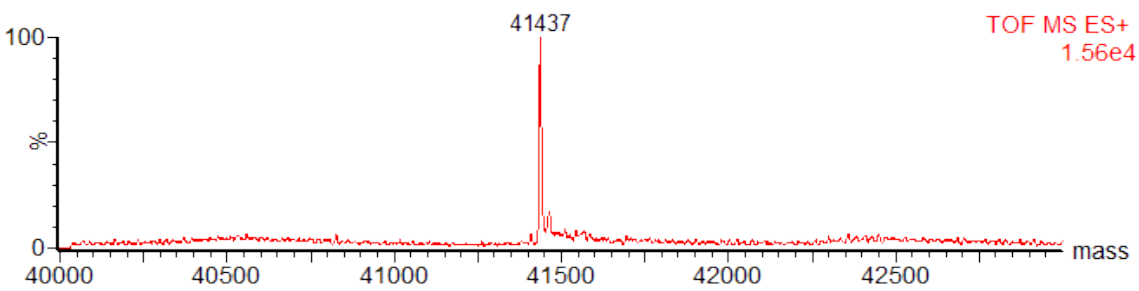
Size Exclusion Chromatography and Storage: The protein was concentrated by Vivaspin (MWCO 10 000) to a final volume ~2.5 mL before filtering. The protein was loaded onto a HiLoad™ 16/60 Superdex 75 gel filtration column and filtered (180 mL, 1.5 CV, 1.2 mL fractions) into p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP). The fractions containing the protein were combined and stored either without further concentration in aliquots (100 µl) or with further concentration by Vivaspin (MWCO 10 000) in aliquots (50 µL). Aliquots were flash frozen in liquid nitrogen before being stored at −80 °C. Yields: p38α variant **3**: 7 mL of a 1.8 mg/mL solution in p38α storage buffer, p38α variant **4**: 3.6 mL of a 3.5 mg/mL solution in p38α storage buffer; p38α variant **3**: (calculated mass: 41 437, observed mass: 41 437), p38α variant **4**: (calculated mass: 41405, observed mass: 41404).

p38α Variant 3:





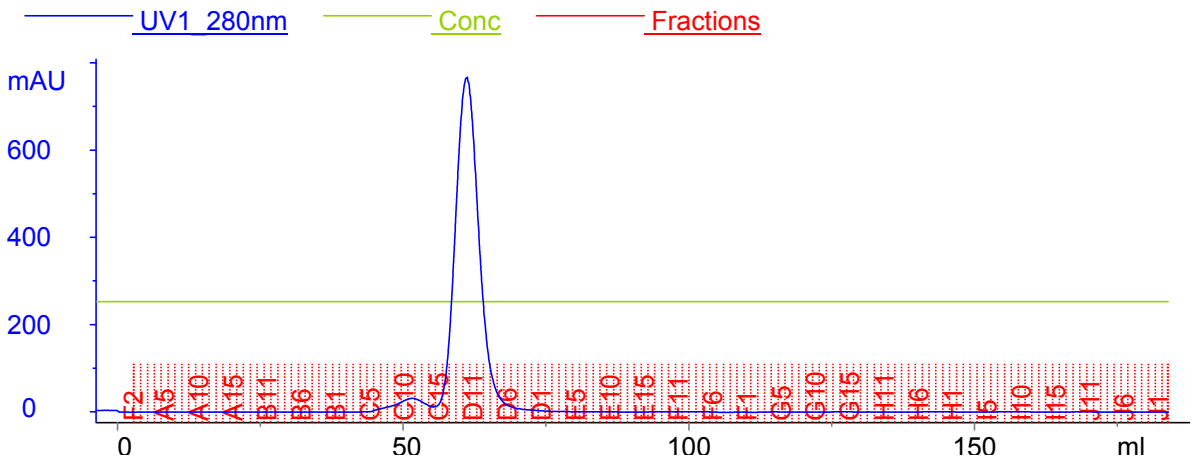


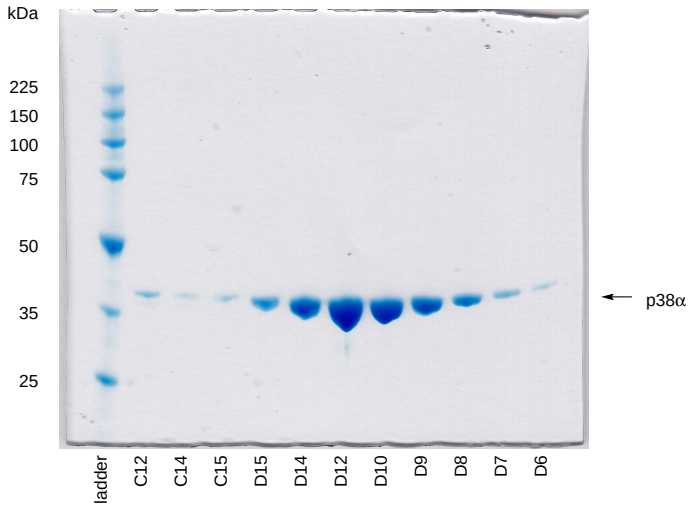


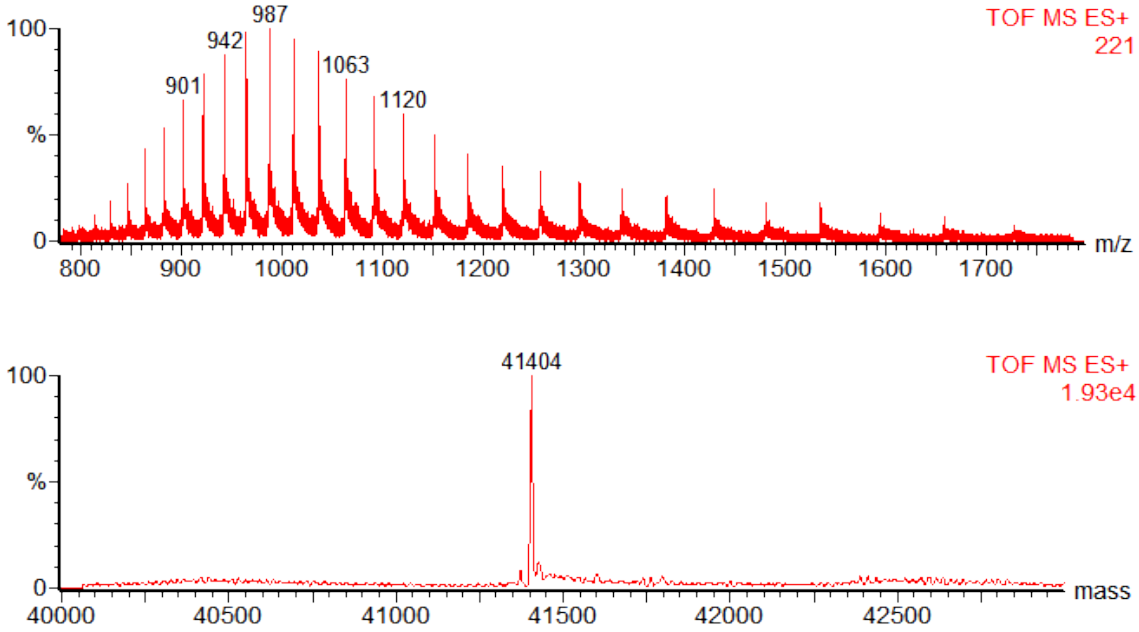
Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                                          GS
  1 MSQERPTFYR QELNKTIWEV PERYQNLSPV GSGAYGSVCA AFDTKTGLRV
 51 AVKKLSRPFQ SIIHAKRTYR ELRLLKHMKH ENVIGLLDVF TPARSLEEFN
101 DVYLVTHLMG ADLNNIVKSQ KLTDDHVQFL IYQILRGLKY IHSADIIHRD
151 LKPSNLAVNE DSELKILDFG LCRHTDDEMT GYVATRWYRA PEIMLNWMHY
201 NQTVDIWSVG CIMAELLTGR TLFPGTDHID QLKLILRLVG TPGAELLKKI
251 SSESARNYIQ SLTQMPKMNF ANVFIGANPL AVDLLEKMLV LDSDKRITAA
301 QALAHAYFAQ YHDPDDEPVA DPYDQSFESR DLLIDEWKSL TYDEVISFVP
351 PPLDQEEMES
```

p38α Variant 4:







Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                                          GS
  1 MSQERPTFYR QELNKTIWEV PERYQNLSPV GSGAYGSVAA AFDTKTGLRV
 51 AVKKLSRPFQ SIIHAKRTYR ELRLLKHMKH ENVIGLLDVF TPARSLEEFN
101 DVYLVTHLMG ADLNNIVKSQ KLTDDHVQFL IYQILRGLKY IHSADIIHRD
151 LKPSNLAVNE DSELKILDFG LCRHTDDEMT GYVATRWYRA PEIMLNWMHY
201 NQTVDIWSVG CIMAELLTGR TLFPGTDHID QLKLILRLVG TPGAELLKKI
251 SSESARNYIQ SLTQMPKMNF ANVFIGANPL AVDLLEKMLV LDSDKRITAA
301 QALAHAYFAQ YHDPDDEPVA DPYDQSFESR DLLIDEWKSL TYDEVISFVP
351 PPLDQEEMES
```

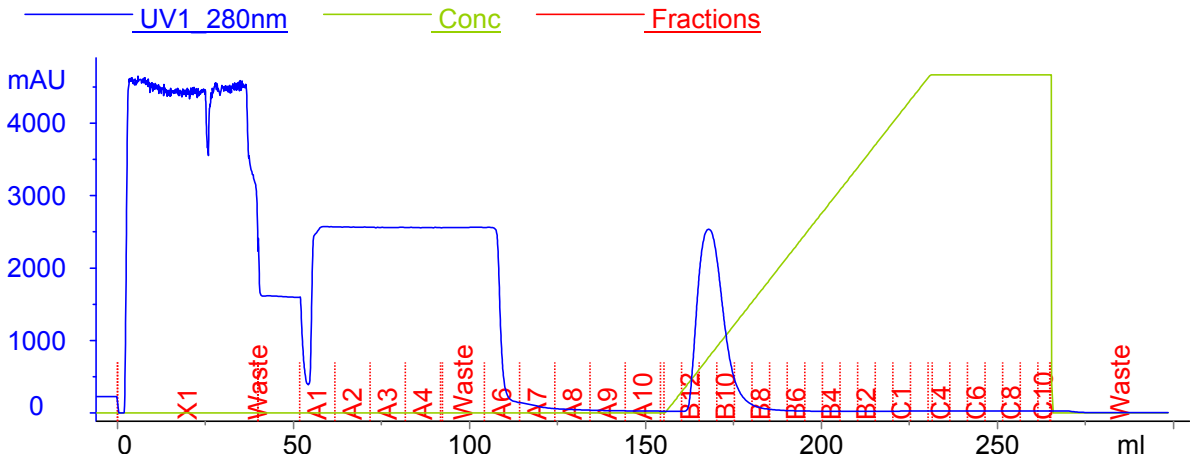
Protein Purification of p38α-Cys180

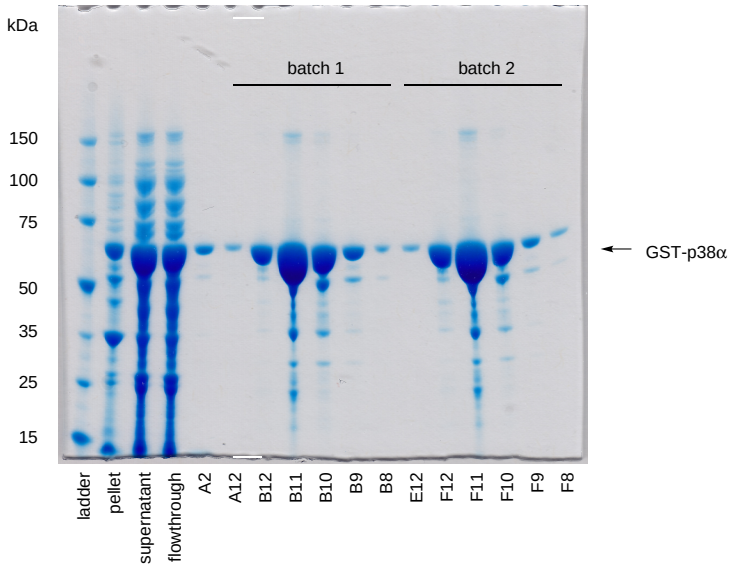
p38α-Cys180 was purified using a similar procedure as used for p38α-Cys172:

Cell Lysis: A pellet (14.0 g) containing p38α-Cys180 was thawed on ice and re-suspended (60 mL) in ice-cold GST lysis buffer (PBS pH 7.3, 0.5 mM TCEP). Lysozyme (70 mg) was then added and the mixture stirred at 4 °C for 2 h before being sonicated with a sonicator equipped with a microtip (15×2 s blasts, 60% amplitude, 1291 J total energy). The lysate was further treated with DNase (1 mg) at 4 °C for 1 h before pelleting by centrifugation (20 000 rpm, JA25.50 rotor, 40 min, 4 °C) and the supernatant taken.

GST Affinity Chromatography: The cleared lysate was further clarified by filtration (0.2 µm filter) and loaded in 2 batches onto a pre-packed GSTrap™ HP GST affinity column (GE Healthcare) on an ÄTKA Purifier™ system. On each run, the lysate was chased through (5 mL, 1 CV) with more lysis buffer before

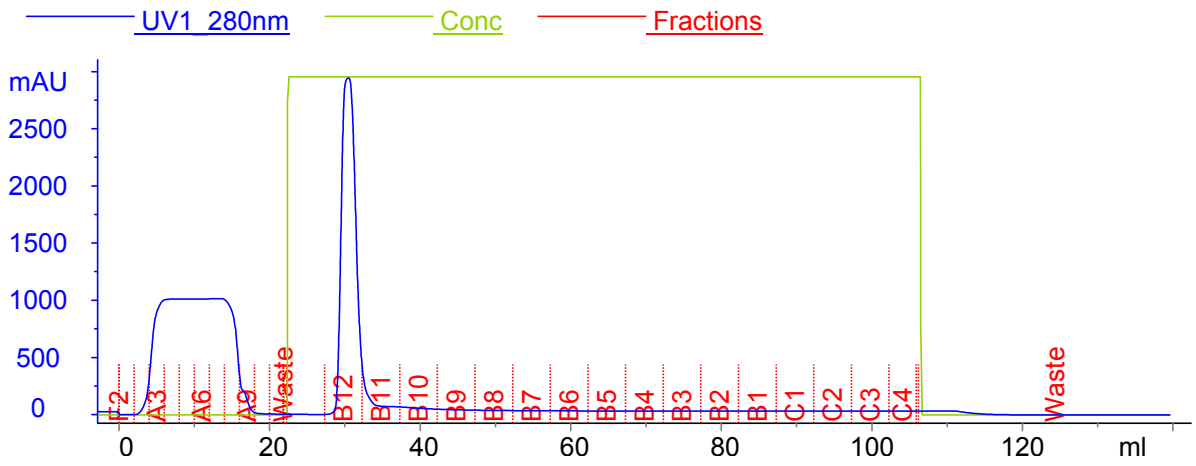
washing, first (40 mL, 8 CV) with GST wash buffer (PBS pH 7.3, 1% Triton X-100), then again (50 mL, 10 CV) with lysis buffer. The bound proteins were eluted (75 mL, 15 CV, 5 mL fractions) by a linear gradient from lysis buffer to GST elution buffer (50 mM Tris pH 8.0, 20 mM L-glutathione). The fractions containing the desired protein from both separations were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and combined.

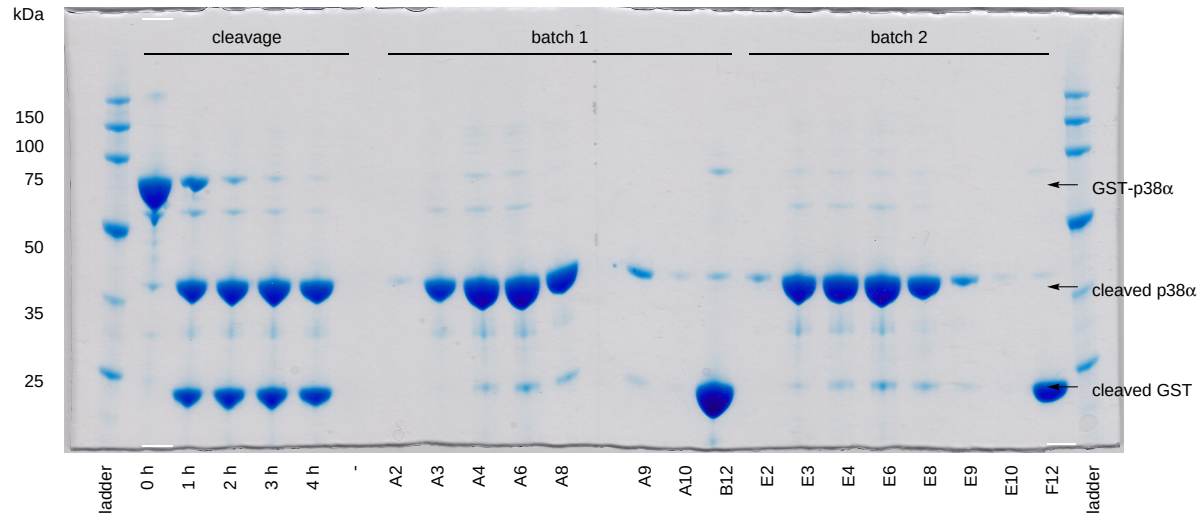




GST Tag Cleavage: The combined fractions were partially concentrated (23 mL final volume) by Vivaspin (10 000 MWCO) before dialysis against PBS at 4 °C for 16 h. Thrombin (400 U of a 1 U/µL solution) in PBS was added and the protein incubated at room temperature with gentle rocking for 4 h before the thrombin was inhibited with PMSF (170 µL of a 10 mg/mL solution) in isopropanol and the protein returned on ice. Small amounts of filamentous precipitate had formed in this reaction.

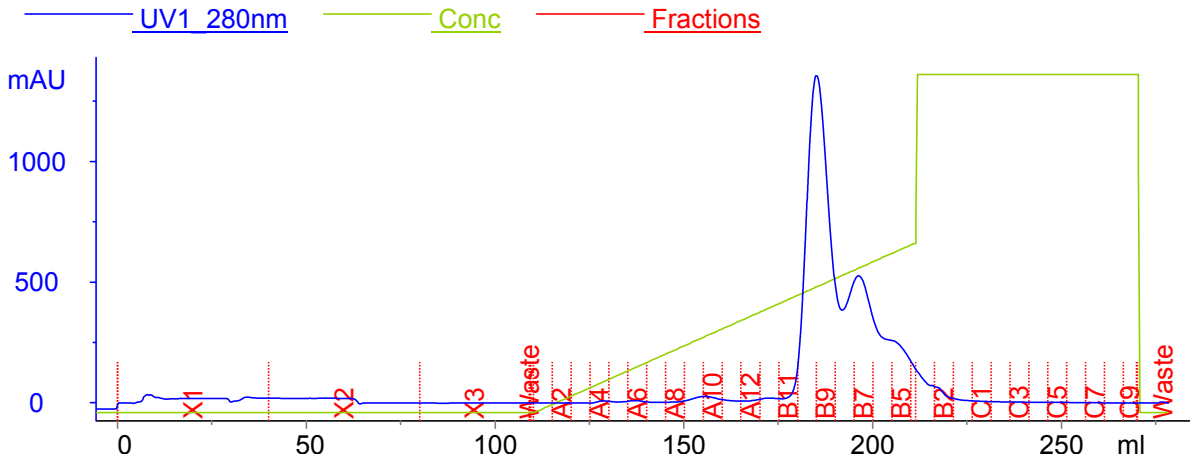
GST Tag Rebinding: After filtering, the protein was re-loaded (2 mL fractions collected) onto the GSTrap™ HP GST affinity column, again in 2 batches and in each run, the column was washed (10 mL, 2 CV) with lysis buffer before elution (75 mL, 15 CV) with 100% elution buffer. The fractions from the flow-through containing the target protein of both batches were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and combined.

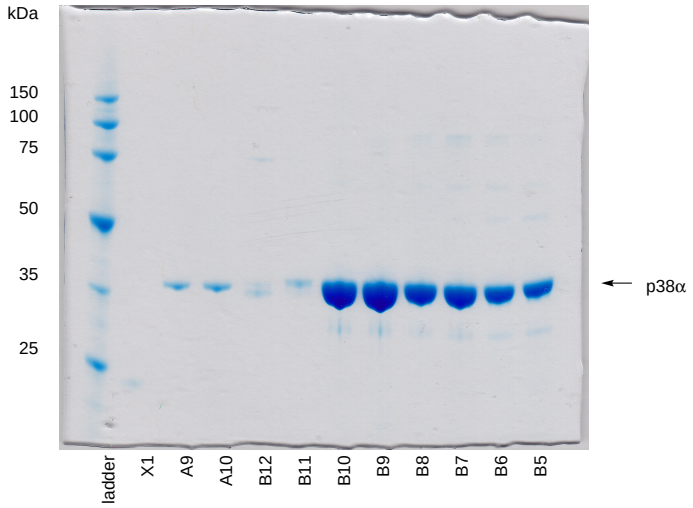




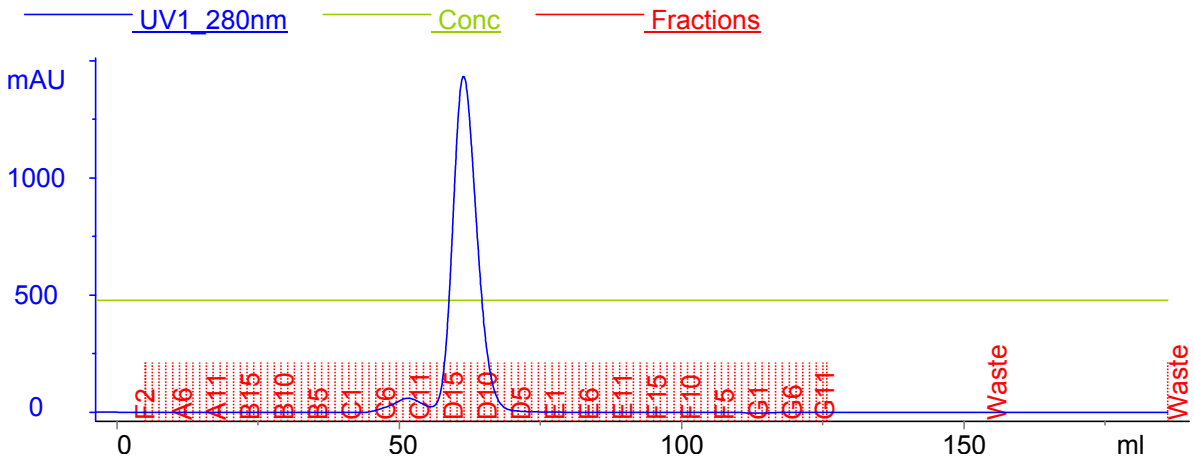
Anion Exchange Chromatography: The combined fractions were diluted 2× with anion exchange start buffer (25 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP), re-filtered, and loaded as a single batch onto a 5 mL HiTrap™ Q HP anion exchange column (GE Healthcare). The column was further washed (50 mL, 10 CV) with start buffer before elution (100 mL, 20 CV, 5 mL fractions) with a linear gradient from start buffer to 50% anion exchange elution buffer (same as start buffer but with additional 1.0 M NaCl). SDS-PAGE analysis of the fractions determined that the desired protein corresponded to 2 distinct peaks in the

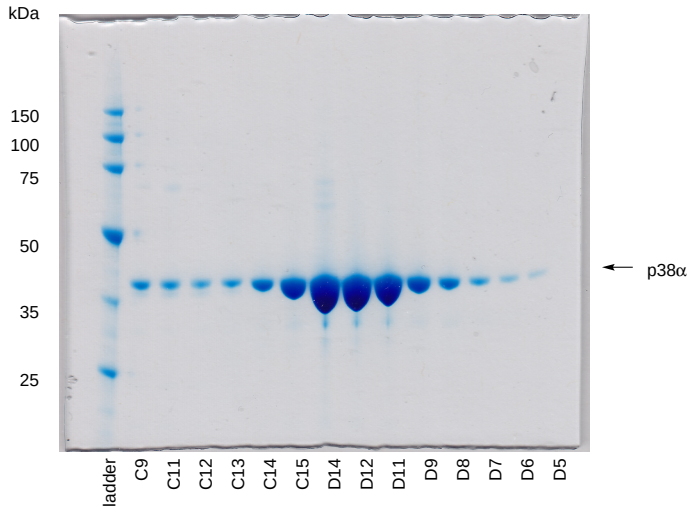
UV trace of the FPLC report file, with the first peak being the major one. Protein corresponding to this major peak was taken as the desired protein and further treated.

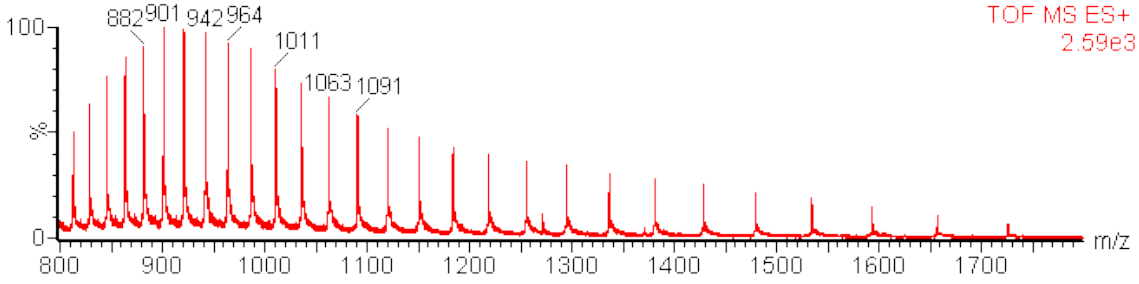


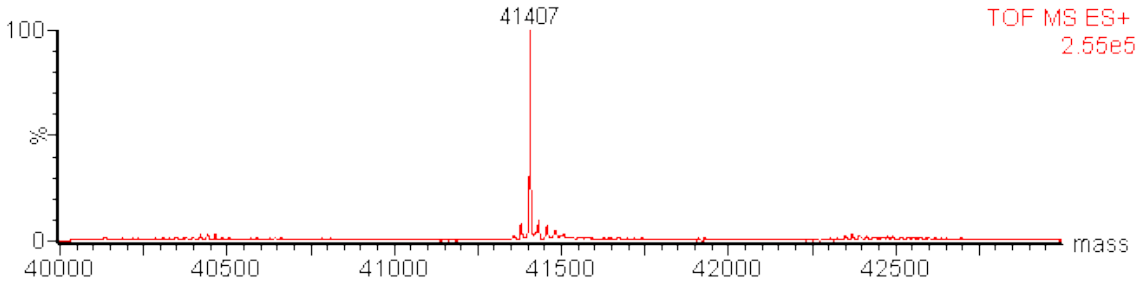


Size Exclusion Chromatography and Storage: The protein was concentrated (∼3 mL) before being loaded onto a HiLoad™ 16/60 Superdex 75 gel filtration column and filtered (180 mL, 1.5 CV) into p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP). The fractions containing the target protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, pooled, and concentrated by Vivaspin. The protein was then divided into aliquots, flash frozen in liquid nitrogen and stored at −80 °C; Yield: 2.8 mL of a 8.2 mg/mL solution in p38α storage buffer; (calculated mass: 41 407, observed mass: 41 404–41 409).


UV1_280nm
Conc
Fractions
mAU
1000
500
0
F2
A6
A11
B15
B10
B5
C1
C6
C11
D15
D10
D5
E1
E6
E11
F15
F10
F5
G1
G6
G11
Waste
Waste
0
50
100
150
ml


kDa
150
100
75
50
35
25
p38α
ladder
C9
C11
C12
C13
C14
C15
D14
D12
D11
D9
D8
D7
D6
D5


TOF MS ES+
2.59e3
882
901
942
964
1011
1063
1091
100
%
0
800
900
1000
1100
1200
1300
1400
1500
1600
1700
m/z

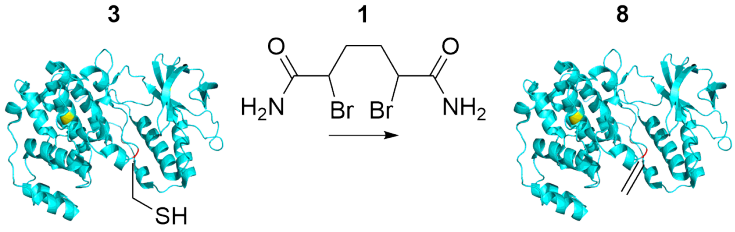

TOF MS ES+
2.55e5
41407
100
%
0
40000
40500
41000
41500
42000
42500
mass

Protein sequence obtained by translation of the corresponding gene sequence:

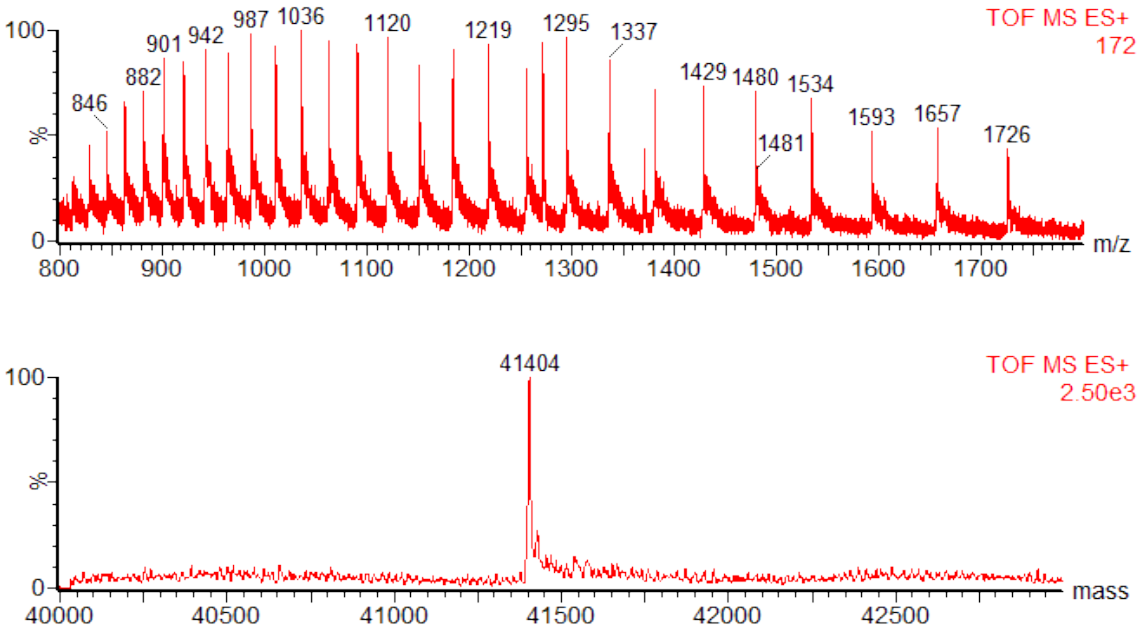
```
tag                                                          GS
  1 MSQERPTFYR QELNKTIWEV PERYQNLSPV GSGAYGSVCA AFDTKTGLRV
 51 AVKKLSRPFQ SIIHAKRTYR ELRLLKHMKH ENVIGLLDVF TPARSLEEFN
101 DVYLVTHLMG ADLNNIVKSQ KLTDDHVQFL IYQILRGLKY IHSADIIHRD
151 LKPSNLAVNE DSELKILDFG LARHTDDEMC GYVATRWYRA PEIMLNWMHY
201 NQTVDIWSVG CIMAELLTGR TLFPGTDHID QLKLILRLVG TPGAELLKKI
251 SSESARNYIQ SLTQMPKMNF ANVFIGANPL AVDLLEKMLV LDSDKRITAA
301 QALAHAYFAQ YHDPDDEPVA DPYDQSFESR DLLIDEWKSL TYDEVISFVP
351 PPLDQEEMES
```

2.11.4 Chemical Modification: Optimised Procedures

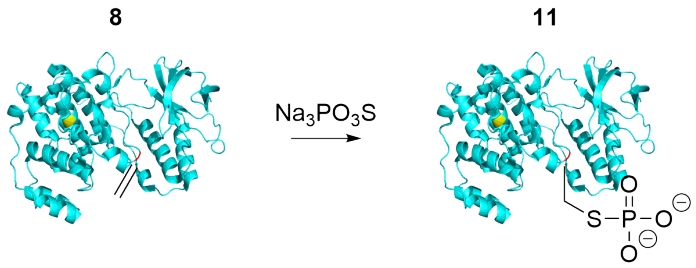
Synthesis of p38α-Dha172



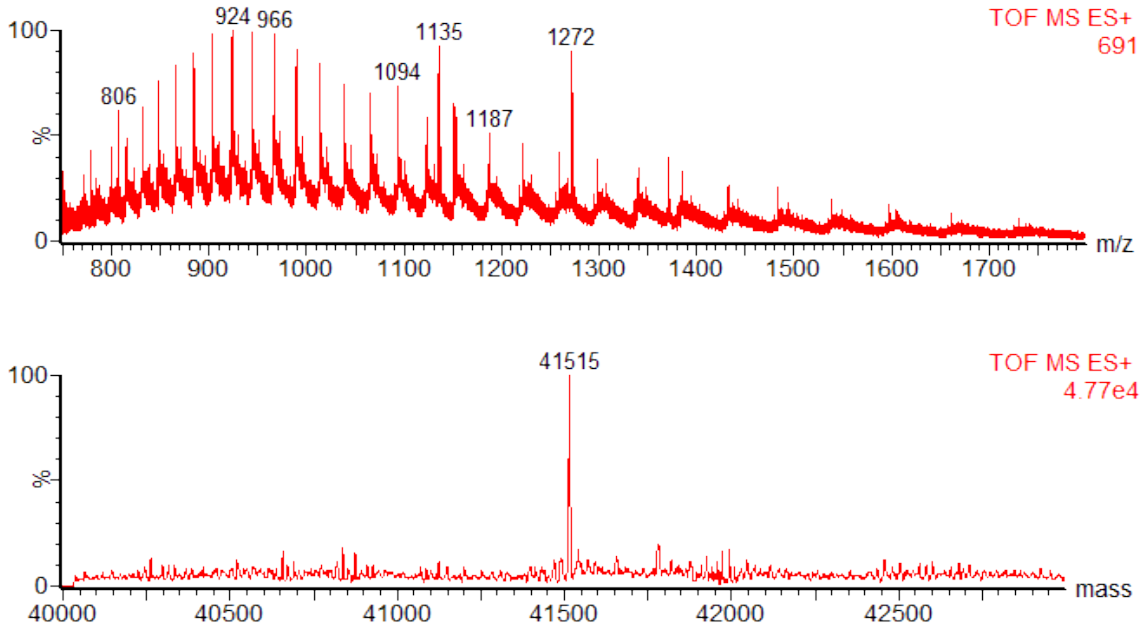
p38α-Cys172 (500 µL of a 1.8 mg/mL solution, 22 nmol, 1.0 eq.) in p38α storage buffer (50 mM HEPES, pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged using a G-25 MiniTrap™ desalting column (GE Healthcare) to p38α reaction buffer (the same as storage buffer but without TCEP). A suspension of **1** (252 µL of a 3.9 mg/mL suspension, 3.3 µmol, 150 eq.) in p38α reaction buffer was added and the mixture was shaken at 37 °C, 550 rpm for 4 h. LC-MS after this time showed >95% conversion to the Dha product. The reaction mixture was desalted using a G-25 MidiTrap column to give p38α-Dha172, which was either kept on ice for immediate use, or flash frozen in liquid nitrogen and stored at −80 °C for short-term use: Yield: 755 µL of a 0.68 mg/mL solution in p38α reaction buffer, 57%, (calculated mass: 41 403, observed mass: 41 404).



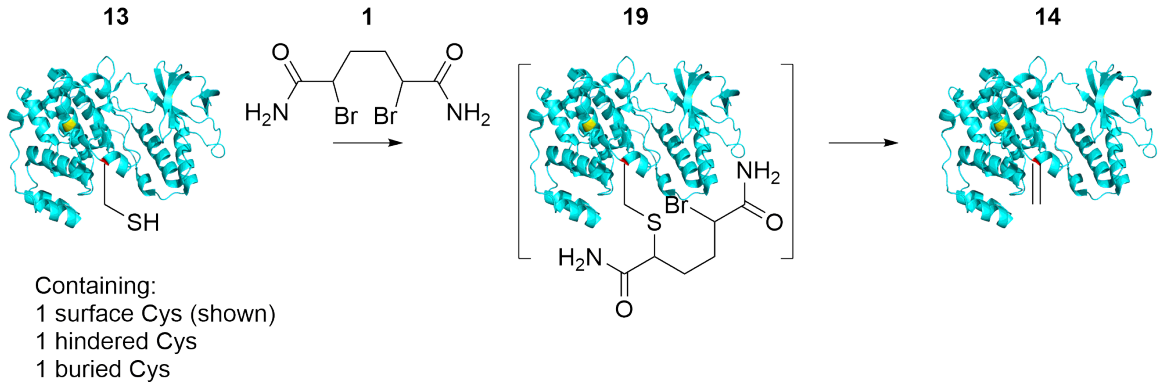
Synthesis of p38α-pCys172



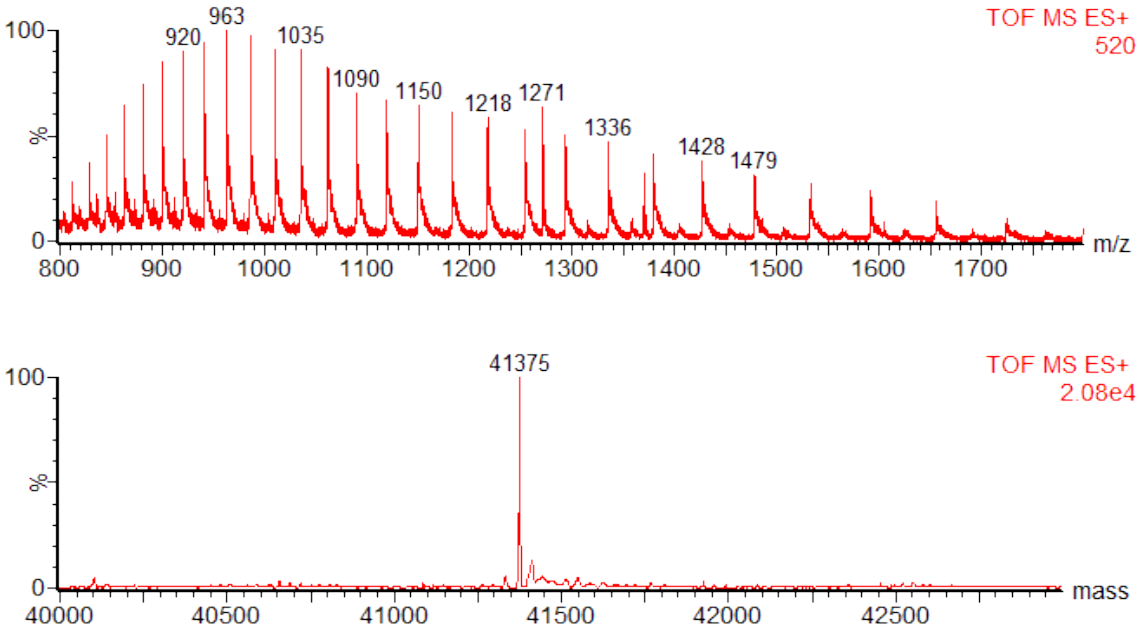
p38α-Dha172 (470 µL of a 0.48 mg/mL solution, 5.4 nmol, 1.0 eq.) in p38α reaction buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 5% glycerol) was thawed on ice. To the thawed protein was added batchwise (5 min intervals) sodium thiophosphate (~pH 8.0, 5×7.11 µL of a 690 mg/mL suspension, 5×27 µmol, 5×5000 eq.) in H_2O/HCl and the mixture shaken at 37 °C, 550 rpm for 4 h. LC-MS after this time showed >95% conversion to the Dha product. The reaction mixture was desalted back to reaction buffer using a G-25 MiniTrap column, followed by repeated concentration/dilution by Vivaspin (MWCO 10 000) to give p38α-pCys172, which was kept on ice for immediate use: (calculated mass: 41 517, observed mass: 41 515).



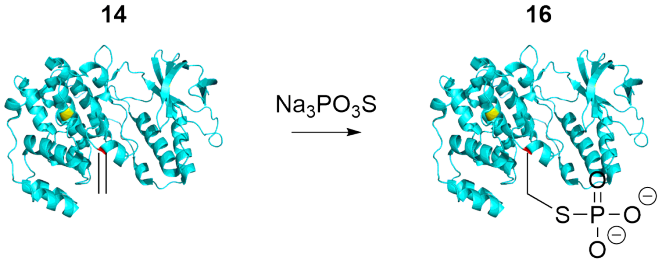
Synthesis of p38α-Dha180



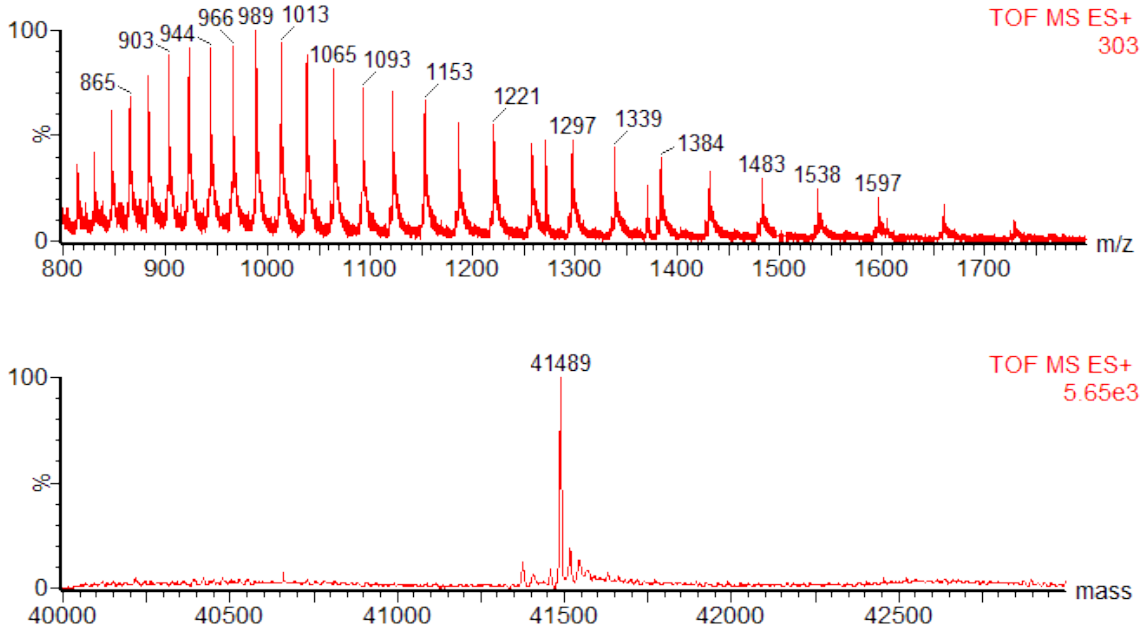
p38α-Cys180 **13** was thawed on ice and the buffer exchanged using a G-25 MiniTrap™ desalting column (GE Healthcare) to p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol). To the resulting protein (580 µL of a 2.51 mg/mL solution, 35 nmol, 1.0 eq.) was added a suspension of dibromide **1** (410 µL of a 3.9 mg/mL suspension, 5.3 µmol, 150 eq.) in the same buffer. The mixture was then shaken at 37 °C, 550 rpm for 2 h. LC-MS showed >95% consumption of the unmodified protein p38α-Cys180 with some residual bromide adduct intermediate **19**. The reaction mixture was desalted again using a G-25 MidiTrap™ desalting column (GE Healthcare) and the protein incubated on ice for 18 h. LC-MS showed >95% conversion to the Dha product p38α-Dha172. The reaction mixture was divided into aliquots and kept on ice for immediate further use: Yield: 1250 µL of a 0.87 mg/mL solution in p38α reaction buffer, 75%, (calculated mass: 41 373, observed mass: 41 375).



Synthesis of p38α-pCys180

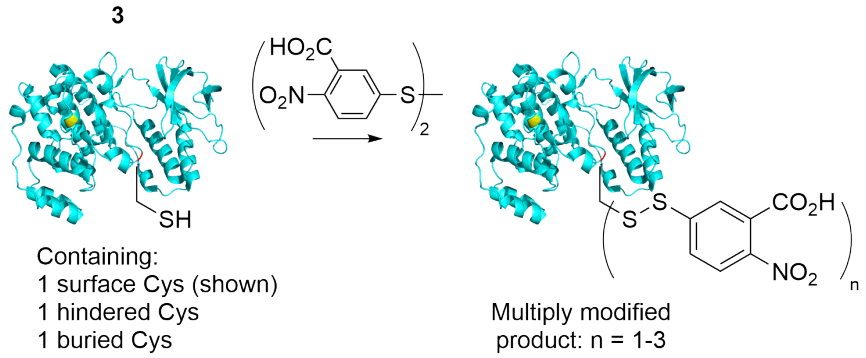


Sodium thiophosphate was added portion-wise (6×13.7 µL of a 690 mg/mL solution at pH 8, 6×53 µmol, 6×5000 eq.) at 5 min intervals to p38α-Dha180 (500 µL of a 0.87 mg/mL solution, 10.5 nmol, 1.0 eq.) in p38α reaction buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 5% glycerol). The sodium thiophosphate solution was made by dissolving sodium thiophosphate (68.2 mg) in water (27.6 µL) and the pH adjusted with 5 M HCl (29.6 µL). The reaction mixture was then heated with shaking at 37 °C, 550 rpm for 8 h. LC-MS after this time showed >90% conversion to the thiophosphate adduct. The reaction mixture was desalted back to reaction buffer using a G-25 MiniTrap column, followed by repeated concentration/dilution by Vivaspin (MWCO 10 000) to give the thiophosphate product p38α-pCys180: (calculated mass: 41 487, observed masses: 41 489).

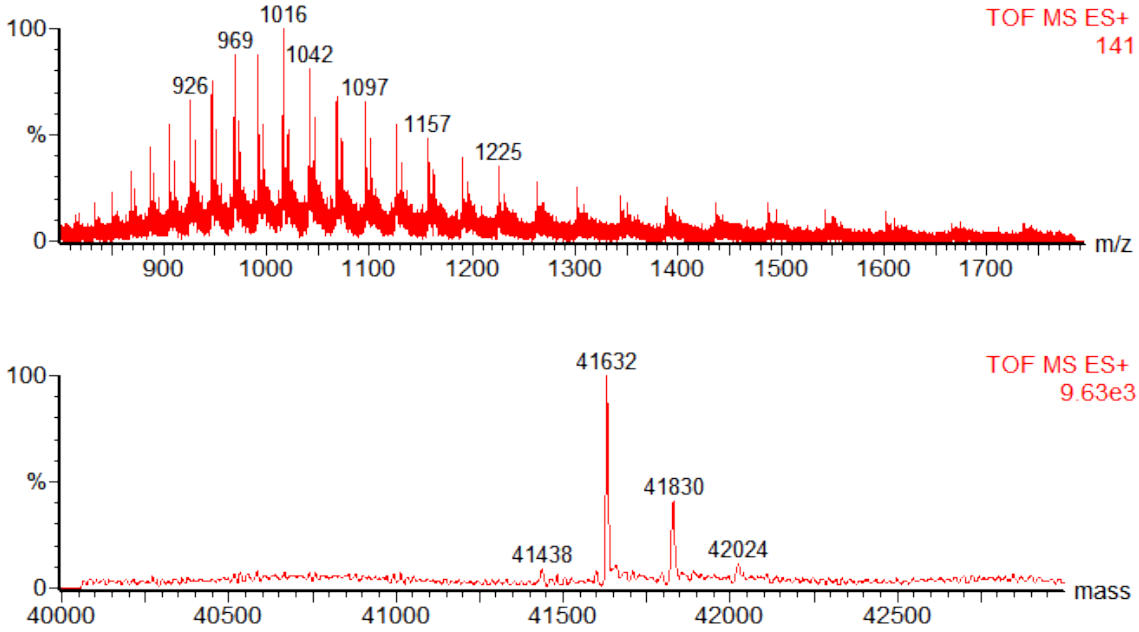


2.11.5 Chemical Modification: Control Experiments

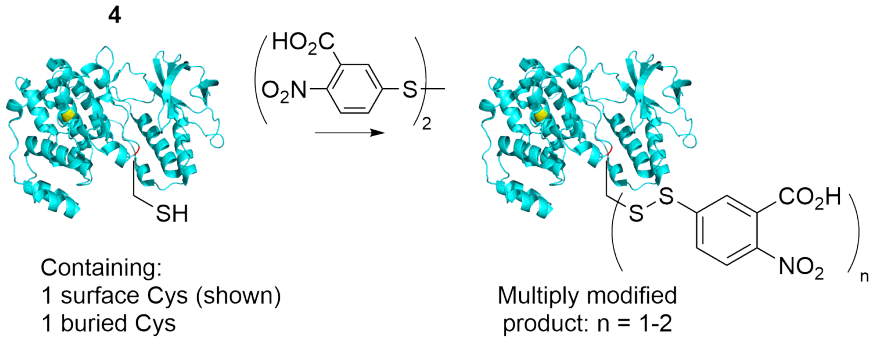
Treatment of p38α Variant 3 with Ellman's Reagent



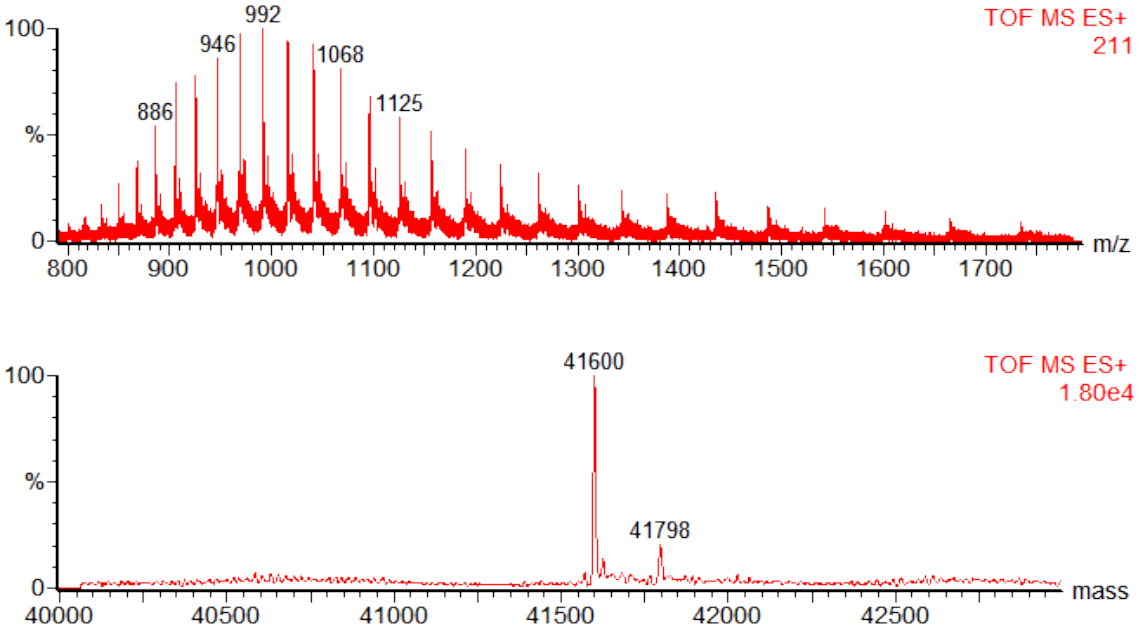
A freshly thawed solution of **3** (80 µL of a 1.8 mg/mL solution, 3.5 nmol, 1.0 eq.) in p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP) was diluted (60 µL) with HEPES buffer (50 mM HEPES pH7.5). A solution of Ellman's reagent (20.7 µL of a 6.65 mg/mL solution, 350 nmol, 100 eq.) in HEPES buffer was then added and the mixture mixed end-over-end at 4 °C for 30 min. After this time, a sample (20 µL) of the reaction mixture was diluted further with HEPES buffer (40 µL) and analysed by LC-MS. LC-MS showed a mixture of Ellman's adducts corresponding to 1–3 modifications being made to the protein, with the singly modified protein being the predominant product; (calculated masses: 41634, 41831 and 42028, observed masses: 41632, 41830 and 42024).



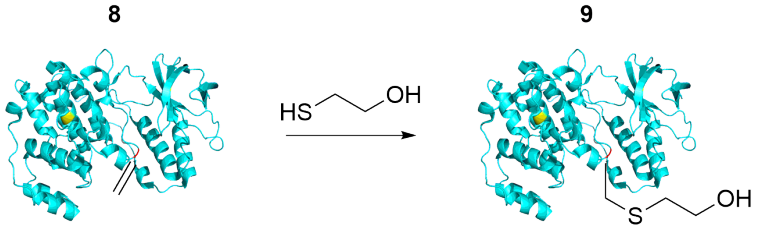
Treatment of p38α Variant 4 with Ellman's Reagent



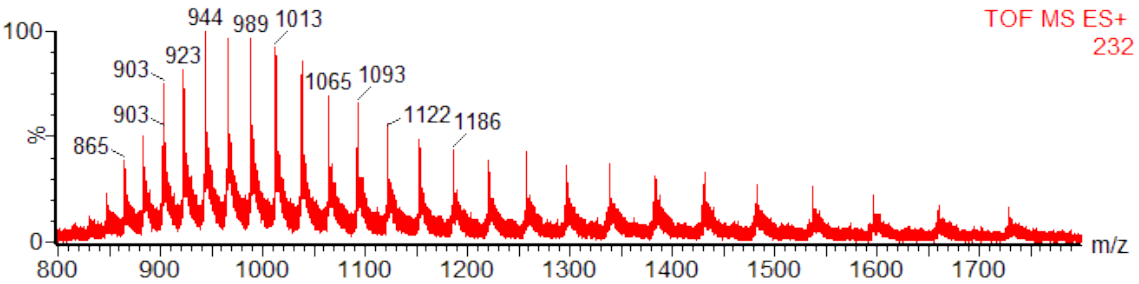
A freshly thawed solution of **4** (40 µL of a 3.5 mg/mL solution, 3.4 nmol, 1.0 eq.) in p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP) was diluted (60 µL) with HEPES buffer (50 mM HEPES pH7.5). A solution of Ellman's reagent (20.2 µL of a 6.65 mg/mL solution, 340 nmol, 100 eq.) in HEPES buffer was then added and the mixture mixed end-over-end at 4 °C for 30 min. After this time, a sample (20 µL) of the reaction mixture was diluted further with HEPES buffer (40 µL) and analysed by LC-MS. LC-MS showed a mixture of Ellman's adducts corresponding to 1–2 modifications being made to the protein, with the singly modified protein being the predominant product; (calculated masses: 41602 and 41799, observed masses: 41600 and 41798).

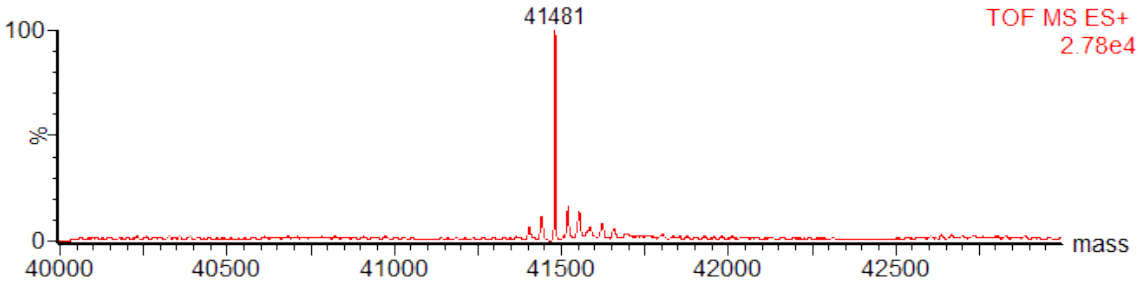


Treatment of p38α-Dha172 (8) with β-mercaptoethanol

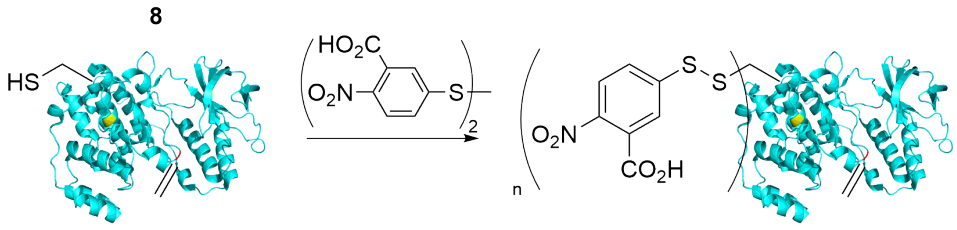


To an aliquot of p38α-Dha172 (**8**) (136 µL of a 0.69 mg/mL solution, 2.3 nmol, 1.0 eq.) was added a solution of β-mercaptoethanol (3.46 µL of a 5% solution, 31.7 µmol, 14 000 eq.) in reaction buffer. The mixture was shaken at room temperature, 500 rpm for 2 h and the reaction followed by LC-MS. LC-MS after 1 h showed ~90% conversion while LC-MS after 2 h showed >95% conversion to the mercaptoethanol adduct **9** (calculated mass: 41481, observed mass: 41481).





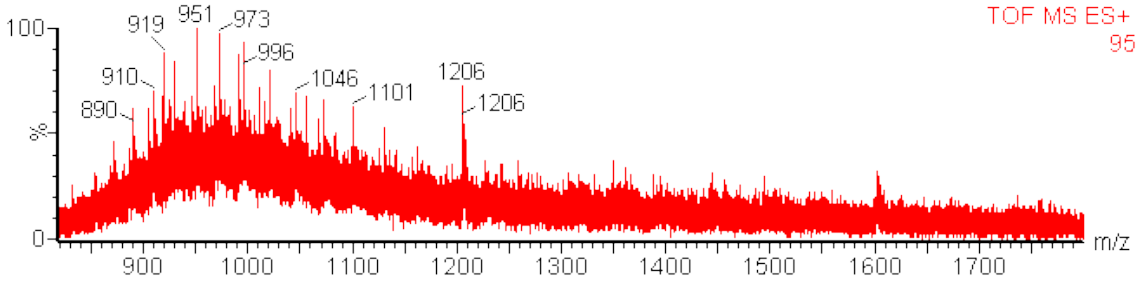
Treatment of p38α-Dha172 (8) with Ellman's Reagent

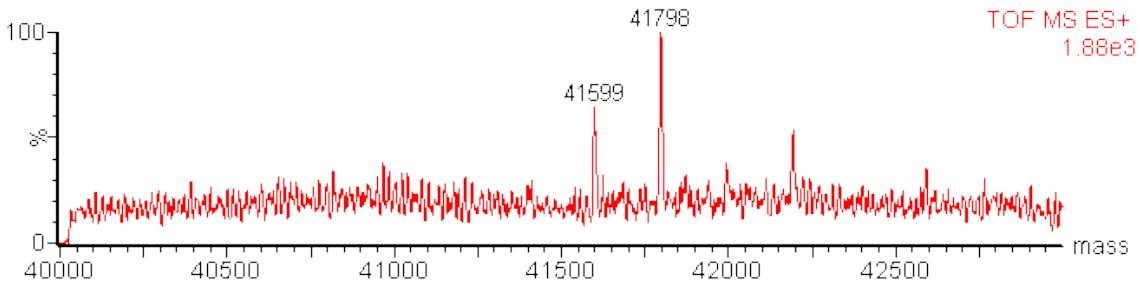


Containing:
1 hindered Cys (shown)
1 buried Cys

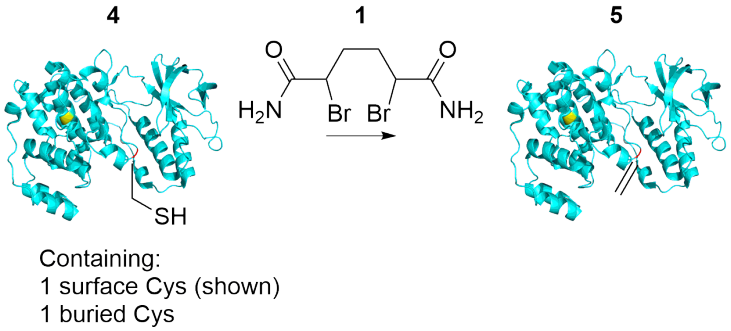
Multiply modified
product: n = 1-2

To an aliquot of p38α-Dha172 (**8**) (136 µL of a 0.69 mg/mL solution, 2.3 nmol, 1.0 eq.) was added a solution of Ellman's reagent (13.8 µL of a 6.5 mg/mL solution, 0.23 µmol, 100 eq.) in reaction buffer. The mixture was shaken at room temperature, 500 rpm for 2 h. LC-MS after this time showed complete conversion to the single and double Ellman's adducts in a 3:5 ratio (calculated masses: 41600 and 41797, observed masses: 41599 and 41798).

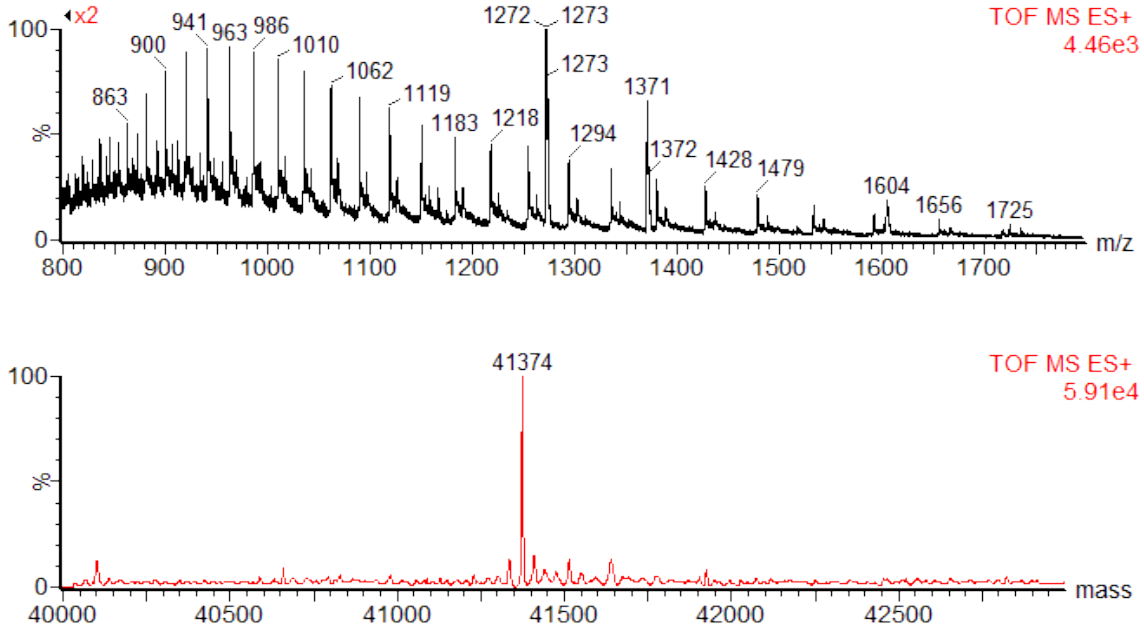




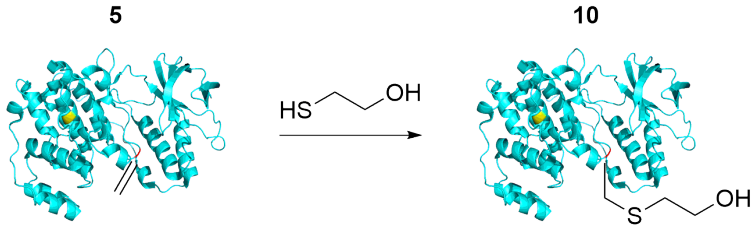
Treatment of p38α Variant 4 with Dibromide 1



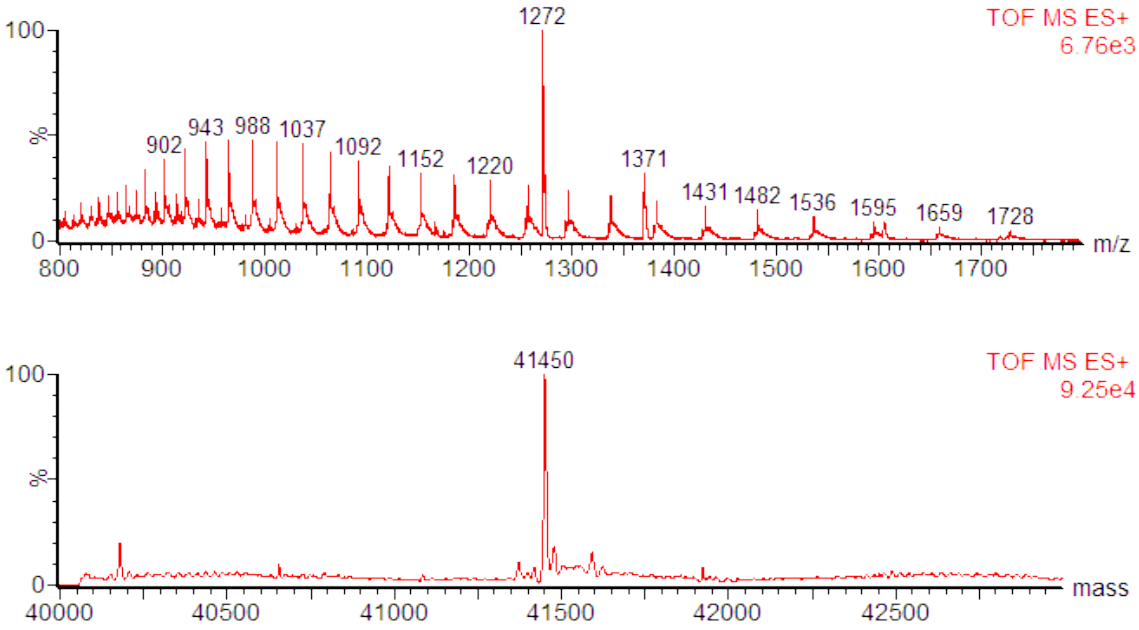
The buffer of a freshly thawed aliquot of p38α variant **4** (150 µL of a 3.5 mg/mL solution, 13 nmol, 1.0 eq.) was exchanged from p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl,5% glycerol, 0.5 mM TCEP) to p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) using a SpinTrap™ G-25 desalting column. To resulting protein aliquot (150 µL) was further diluted with more reaction buffer (270 µL) before a suspension of **1** (147 µL of a 3.9 mg/mL suspension, 1.9 µmol, 150 eq.) in reaction buffer was added and the reaction shaken at 37 °C, 500 rpm for 4 h. After this time, the reaction mixture was filtered and the product concentrated (2.1 mg/mL) by Vivaspin. The product was kept on ice and used directly in subsequent reactions. LC-MS showed >95% conversion to the singly modified Dha product **5** (calculated mass: 41371, observed mass: 41374).



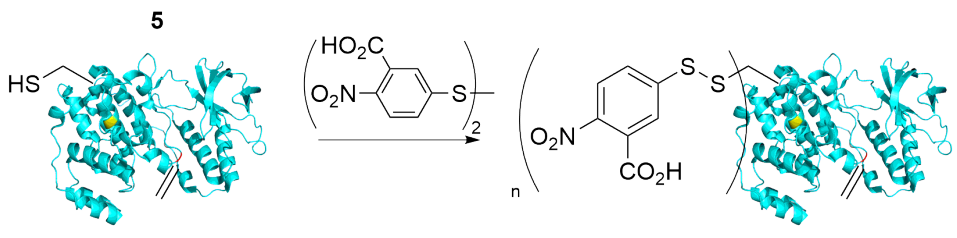
Treatment of Dha Product 5 with β-mercaptoethanol



This procedure is similar to the one used on p38α-Dha172. To an aliquot of Dha product **5** (58 µL of a 2.1 mg/mL solution, 2.9 nmol, 1.0 eq.) was added a solution of β-mercaptoethanol (10 µL of a 2.5% solution, 45.8 µmol, 15 700 eq.) in reaction buffer. The mixture was shaken at room temperature, 500 rpm and the reaction followed by LC-MS over 2 h. LC-MS after 1 h showed >95% conversion to the mercaptoethanol adduct **10** (calculated mass: 41449, observed mass: 41450).



Treatment of Dha Product 5 with Ellman's Reagent

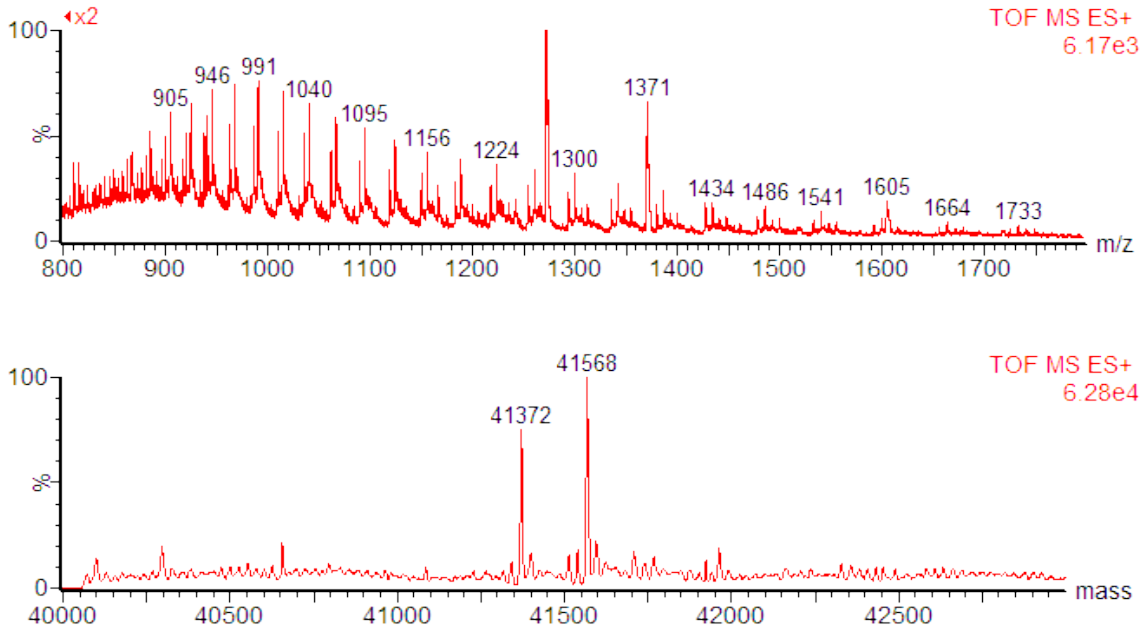


Containing:
1 buried Cys (shown)

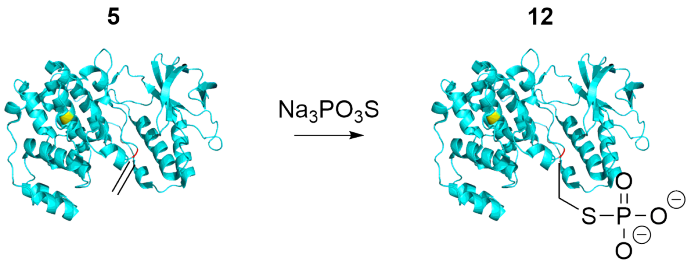
Multiply modified
product: n = 0-1

This procedure is also similar to the one used on p38α-Dha172. To an aliquot of Dha product **5** (58 µL of a 2.1 mg/mL solution, 2.9 nmol, 1.0 eq.) was added a solution of Ellman's reagent (17.8 µL of a 6.5 mg/mL solution, 0.29 µmol, 100 eq.) in reaction buffer. The mixture was shaken at room temperature, 500 rpm and the reaction followed by LC-MS over 2 h. LC-MS after 1 h showed ~50% conversion to the single Ellman's adduct with further conversion (~60%) after 2 h (calculated mass: 41568, observed mass: 41568).

After 2 h:

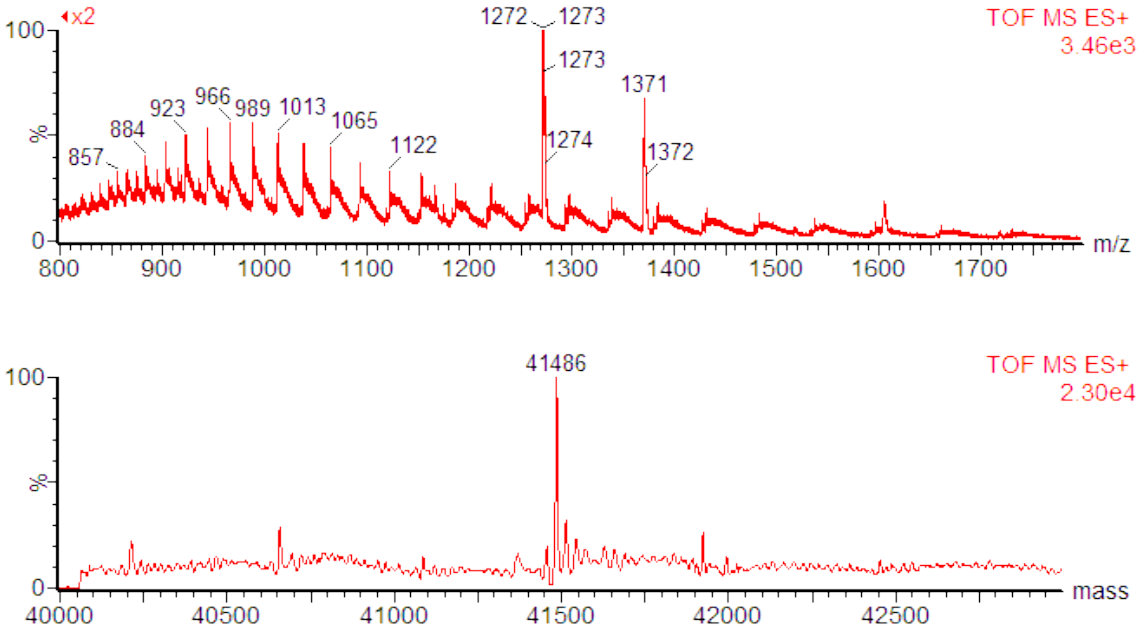


Treatment of Dha Product 5 with Sodium Thiophosphate

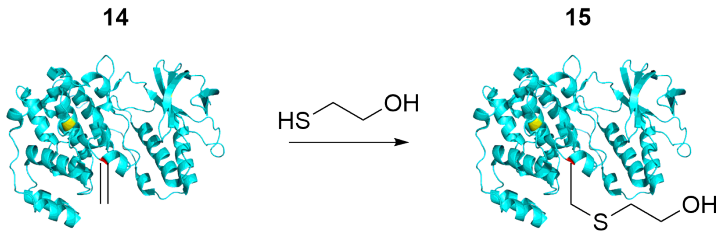


To an aliquot of Dha product **5** (58 µL of a 2.1 mg/mL solution, 2.9 nmol, 1.0 eq.) was added a solution of sodium thiophosphate (19 µL of a 700 mg/mL solution, 73 µmol, 25 000 eq.) in water. The thiophosphate solution was made prior addition by dissolving sodium thiophosphate (64.2 mg) in water (26.0 µL) and balancing the pH with HCl (27.9 µL of a 5 M solution in water) to pH 7.5. The mixture was shaken at

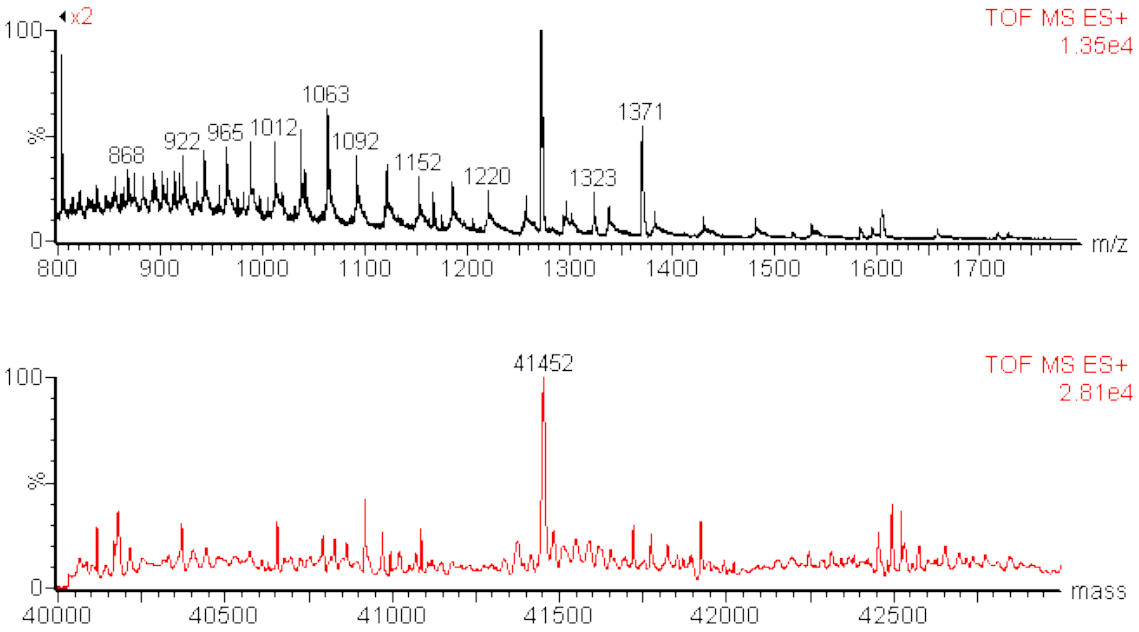
37 °C, 500 rpm and the reaction followed by LC-MS over 3.5 h. LC-MS after 3 h showed >95% consumption of **5**, with white precipitate being observed after 3.5 h. LC-MS after 3 h also showed >90% conversion to the thiophosphate adduct **12** (calculated mass: 41485, observed mass: 41486).



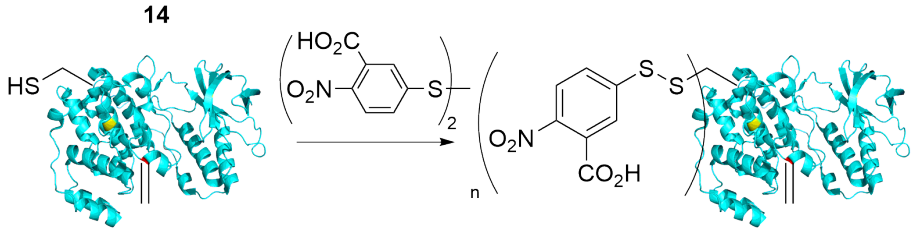
Treatment of p38α-Dha180 (14) with β-mercaptoethanol



To an aliquot of p38α-Dha180 (**14**) (163 µL of a 0.9 mg/mL solution, 3.5 nmol, 1.0 eq.) in p38α reaction buffer was added a solution of β-mercaptoethanol (7.4 µL of a 5% solution, 5.3 µmol, 1500 eq.) in reaction buffer. The reaction was incubated with shaking at room temperature, 500 rpm and the progress of reaction monitored by LC-MS over 3 h. LC-MS after 1 h showed >90% conversion of **14** to the mercaptoethanol adduct **15** with no further observable changes over a prolonged reaction time; (calculated mass: 41452, observed mass: 41451).



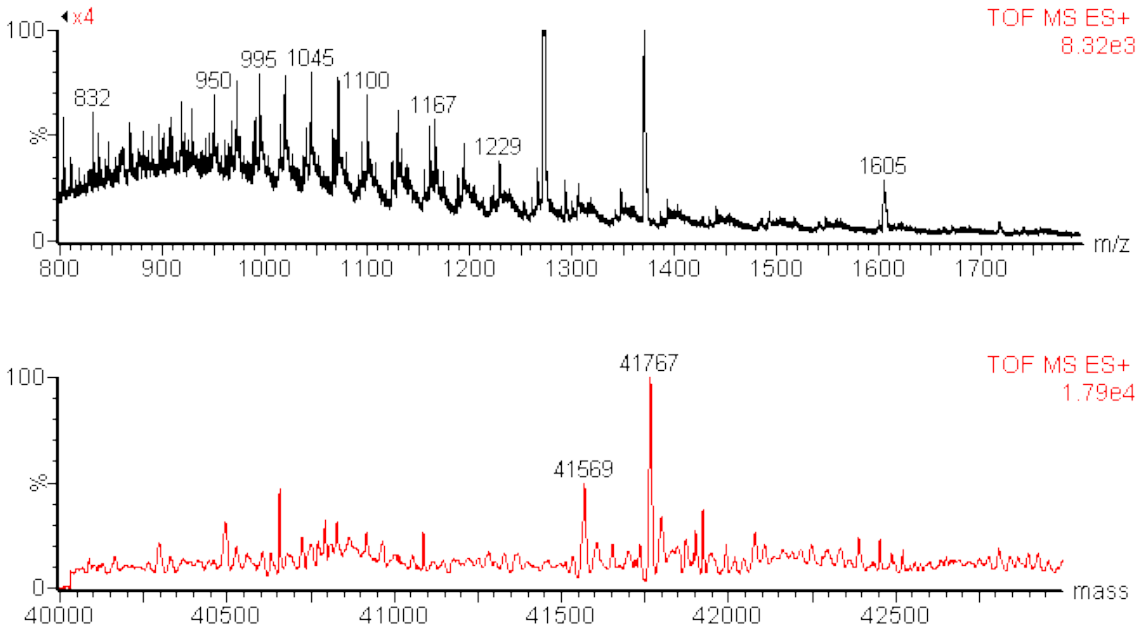
Treatment of p38α-Dha180 (14) with Ellman's Reagent



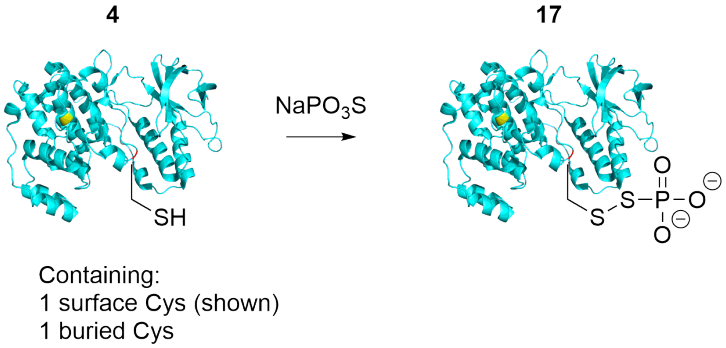
Containing:
1 hindered Cys (shown)
1 buried Cys

Multiply modified
product: n = 1-2

To an aliquot of p38α-Dha180 **14** (163 µL of a 0.9 mg/mL solution, 3.5 nmol, 1.0 eq.) in p38α reaction buffer was added a solution of Ellman's reagent (21.6 µL of a 6.5 mg/mL solution, 0.35 µmol, 100 eq.) in reaction buffer. The reaction was incubated with shaking at room temperature, 500 rpm and the progress of reaction monitored by LC-MS over 3 h. LC-MS after 1 h showed the complete conversion of **14** to either the mono- or di-functionalised Ellman's adduct in a 1:2 ratio with no further change on prolonged incubation; (calculated mass: 41570 and 41767, observed masses: 41569 and 41767).

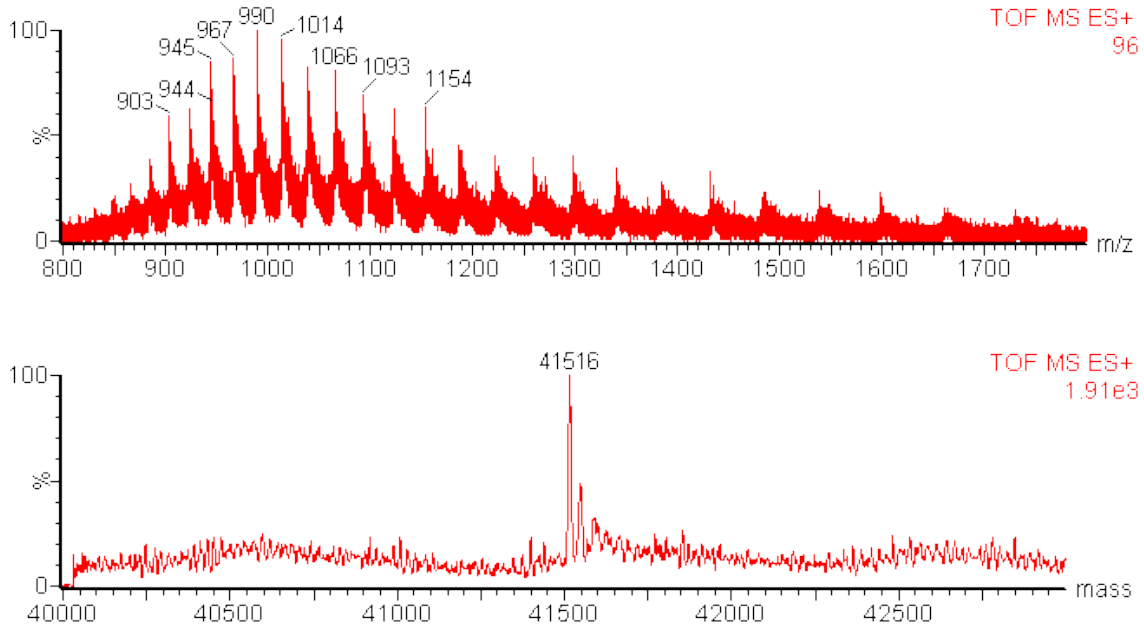


Treatment of p38α Variant 4 with Sodium Thiophosphate



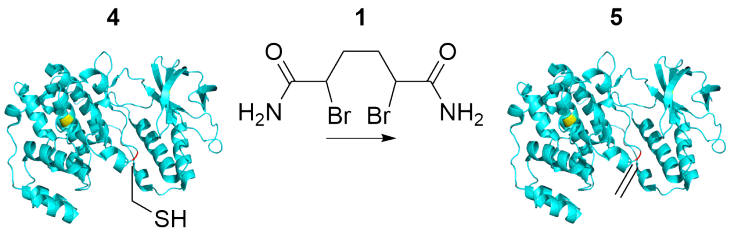
Containing:
1 surface Cys (shown)
1 buried Cys

The buffer of a freshly thawed aliquot of **4** (50 µL of a 3.5 mg/mL solution, 4.2 nmol, 1.0 eq.) was exchanged from p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl,5% glycerol, 0.5 mM TCEP) to p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) using a SpinTrap™ G-25 desalting column (GE Healthcare). After protein samples were taken, a solution of sodium thiophosphate (0.97 µL of a 700 mg/mL solution, 3.8 µmol, 1000 eq.) in water was added. The thiophosphate solution was made prior addition by dissolving sodium thiophosphate (25.8 mg) in water (10.5 µL) and balancing the pH with HCl (11.2 µL of a 5 M solution in water) to pH 7.5. After 2 h, LC-MS still showed no conversion of the starting material but after 3 h, LC-MS showed that all the start material had been consumed and converted to the disulphide thiophosphate adduct **17** (calculated mass: 41517, observed mass: 41516).



2.11.6 Chemical Modification: Optimisation Experiments

General Procedure for Treatment of p38α Variant 4 with Dibromide for Buffer Screening



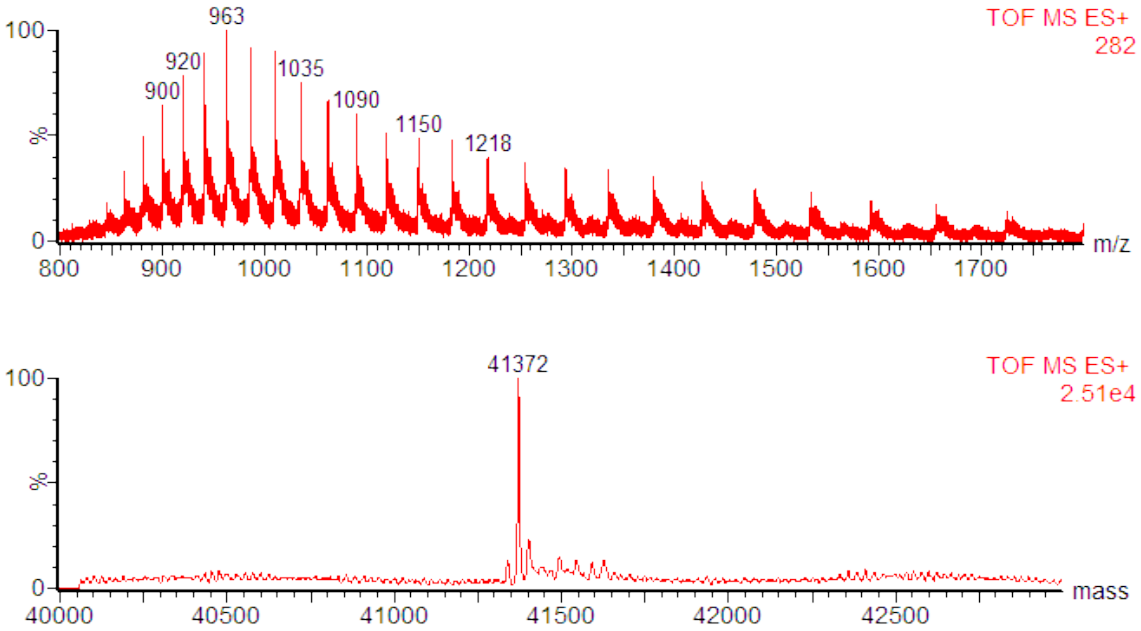
Containing:
1 surface Cys (shown)
1 buried Cys

A freshly thawed aliquot of p38α variant **4** (50 µL of a 3.5 mg/mL solution, 4.2 nmol, 1.0 eq.) in p38α storage buffer (50 mM HEPES pH 7.8, 50 mM NaCl, 5% glycerol, 0.5 mM TCEP) was diluted (50 µL) with p38α reaction buffer (either 50 mM HEPES pH 7.5, or same as p38α storage buffer but pH 7.5 and without TCEP). The diluted protein was loaded onto a SpinTrap™ G-25 desalting column with more of the appropriate reaction buffer (40 µL) as a stacker. The column had been pre-equilibrated according to the manufacturer's instructions. The protein was eluted from the column by centrifugation (800 $\times g$, 1 min, 4 °C).

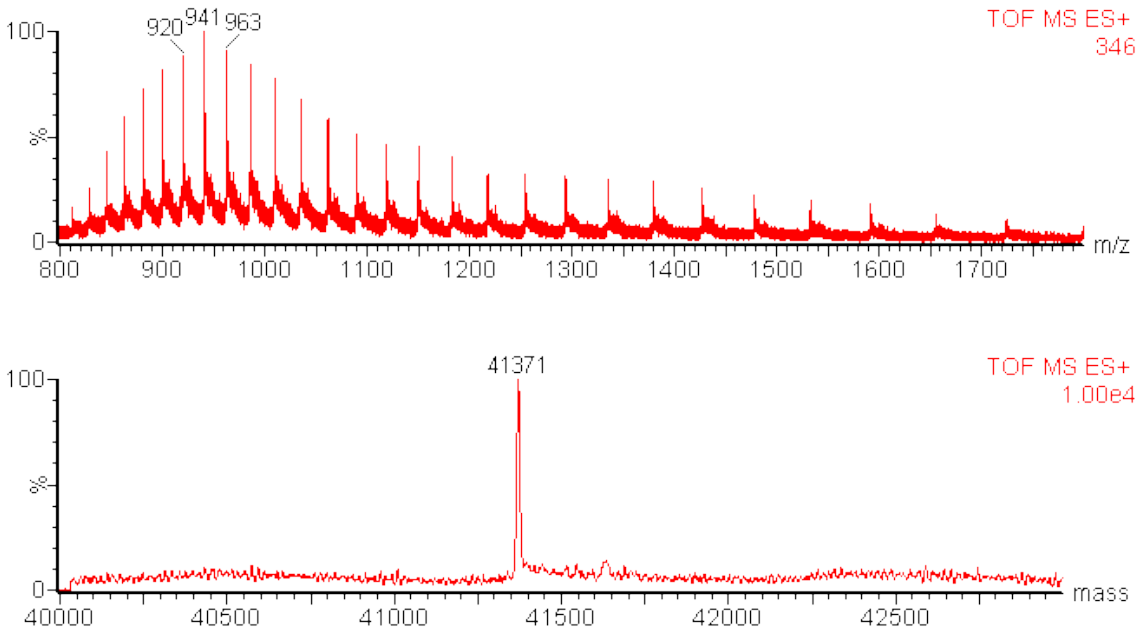
A suspension of dibromide **1** (49.1 µL of a 2.6 mg/mL suspension in reaction buffer, 0.42 µmol, 100 eq.) was added and the reaction shaken at 37 °C, 1000–1400 rpm for 4 h. The suspension had been made sep-

arately by vortexing and sonicating the dibromide in the buffer. The reaction was followed with samples taken after 1, 2.5, 3.5 and 4 h. Analysis of the samples by LC-MS showed quantitative conversion to the Dha product **5** after 4 h; (calculated mass: 41371, observed mass: 41371±1).

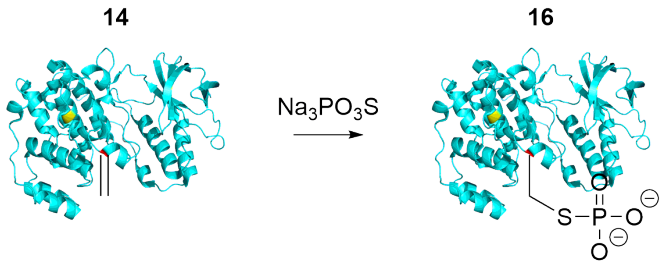
In 50 mM HEPES pH7.5:



In reaction buffer (50 mM HEPES pH7.5, 50 mM NaCl, 5% glycerol):



General Procedure for Treatment of p38α-Dha180 (14) with Sodium Thiophosphate for Buffer Composition Screening



An aliquot of p38α-Dha180 (**14**) (143 µL of a 1.15 mg/mL solution, 3.9 nmol, 1.0 eq.) in an appropriate buffer was thawed on ice before sodium thiophosphate was added portion-wise (5×5.4 µL of a 660 mg/mL solution at pH 8.0, 5×20 µmol, 5×5000 eq.) at 5 min intervals. The sodium thiophosphate solution had been made by dissolving sodium thiophosphate (107 mg) in water (43.4 µL) and the pH balanced with 5 M HCl (46.7 µL). The reaction mixture was then heated with shaking at 37 °C, 550 rpm for 3 h, following the reaction by LC-MS. LC-MS showed >70% in all cases conversion. Prolonged reaction time (6 h total) gave rise to ∼85% of the thiophosphate product **16**: (calculated mass: 41487, observed masses: 41 486–41 488).

The buffers used were:

1. p38α reaction buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 5% glycerol)
2. p38α reaction buffer with added NaCl (50 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol)
3. HEPES buffer (50 mM HEPES, pH 8.0)
4. phosphate buffer (50 mM sodium phosphate, pH 8.0)

2.11.7 Calibration Curve Determination for Species Quantification in MS

The concentration of pure samples of p38α-Dha180 and p38α-pCys180 in p38α reaction buffer (HEPES pH 8.0, 50 mM NaCl, 5% glycerol) were carefully measured using the BSA assay (Thermo Scientific). After normalisation of the protein concentration (final concentration: 0.5 mg/mL), the two chemical variants were mixed in known quantities (0, 20, 40, 60, 80 and 100% p38α-Dha180 in p38α-pCys180). Samples

of these mixtures were then analysed by LC-MS. The individual species were then quantified from the deconvoluted spectra as previous described (see Section 2.11.1).

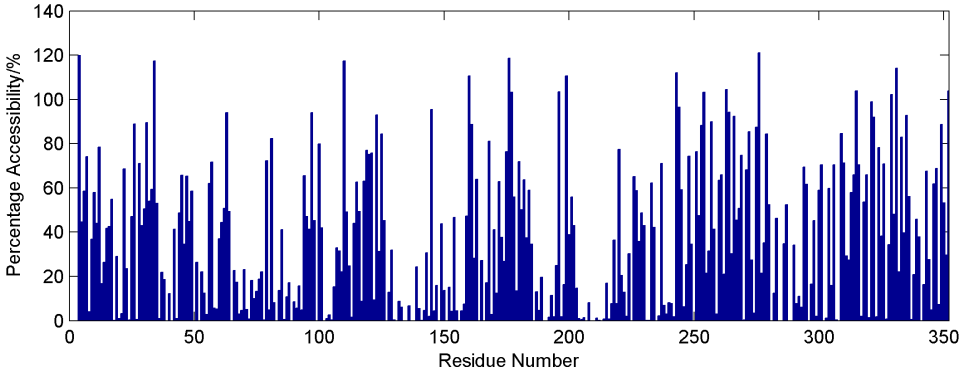
2.11.8 Model Surface and Homology Calculations

Electrostatic Calculations

The crystal structure file (1R3C.pdb) was prepared for electrostatic surface calculation using PDB2PQR[91,92] (CHARMM forcefield) before use of the ABPS PyMOL plugin[93] for calculation of the surface itself. The resultant images were manipulated in PyMOL.

Surface Accessibility Calculations

Surface accessibility was assessed using the Naccess program,[94,95] calculated using crystal strcuture file 1R3C.pdb. The default probe size (1.4 Å) was used. Conversion of the percentage data from Naccess into colour information for the figures was done using scripts written in Python. The relative percentages of total side-chain accessibility were used for this conversion. The percentage data used is plotted below. Due to the methods used in the calculation, percentages of over 100% are possible.



2.11.9 Methods for Characterising (Modified) Proteins

General Procedure for Measurement of Circular Dichroism Spectra

Prior to measuring the spectra, samples not already in p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) were desalted into this using a combination of G-25 SpinTrap™ desalting column (GE Healthcare) and repeated concentration/dilution by Vivaspin (MWCO 10 000). Samples were then diluted (0.36–0.72 mg/mL) to an appropriate final volume (190–220 µL), loaded into a cuvette with

thin pathlength (1.0 mm) before the spectra were collected using a Chirascan (Applied Photophysics). (Collection parameters: Wavelength: 180–260 nm, Wavelength step size: 1.0 nm, Repeat scans: 5). The spectra were collected as a temperature melts (Temperature range: 10 °C→90 °C→10 °C, Temperature step: 10 °C; Temperature equilibration time: 450 s). Data collected were exported as raw data and re-processed in MATLAB to give the final 3D plots (mean average of repeat scans taken, spectra smoothed using Savitzky-Golay filtering. Smoothing parameters: MATLAB function: `sgolayfilt` from Signal Processing toolbox; Polynomial order: 1; Window size: 5). Examples of the MATLAB scripts used for the data processing and plotting are given in Appendix C.

General Procedure for the Measurement of Protein Melting Temperature by Differential Scanning Fluorimetry

Samples of unmodified and Dha products derived from p38α-Cys172 and p38α-Cys180 were diluted (final stock concentration: 10 µM) with p38α reaction buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 5% glycerol). The protein was further diluted with p38α reaction buffer (7.5 µL) before 2× SYPRO orange dye (32.5 µL) was added (final sample volume: 50 µL). Samples were made up on a white PCR plate. Fluorescence for all the samples was then measured using a qPCR instrument. Assay parameters: Temperature range: 25→95 °C; Temperature step size: 1 °C; Temperature equilibration time: 1 min. Instrument: MJ Mini Personal Thermal Cycler for heating, coupled to a Mini Opticon for detection (Bio-rad). Instrument driven using software: Opticon Monitor v. 3.1.32. The collected data were then fitted to the 4-parameter logistic function in MATLAB.[84] The MATLAB scripts used in the analysis is given in Appendix D.

General Procedure for LC-MS/MS Analysis including Sample Preparation

The protocol is adapted from the one associated with the "in-solution tryptic digestion and guanidination kit" (Thermo Scientific), with modifications made to the reduction step.

Sample Preparation: A sample of the protein of interest (10.5 µL) was diluted with ammonium bicarbonate (15 µL of a 50 mM solution) in water. The volume was made up with water (1.5 µL) and the mixture heated at 60 °C for 10 min. Iodoacetamide (3 µL of a 100 mM solution in water) was then added, the reaction mixture protected from light and incubated at room temperature for 30 min. The sample was digested by the addition of trypsin (1.0 µL of a 0.1 µg/µL solution), incubating at 37 °C for 2 h. A second

batch of trypsin (1.0 µL) was added and incubation continued at either 30 °C for 18 h, or 37 °C for 2 h. The digested sample was flash frozen in liquid nitrogen and stored at −80 °C for further purification of the peptides.

Sample Purification: Peptides obtained from digestion were desalted using a C18 Sep-Pak cartridge (Waters). The cartridge was equilibrated by washing (5 mL) with solution B (65% acetonitrile, 35% water, 0.1% formic acid), followed by washing (10 mL) with solution A (98% water, 2% acetonitrile, 0.1% formic acid). The sample was loaded onto the cartridge and the cartridge further washed (10 mL) with solution A. The peptides were then eluted (2×1 mL) with solution B, collecting the two fractions separately. The majority of the acetonitrile was removed *in vacuo* before the peptides were dried by lyophilisation. Finally, the dried peptides were redissolved in solution A (20 µL).

Data Collection and LC-MS/MS Analysis: The sample of redissolved, purified peptides was given to the Mass Spectrometry Service to analyse by LC-MS/MS. In this service (run by Lingzhi Gong), the protein digests were analysed using a nanoACQUITY UPLC coupled to SYNAPT HDMS interfaced with a nanoelectrospray source (Waters Corporation, Milford, MA, USA). Protein digests were injected on a 5 mm symmetry C18 trap column (180 mm×20 mm) and washed for 1 min at 15 mL/min with 0.1% (v/v) formic acid. Peptides were then separated and eluted for MS analysis using a gradient of acetonitrile containing 0.1% (v/v) formic acid at 300 nL/min over 23 min on a nanoACQUITY UPLC column (BEH130 C18 1.7 µm particle size, 75 mm inner diameter × 250 mm length). The column temperature was set at 35 °C. The reference for the nanolockspray was set to the doubly charged peak of Glu-Fiprinopeptide B at a concentration of 500 fmol/µL, flowing at 400 nL/min. The reference was constantly infused and sampled at 30 s intervals. The eluted peptides were analyzed in the positive ionization mode over a mass range of 50–1990 m/z with a scan time of 0.6 s. The online-eluted peptides were analysed using an MSE method collecting MS/MS data using collision energy ramping from 15–35 eV.

Data Analysis: The data were collected using MassLynx (v. 4.1) and any necessary file reformatting (to .PKL file format) was done using ProteinLynx (v. 2.2.5). The data were then analysed using the MASCOT database search engine (MatrixScience, v. 2.4 CBRG Cluster): (Search parameters: Peptide tolerance: ±50 ppm, Number of ^{13}C: 2, MS/MS tolerance: 0.2 Da, Peptide charge: 2+, 3+ and 4+, Variable modifications considered: Acetyl (Protein N-term), Carbamidomethyl (C), Cys→Dha (C), Deamidated (N), Deamidated (Q), Oxidation (C), Oxidation (M) and Phospho (C)). The data displayed below are: 1.

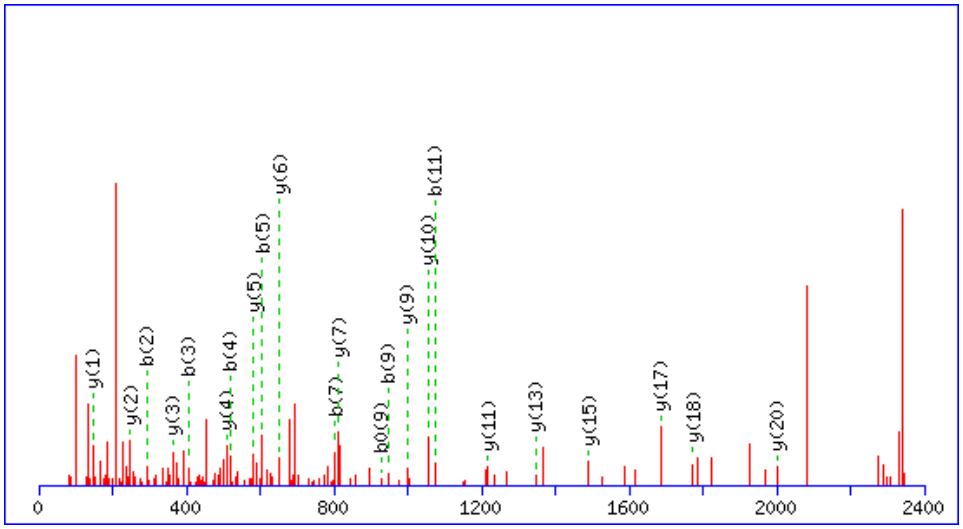
Sequence coverage of each analysis with the peptides detected in red, 2. Highlighted peptides of interest that are present, 3. MS fragmentation spectra for the corresponding peptides.

p38α-Cys172:

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLCRHTDDE MTGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

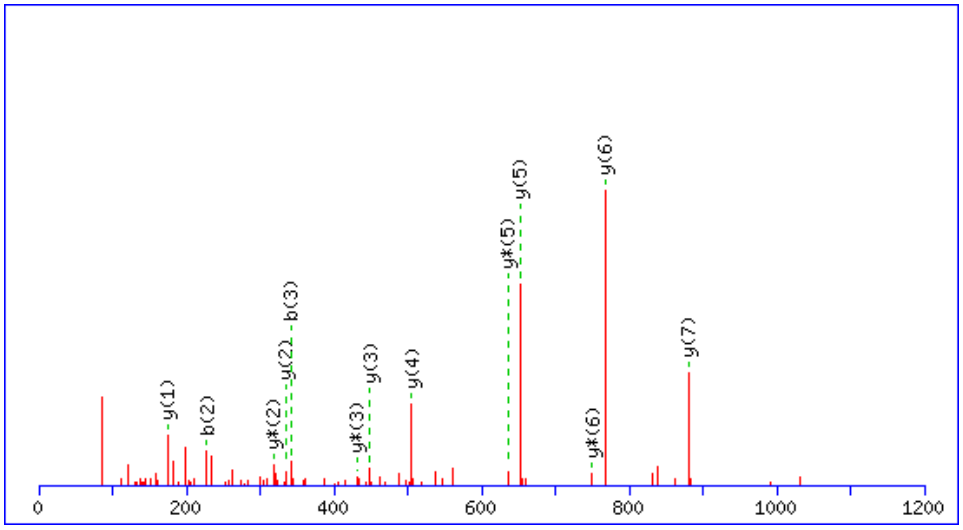
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0771	2291.0698	2291.0529	7.41	90	3.9e-09



Peptide: K.ILDFGLCR.H + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
168	175	994.5047	993.4974	992.5113	994	49	9e-05

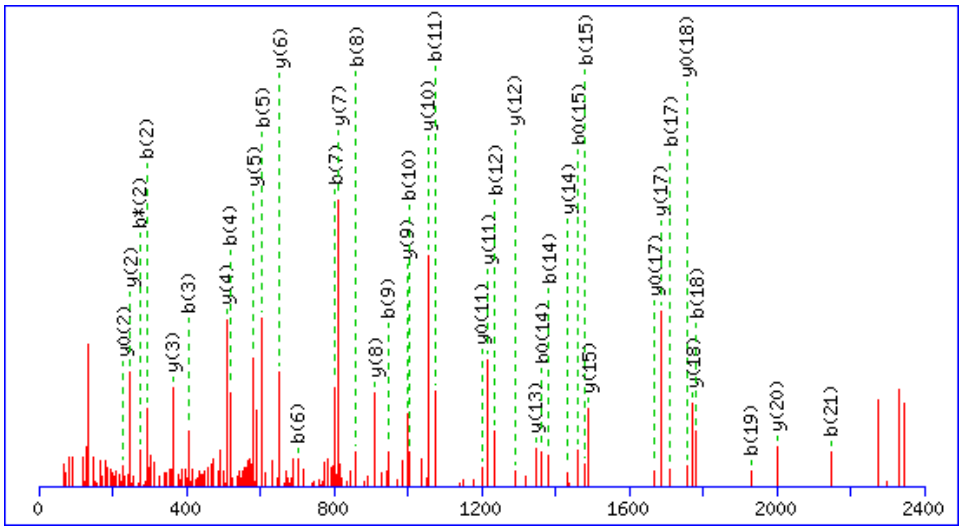


p38α-Dha172:

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLCRHTDDE MTGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

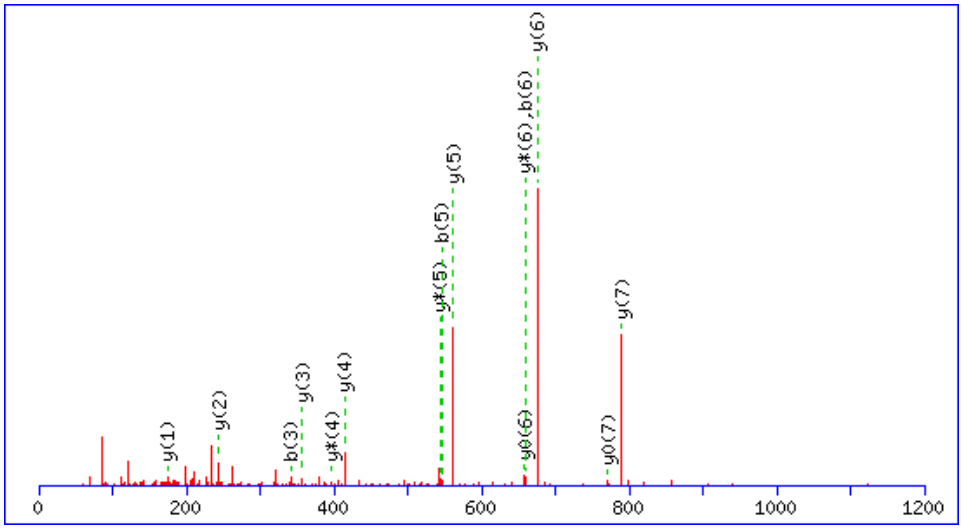
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0630	2291.0557	2291.0529	1.25	165	6.6e-16



Peptide: K.ILDFGLCR.H + Cys->Dha (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
168	175	902.5085	901.5012	901.5022	-1.05	36	0.0074

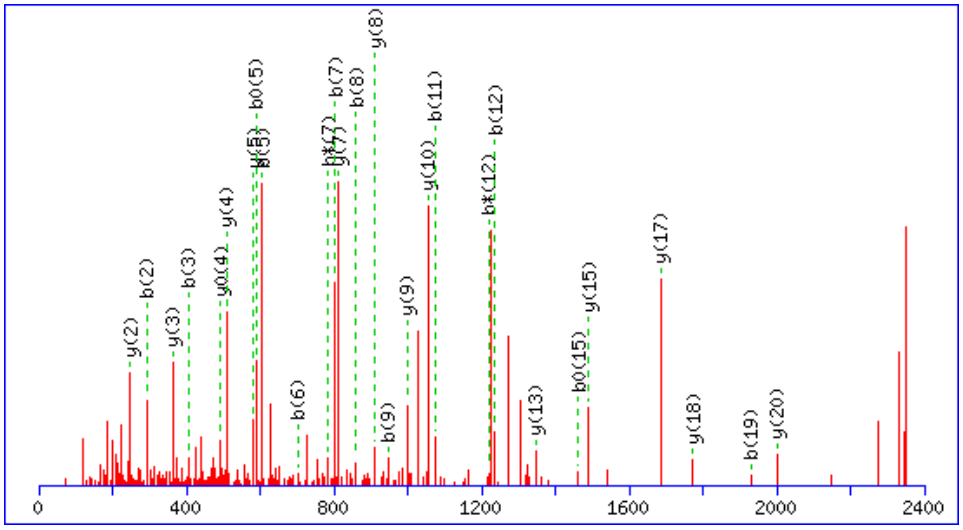


p38α-pCys172:

```
   1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
  51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
 101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
 151 RDLKPSNLAV NEDSELKILD FGLCRHTDDE MTGYVATRWY RAPEIMLNWM
 201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
 251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
 301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
 351 VPPPLDQEEM ES
```

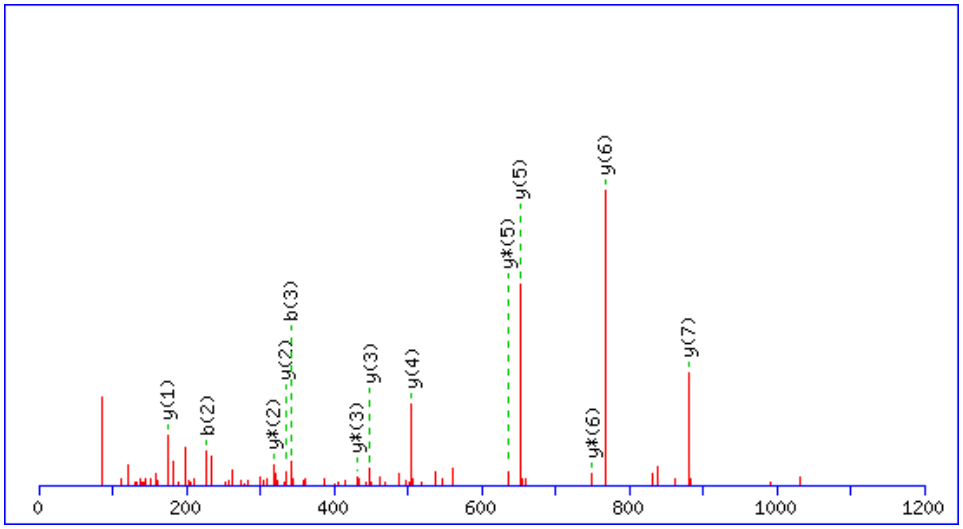
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0637	2291.0564	2291.0529	1.56	75	1e-07



Peptide: K.ILDFGLCR.H + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
168	175	994.5047	993.4974	992.5113	994	49	9e-05

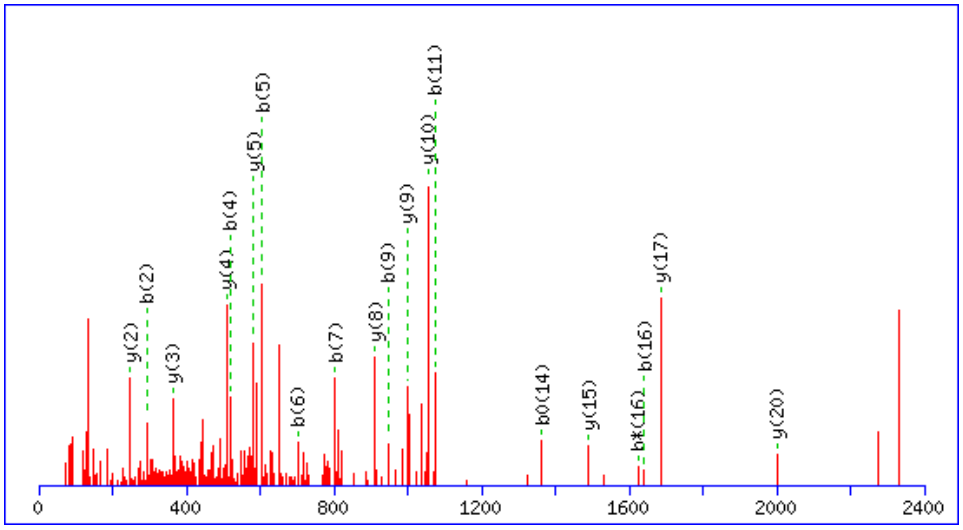


p38α-Cys180:

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

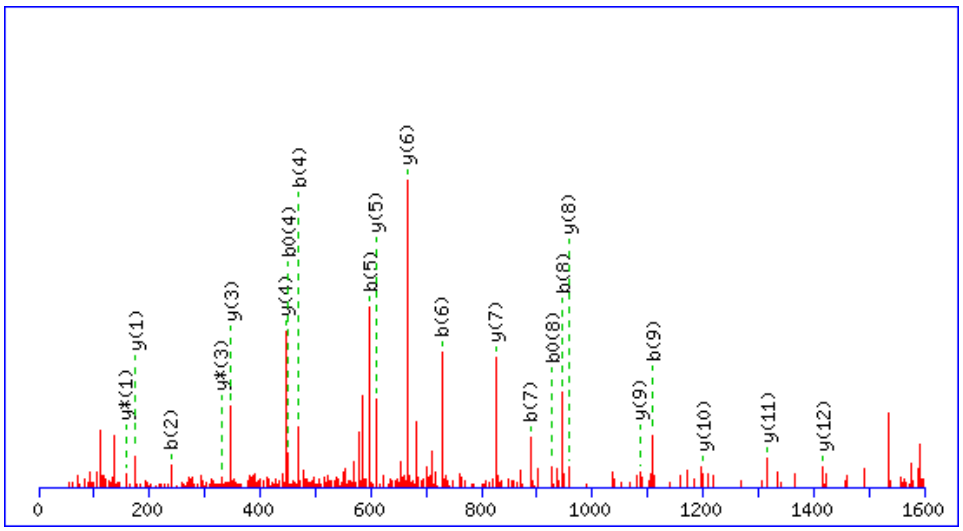
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0615	2291.0542	2291.0529	0.60	83	1.6e-08



Peptide: R.HTDDEMCGYVATR.W + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
176	188	1554.6293	1553.6220	1553.6239	-1.20	73	4.2e-07

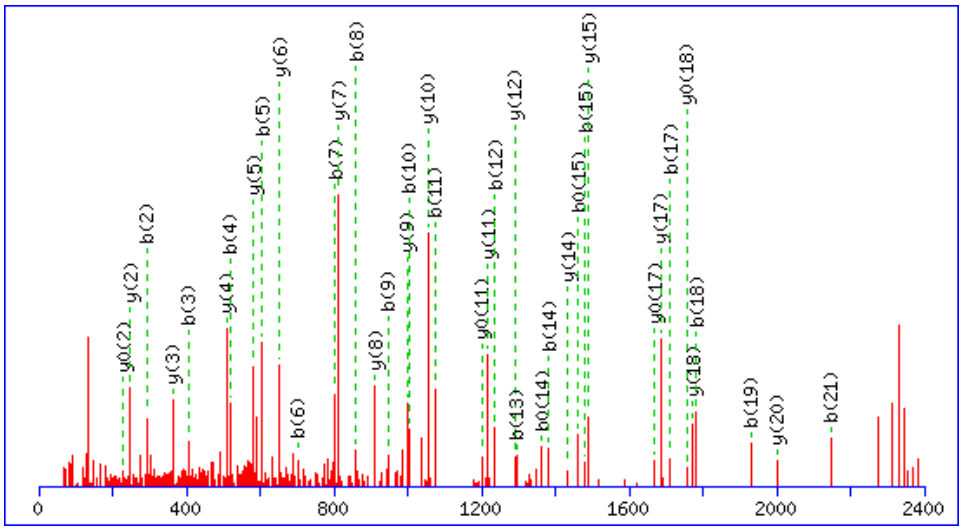


p38α-Dha180:

```
  1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
 51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
351 VPPPLDQEEM ES
```

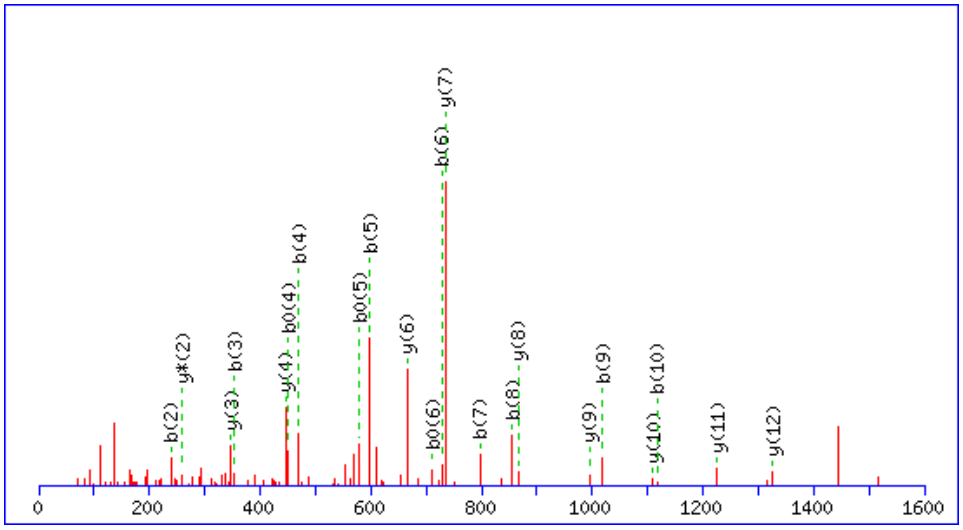
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0593	2291.0520	2291.0529	-0.36	159	3.2e-15



Peptide: R.HTDDEMCGYVATR.W + Cys->Dha (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
176	188	1463.6230	1462.6157	1462.6147	0.69	75	1.8e-07

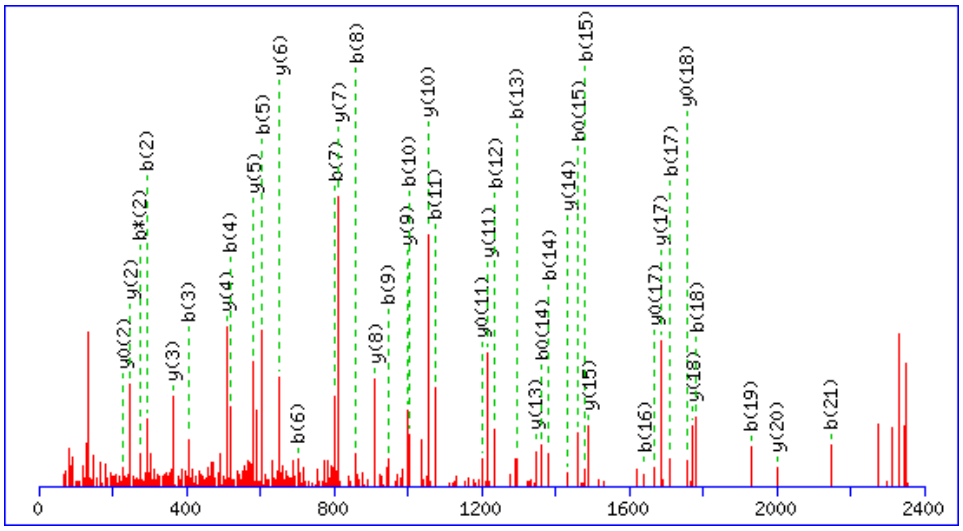


p38α-pCys180:

```
   1 GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
  51 RVAVKKLSRP FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE
 101 FNDVYLVTHL MGADLNNIVK SQKLTDDHVQ FLIYQILRGL KYIHSADIIH
 151 RDLKPSNLAV NEDSELKILD FGLARHTDDE MCGYVATRWY RAPEIMLNWM
 201 HYNQTVDIWS VGCIMAELLT GRTLFPGTDH IDQLKLILRL VGTPGAELLK
 251 KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM LVLDSDKRIT
 301 AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
 351 VPPPLDQEEM ES
```

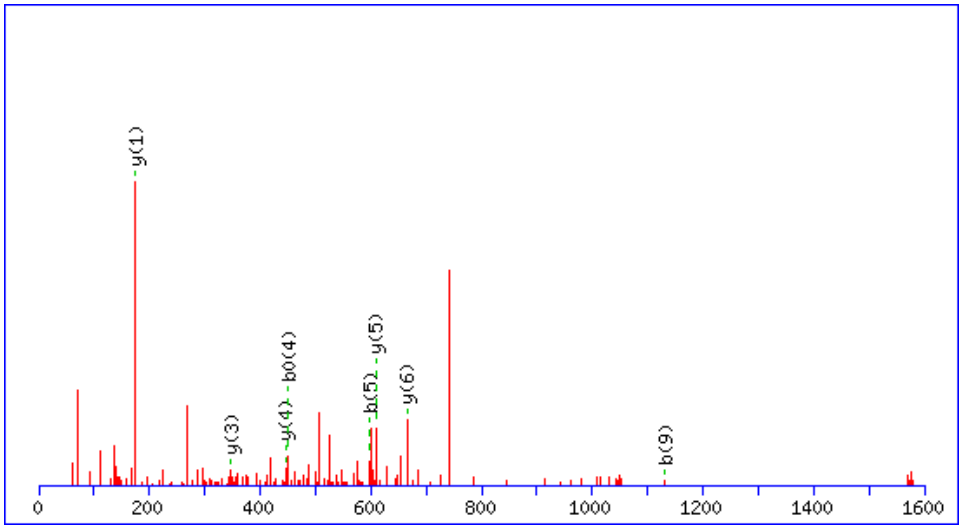
Peptide: R.YQNLSPVGSGAYGSVCAAFDTK.T + Carbamidomethyl (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
26	47	2292.0618	2291.0545	2291.0529	0.73	173	9.6e-17



Peptide: R.HTDDEMCGYVATR.W + Phospho (C)

Start	End	Observed	Mr(expt)	Mr(calc)	Δppm	Score	Expect
176	188	1577.5786	1576.5713	1576.5687	1.64	6	0.28



LC-MS/MS Analysis for Detection of Phosphopeptides from p38α Variants

Protein digestion, LC-MS/MS analysis and subsequent analysis of the data was done by Ritu Raj.

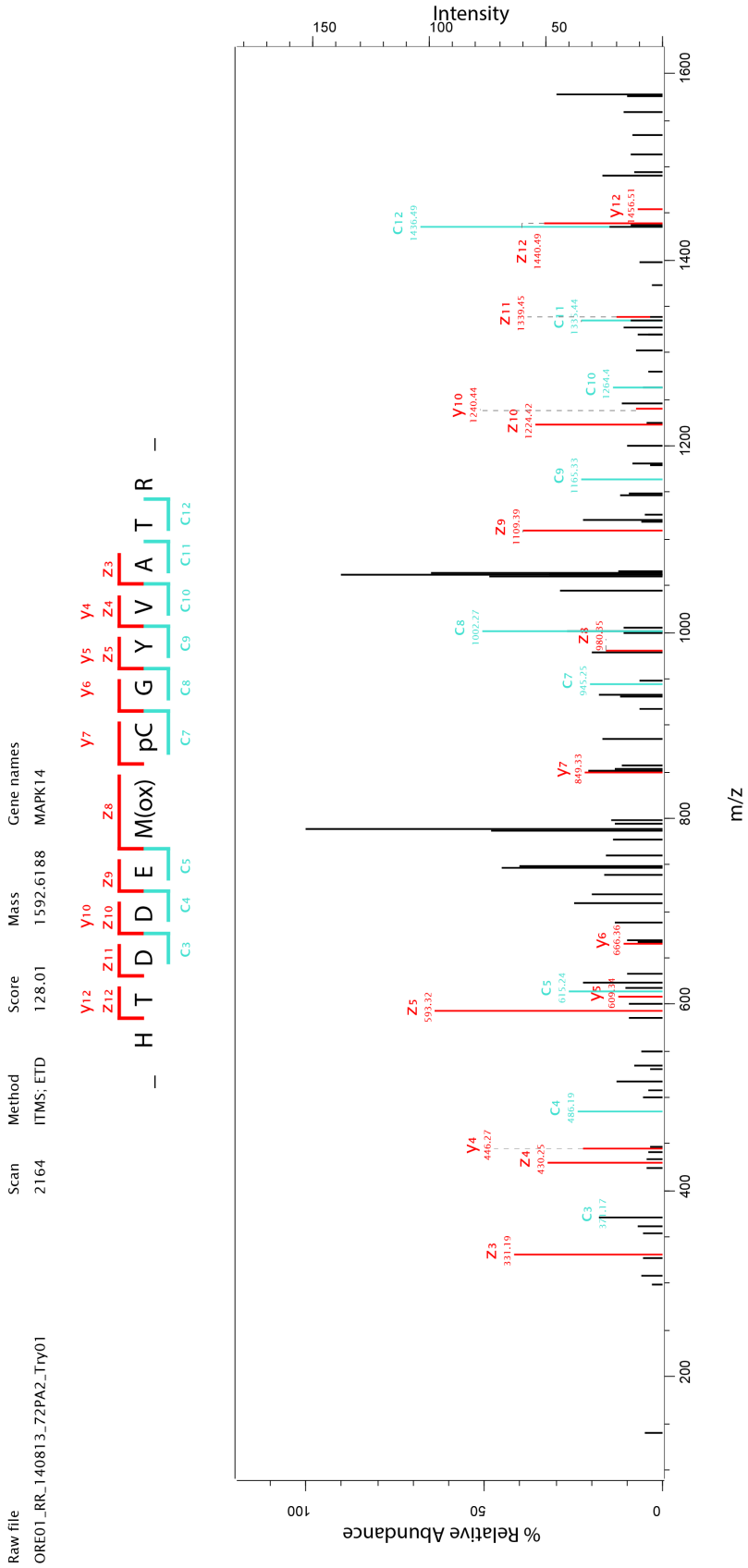
Protein Digestion: p38α-pCys180 (40 µg) was buffer exchanged using Vivaspin™ (MWCO 10 000 Da) against ammonium bicarbonate buffer (50 mM solution) to a final protein concentration of 0.4 mg/mL. Urea (48 mg, 8 M final concentration) was then added and the sample incubated at room temperature for 10 min with occasional vortexing. The protein was reduced with DTT (200 mM solution in water) at 56 °C for 25 min and alkylated with iodoacetamide (400 mM solution in water) at room temperature for 30 min in the dark. The sample was diluted 4× with ammonium bicarbonate buffer and trypsin added (final enzyme:protein 1:50 (w/w) ratio). The sample was incubated overnight at 37 °C. The digested protein was diluted (final concentration: 100 fmol/µL) with water. Variant p38α-pCys172 was treated similarly.

LC-MS/MS Analysis: The samples were analysed on an Orbitrap Elite (Thermo Fisher Scientific, DE) connected to a UHPLC Proxeon EASY-nLC 1000 and an EASY-Spray nano-electrospray ion source with EASY-Spray column (Thermo Fischer Scientific, DK). Peptides were trapped on an Acclaim PepMap® trapping column (100 µm i.d. × 20 mm, 5 µm C18) and separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 500 mm, RSLC C18, 2 µm, 100 Å). Samples were loaded at a pressure of 500 bar with 100% solvent A (0.1% formic acid in water) and the peptides separated by a linear gradient (length: 15 min, 7% to 30% solvent B (0.1% formic acid in acetonitrile), flow rate: 200 nL/min). Full

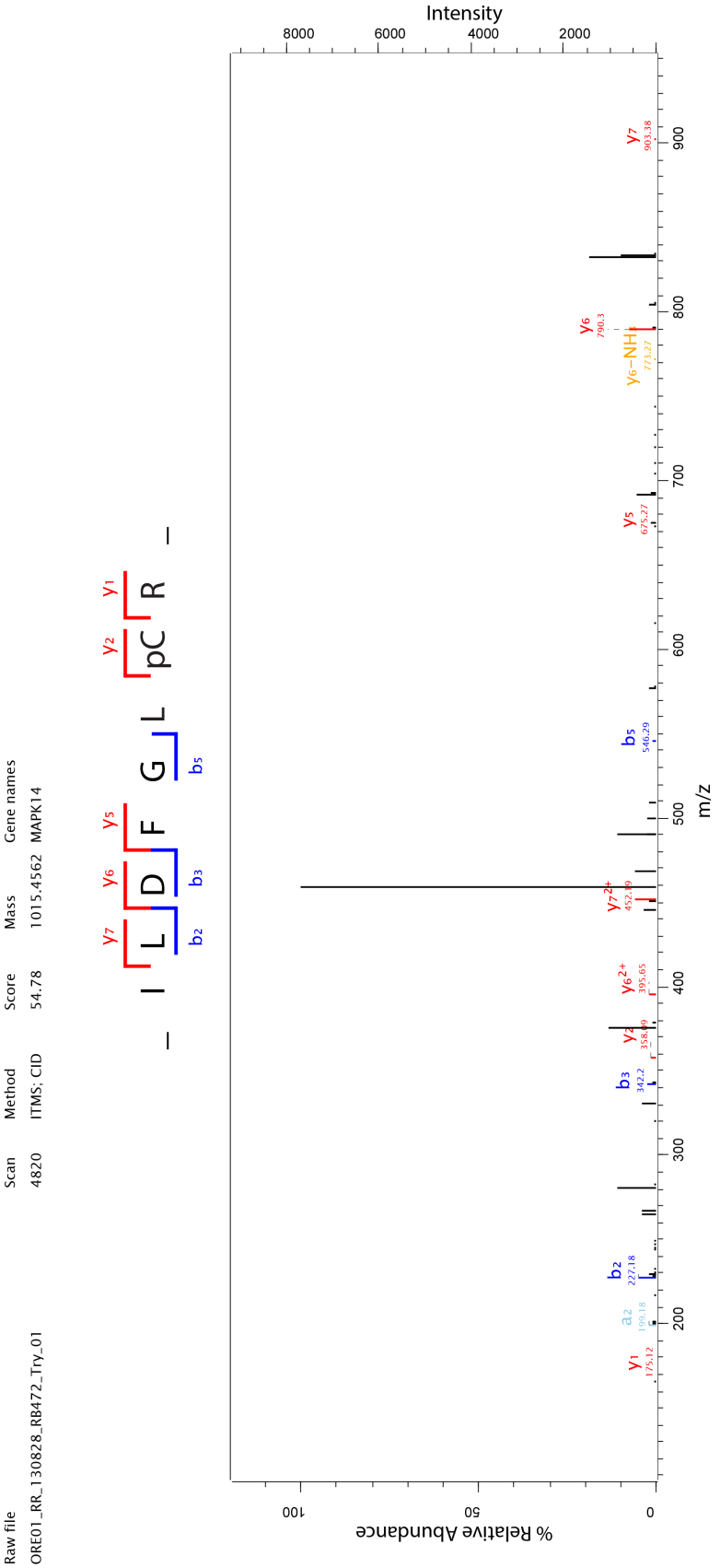
scan MS were acquired in the Orbitrap (350-1500 m/z, resolution 120 000, AGC target 1×10^6, maximum injection time 250 ms). CID and ETD spectra were acquired in Ion Trap (resolution 7500, AGC cation target 3×10^4, AGC Anion target 2×10^5, maximum injection time 100 ms). After the MS scans, the 5 most intense peaks were selected for fragmentation based on data-dependent decision tree (DDDT). The signal threshold for fragmentation of parent ion was set to 500 ion counts. For CID fragmentation, the normalized collision energy and default charge state was set to 35% and 2 respectively. For ETD fragmentation, reaction time was set to 50 ms and supplemental activation was enabled. Charge state screening was enabled; parent ion with unassigned charge state and charge state 1 was rejected. Dynamic exclusion was enabled (exclusion size list 75, exclusion duration 5 s). The setting for DDDT was as such: ETD fragmentation was followed for charge states 3, 4 and 5 with m/z less than 750. For charge states greater than 5, ETD fragmentation was always performed. In all other cases, CID activation was performed. In a separate experiment, HCD was used. All the settings for the instrument were kept the same, except HCD spectra were acquired in Orbitrap and the 3 most intense peaks were selected for fragmentation and normalized collision energy was set to 32%.

Data Analysis: The raw data files generated were processed using MaxQuant software (Version 1.4.1.2), integrated with Andromeda search engine as described elsewhere.[96,97] For identification of phosphocysteine peptides, Andromeda searched peak lists against the sequence of p38α, as well as against a list of common contaminants. Trypsin was selected as specific digestion mode with maximum number of missed cleavages set to 4. In case of p38α-pCys180 protein; acetylation (N-term), carbamidomethylation (C), cysteine to serine, oxidation (M), phosphorylation (STY), threonine to carbamidomethylation, threonine to cysteine, threonine to Dha, and threonine to phosphocysteine were used as variable modifications. For p38α-pCys172, similar rational variable modifications were used. All spectra were manually validated.

p38α-pCys180:



p38α-pCys172:

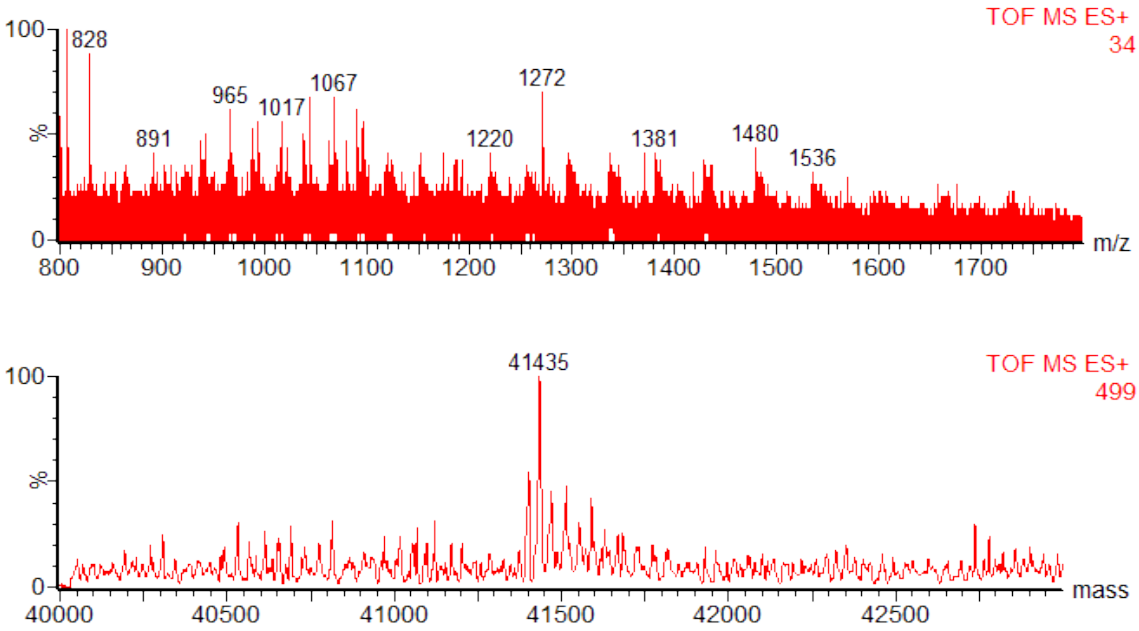


2.11.10 Procedures Used in Optimising LC-MS/MS Analysis Protocol

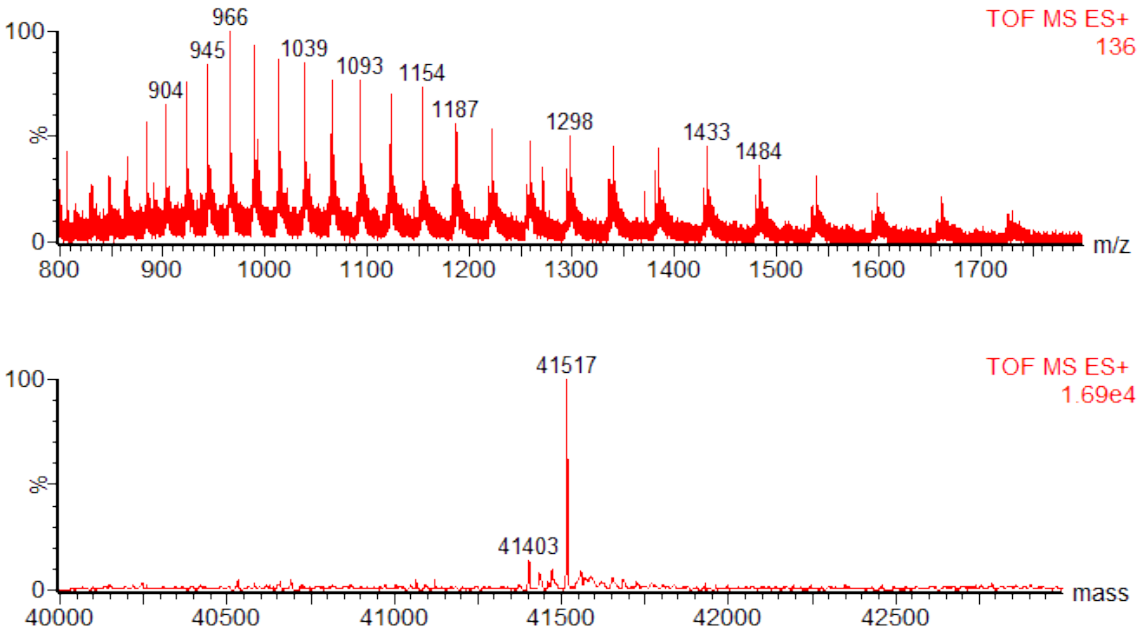
Test for Phosphocysteine Stability against Temperature

Samples of p38α-pCys172 were incubated at either 60 °C or 95 °C for 10 min. The samples were then analysed by LC-MS.

After treatment at 95 °C:



After treatment at 60 °C:



General Procedure for Investigating the Necessity of DTT in Sample Preparation

To a solution of p38α-pCys180 (10.5 µL) in p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) was added DTT solution (5 µL of a 200 mM solution in 100 mM Tris pH 7.8) and the mixture incubated at room temperature for 1 h. Iodoacetamide solution (20 µL of a 200 mM solution in 100 mM Tris pH 7.8) was added and the mixture incubated at room temperature for a further 1 h. Finally, more DTT solution (20 µL) was added and the mixture incubated for a final 1 h. Trypsin (2 µL of a 0.1 µg/µL solution in trypsin storage buffer) was added and the reaction put to 37 °C for 20 h. The reaction was flash frozen in liquid nitrogen and stored at −80 °C before subsequent purification by solid phase extraction. The steps involving DTT were variably omitted and in the case of omission, heating at 60 °C for 10 min was used for denaturation.

General Procedure for Proteolytic Digest with Alternative Proteases

A solution of p38α-Cys180 kinase ($\sim$ 1 mg/mL) in p38α reaction buffer (50 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol) was diluted with ammonium bicarbonate solution. The mixture was heated at 60 °C for 10 min. After cooling to room temperature, iodoacetamide solution was added, the reaction protected from light and incubated at room temperature for 30 min. For digestions with pepsin, hydrochloric acid (1 µL of a 2.5 M solution in water, final pH: $\sim$1.5) was added. The kinase was then digested by addition of proteolytic enzyme, incubating at 37 °C. After digestion, samples were either diluted and directly analysed by LC-MS/MS by the Mass Spectrometry service, or desalted before the samples were analysed, as described previously (see Section 2.11.9).

The exact quantities and enzymes used for all the trials are summarised in the table:

Entry	Enzyme	Kinase Volume/µL	[$(NH_4)HCO_3$]/mM	[$(NH_4)HCO_3$] Volume/µL
1	pepsin	10.5	50	15
2	Asp-N	20	100	88
3	Glu-C	20	100	88

Entry	Iodoacetamide Volume/µL	Reaction Time/h	[Enzyme]/mgmL^{-1}	Enzyme volume/µL
1	3	1	0.5	1
2	12	18	0.04	5
3	12	18	0.1	4

Chapter 3

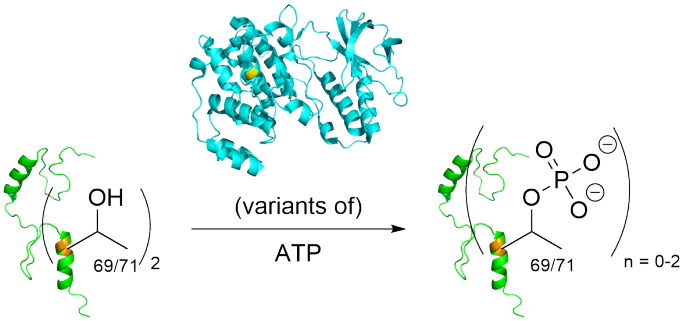
Functional Evaluation of p38α Kinase after Chemical Modification

Having successfully installed phosphocysteine in p38α-Cys172 and p38α-Cys180 in the previous chapter, the enzymatic activity of the resulting chemical variants, p38α-pCys172 and p38α-pCys180 needed to be tested. To do this, a suitable assay had to be developed.

3.1 p38α Preliminary Activity Assay

ATF2 (activating transcription factor 2) was chosen as the protein substrate for the activity assay as it is a natural substrate of p38α, and because the protein was readily available. A protein, rather than a peptide,[98,99] was seen to be important as a substrate since one of the questions to be answered is whether phosphocysteine is a sufficient mimic of phosphothreonine, in the context of mimicking kinase phosphoforms. For this question to be sufficiently investigated and challenged, the chemically modified kinase must be biologically recognised, which requires a protein substrate.

The preliminary activity assay aimed to determine if there was any enzymatic activity at all among the chemically modified derivatives (Dha and pCys) derived from p38α-Cys172 and p38α-Cys180, as well as any activity in the parent species themselves. A number of factors caused doubt as to whether any of the p38α chemical variants would display enzymatic activity:



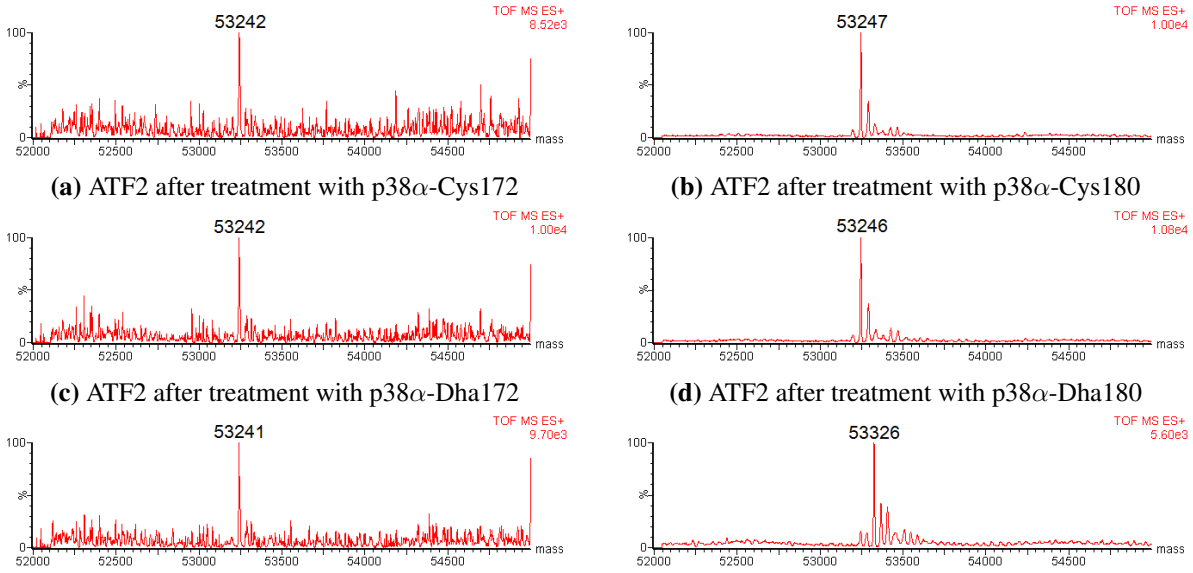
Scheme 3.1 ATF2 is phosphorylated by p38α using ATP as a co-factor.

1. phosphocysteine is a suitable mimic of phosphothreonine electrostatically, but not sterically; phosphothreonine has a methyl group which is not present in phosphocysteine. Enzymatic activity would be difficult to discern if sterics is the more important factor in giving an active conformation in p38α.

2. The p38α chemical variants were at most only phosphorylated once, whereas p38α requires dual phosphorylation for full activation.

Considering that these two factors could potentially reduce the signs of a p38α chemical variant being active, a super-stoichiometric quantity of kinase ([p38α variant] $=$ 4 µM, [ATF2] $=$ 3 µM) was used to ensure that any activity near the basal level would be detected. Basal activity is the low level of enzymatic activity that some kinases possess, even without activation by phosphorlyation, and can thus be variable in magnitude between kinases (also see Section 1.3). MS was used as the substrate detection method, having been already used extensively in the chemical modification stage of the work. All of the different p38α species were then tested to determine whether they would phosphorylate the ATF2 substrate (Scheme 3.1).

No detectable levels of ATF2 phosphorylation were observed for the chemical variants from p38α-Cys172, while for the p38α-Cys180 derivatives, phosphorylated ATF2 was gratifyingly only found in the presence of p38α-pCys180 (Fig. 3.1). This result also suggested that all the inactive chemical variants did not possess any basal activity. MS detection of enzymatic reaction products had thus shown that p38α-pCys180 had activity, but given the variable noise/signal ratio of this detection method, any extremely low levels of activity from the other chemical variants could not be ruled out.



(a) ATF2 after treatment with p38α-Cys172

(b) ATF2 after treatment with p38α-Cys180

(c) ATF2 after treatment with p38α-Dha172

(d) ATF2 after treatment with p38α-Dha180

(e) ATF2 after treatment with p38α-pCys172

(f) ATF2 after treatment with p38α-pCys180

Fig. 3.1 Deconvoluted MS spectra of ATF2 after 3 h of treatment with various p38α chemical variants. Only **(f)** shows a mass corresponding to single phosphorylation of ATF2. Expected mass change from unmodified ATF2 upon single phosphorylation: $+80$.

3.2 Preliminary Radioassay for p38α Chemical Variants

To confirm the findings of the preliminary MS study, and to confirm that the inactive p38α chemical variants did not possess basal activity, incorporation of radioactive phosphate was used as a complementary assay method. In this method, [γ-^{32}P]-ATP is included in the enzymatic reaction mixture. Since it is the γ-phosphate of ATP that is transferred onto the substrate, the use of [γ-^{32}P]-ATP causes the substrate to be radioactively labelled. Subsequent separation of the enzymatic reaction mixture by SDS-PAGE and gel autoradiography then completes the phospho-substrate detection, a process which will henceforth be referred to as electrophoretic radioassay (ERA). The choice of ERA was for a number of reasons:

1. Due to the sensitivity of this method, it is regarded as a "gold standard" for kinase assays[100] and therefore suitable for detection of low levels of enzymatic activity.

2. With the phospho-substrates being directly detected, the results from ERA could be directly compared with those from MS detection.

3. With fewer assay components and therefore fewer variables, assays with direct detection can be quantified with more certainty.

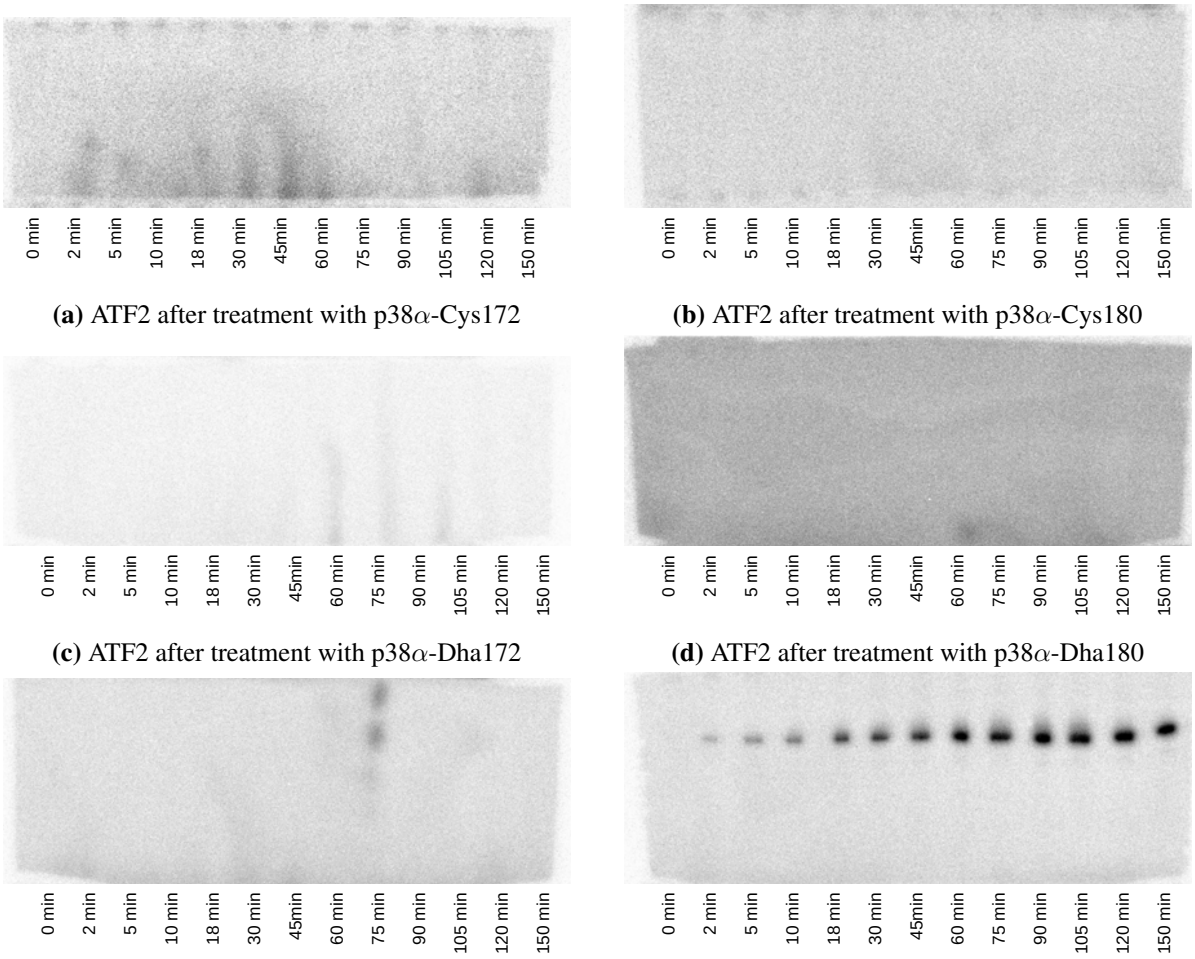
Use of the ERA, over similar enzymatic reaction timescales as those used for the MS detection assay, confirmed that all the p38α chemical variants did not have basal activity, except for p38α-pCys180, which was the only enzymatically active chemical variant. p38α-pCys180 was thus subsequently the only chemical variant taken on for further studies (Fig. 3.2). The activity studies on the p38α-Cys172 derived chemical variants acted as a negative control, since they showed that phosphorylation at the native phosphorylation site was required for p38α activation, and that mere phosphorylation on the activation loop of p38α was not sufficient. It is interesting to note that the Dha variant p38α-Dha180 also did not have activity. The sp^2 centre introduced on Dha incorporation would have a rigidifying effect on the activation loop, which in turn could cause conformation change in the p38α variant and could thus confer activity. Finally, the sensitivity of ERA was also demonstrated with ^{32}P incorporation being clearly visible, even as soon after 2 min in the timecourse of the p38α-pCys180 experiment (Fig. 3.2f, left-most dark spot). In comparison, MS from the p38α-pCys180 experiment was only able to conclusively detect phosphorylated ATF2 after $\sim$30 min (Fig. 3.3).

3.3 Comparison of MS and Radiolabel Detection Assay Methods

Working on p38α-pCys180 as the remaining p38α chemical variant of interest meant that the high sensitivity of ERA was no longer required, since its activity could be suitably detected by MS. Furthermore, since ATF2 has two p38α phosphorylation sites, Thr69 and Thr71, the use of MS detection had the advantage that it could discern unmodified substrate from mono- and bis-phosphorylated ATF2 (ATF2-P and ATF2-PP respectively). This would potentially give a deeper insight into the mechanistic function of p38α, so the assay by MS detection was developed further.

As was the case in the chemical modification stage of this work (Section 2.8.1), the quantitative nature of the MS data had to be tested and verified. Since pure phosphoforms of ATF2 cannot be obtained, the calibration curve could not be made using MS alone. Instead, another MS detection assay was run with p38α-pCys180, using the same conditions and timepoints as the ERA experiment. The data from the two experiments were then directly compared. In doing so, the following assumptions were made:

1. All of the ATF2 species ionise in a similar fashion (same as in Section 2.8.1). Given the different charges of the ATF2 species at different levels of phosphorylation (ATF2, ATF2-P and ATF2-PP), this is a significant assumption.



(a) ATF2 after treatment with p38α-Cys172

(b) ATF2 after treatment with p38α-Cys180

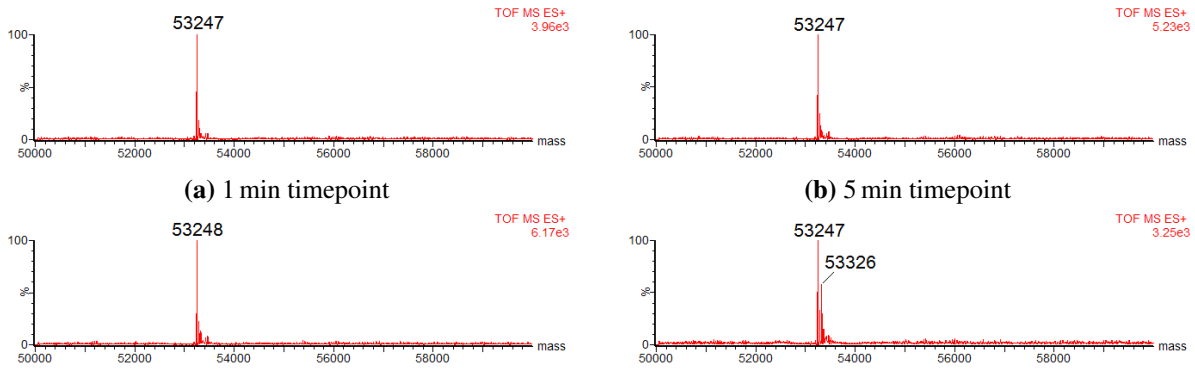
(c) ATF2 after treatment with p38α-Dha172

(d) ATF2 after treatment with p38α-Dha180

(e) ATF2 after treatment with p38α-pCys172

(f) ATF2 after treatment with p38α-pCys180

Fig. 3.2 Gel autoradiographs of timecourse experiments over 2.5 h where ATF2 was treated with various p38α chemical variants. Only **(f)** shows radiolabel incorporation onto ATF2, being the only gel with dark spots.



(a) 1 min timepoint

(b) 5 min timepoint

(c) 10 min timepoint

(d) 30 min timepoint

Fig. 3.3 Deconvoluted MS spectra taken over the timecourse experiment of ATF2 treatment with p38α-pCys180. The appearance of a new peak is only apparent in **(d)**.

2. ATF2-PP in ERA gives twice the radioactive signal as ATF2-P, since it has been doubly labelled.

3. The component signal from ATF2-PP in ERA corresponds to double the MS signal from the same ATF2 species, again because this species has been doubly labelled in ERA.

It then follows that:

$$[\gamma\text{-}^{32}\mathrm{P} \text{ on ATF2}]_{\mathrm{ERA}} = [\text{ATF2-P}]_{\mathrm{MS}} + 2[\text{ATF2-PP}]_{\mathrm{MS}} \tag{3.1}$$

where the quantities on both sides of Equation 3.1 can be termed as "effective phosphorylation signal". The direct comparison (Fig. 3.4) showed that there was a reasonable profile match in the effective phosphorylation signal from both ERA and MS. Further justification on the continued use of MS detection lay in its potential for richer datasets, hence it was adopted for more extensive use.

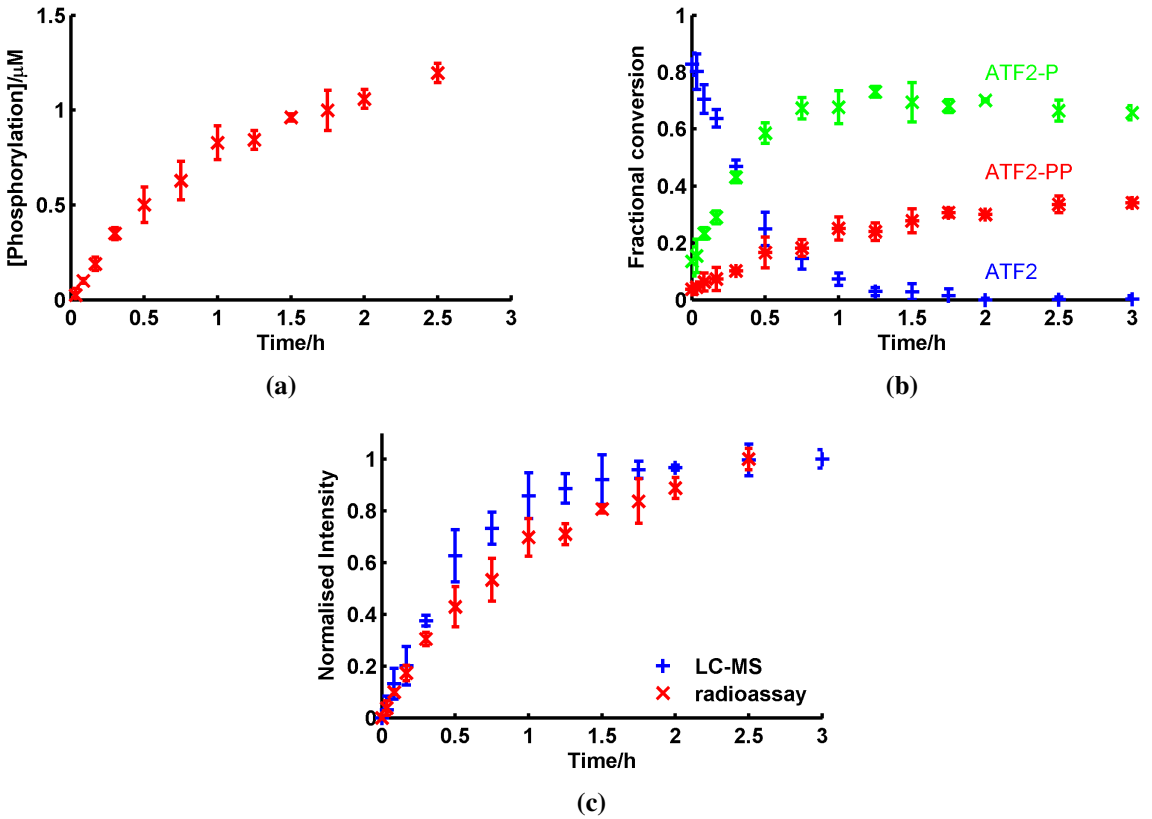
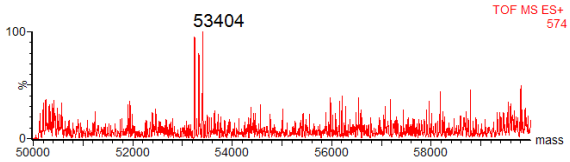


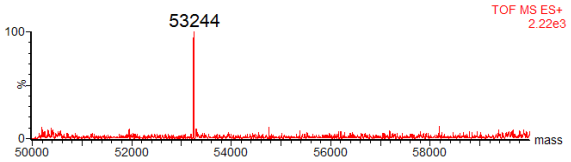
Fig. 3.4 The datasets from **(a)** ERA and **(b)** MS were **(c)** directly compared.

3.4 Wild-Type Controls

A wild-type comparison was needed for the enzymatically active, p38α chemical variant p38α-pCys180. p38α-pThr180 would be the direct analogue of the synthetic mimic, however, its production involves extensive protein purification and can thus be difficult to produce.[33] Further, a production method reliant on biological methods and extensive post-expression purification may not always lead to species homogeneity.[34,101] Rather than face potential technical difficulties in producing homogeneous p38α-pThr180, the most widely available form of active p38α was taken as the standard, which would also be a more representative comparison. Thus, commercially available, enzymatically activated wild-type p38α (termed "active p38α-WT" and p38α-pThrpTyr) was used. As expected, p38α-pThrpTyr was shown to be active upon its use in catalytic quantities (Fig. 3.5a) while inactive p38α-WT (Fig. 3.5b) was not.



(a) ATF2 after treatment with active p38α-WT

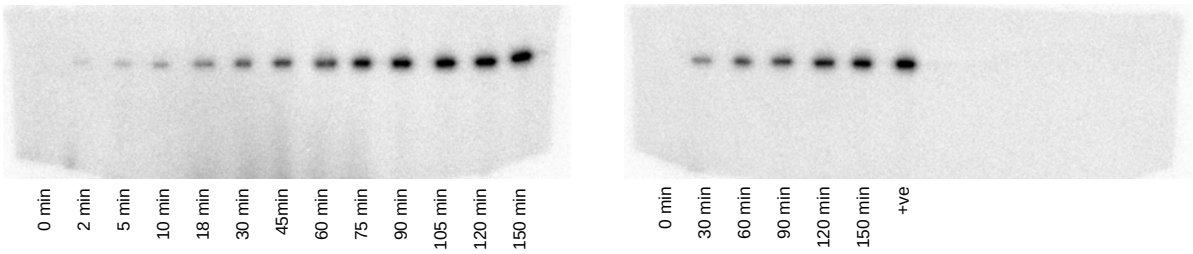


(b) ATF2 after treatment with inactive p38α-WT

Fig. 3.5 Deconvoluted MS spectra of ATF2 after 0.5 h treatment at room temperature with p38α-WT of various activation states. Only **(a)** shows masses corresponding to mono- and bis-phosphorylation of ATF2. Expected mass change from unmodified ATF2 upon mono-phosphorylation: +80.

A more stringent assessment of activity for the wild-type species was subsequently made with ERA. Active p38α-WT in the same catalytic concentration ([p38α-WT] = 80 nM, [ATF2] = 3 nM) as that used in the MS assay was tested and shown to be active, as expected (Fig. 3.6a). Conversely, the result from the MS assay suggested detection of activity with inactive p38α-WT was not expected, so on repeating the experiment using the ERA, the stringency on conferring inactivity was increased by increasing the concentration of inactive p38α-WT used ([inactive p38α-WT] = 800 nM, [ATF2] = 3 nM). A 10$\times$ higher concentration of inactive p38α-WT was used, compared to that used for active p38α-WT. Surprisingly however, activity was observed in this "inactive" form (Fig. 3.6b).

MS of the two forms of p38α-WT showed that while the active form was predominantly bis-phosphorylated (Fig. 3.7a), the "inactive" form was in fact partially mono-phosphorylated (Fig. 3.7c). This partial phosphorylation would explain why this batch of p38α-WT was partially active. It is also likely that completely unphosphorylated p38α-WT does not have any basal activity, and there are no reports of basal activity in the literature. The exact cause of partial phosphorylation is mere speculation



(a) ATF2 after treatment with active p38α-WT **(b)** ATF2 after treatment with inactive p38α-WT

Fig. 3.6 Gel autoradiographs of timecourse experiments over 2.5 h where ATF2 was treated with p38α-WT of varying activation state. Both phosphorylation states of p38α-WT showed radiolabel incorporation onto ATF2.

since no details are given in the product manual,[102] but it is likely to be due to the bacterial strain used in expressing the protein.[103] The mixture of phosphoforms found in the commercial p38α-WT samples, particularly in active p38α-WT, only highlights the difficulty in producing highly homogeneous batches of kinase phosphoforms by biological means, a problem which a semi-synthetic, chemical modification approach is able to overcome (Fig. 3.7b and 3.7d).

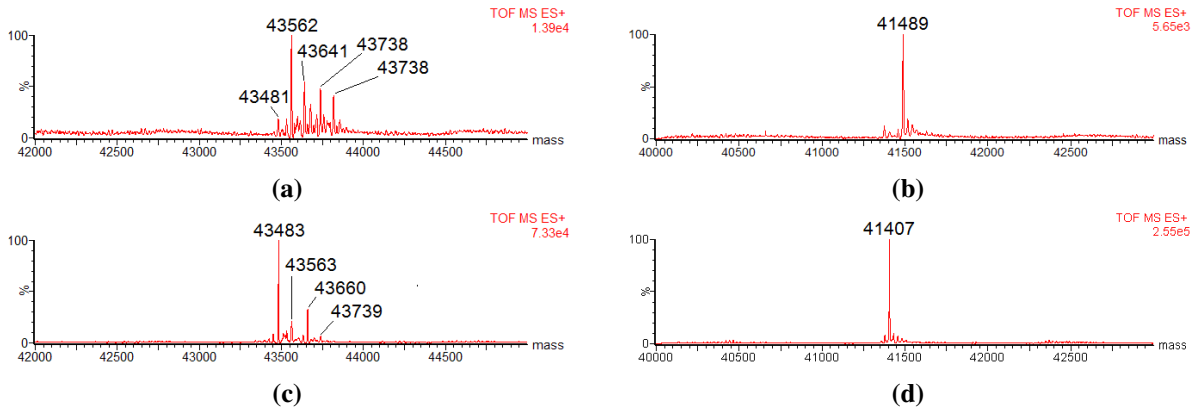


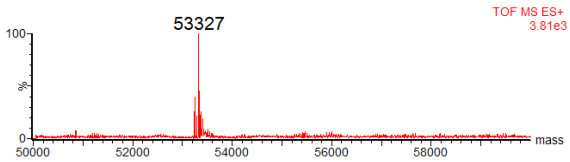
Fig. 3.7 Deconvoluted MS of biologically and semi-synthetically derived p38α, both before and after activation. **(a)** Enzymatically activated kinase exhibits three different phosphorylation states (observed masses 43481, 43562 and 43641 for the ungluconoylated protein)[104] while conversely, **(b)** the chemically modified one only displays one phosphoform. **(c)** Even prior to activation, p38α-WT displays partial phosphorylation due to phosphorylation during expression whilst **(d)** the synthetic precursor is pure. Figure reproduced from supporting information.[1]

3.5 Kinetic Comparison of Active Forms of p38α

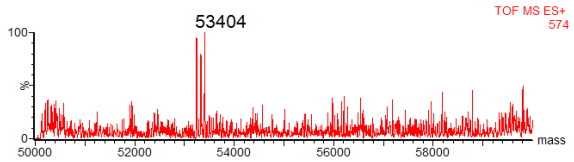
3.5.1 Use of MS Detection in Kinetic Assay

In order to make a valid kinetic comparison between the two activated forms of p38α, the concentrations used in both sets of experiments would have to be identical, in contrast to the preliminary ex-

periments described so far. The main reason for using different p38α concentrations in previous experiments was because the chemical variant p38α-pCys180 was much less active than active p38α-WT (Fig. 3.8). This difference in activity whereby the mono-phosphorylated p38α is much less active than the bis-phosphorylated form is also corroborated in the literature.[33,52] Therefore, to complete the assay in a practical time while still allowing for some gradation of assay signal, different concentrations of p38α had been used: 4 µM for p38α-pCys180 and 80 nM for active p38α-WT.



(a) ATF2 after treatment with p38α-pCys180



(b) ATF2 after treatment with active p38α-WT

Fig. 3.8 Deconvoluted MS showing similar extents of substrate phosphorylation when ATF2 was treated with **(a)** p38α-pCys180 and **(b)** active p38α-WT, even though the p38α-WT experiment had lower p38α variant concentration and was taken after a shorter incubation time (4 µM and 1 h for **(a)**, compared to 80 nM and 30 min for **(b)**). **(b)** reproduced from Fig. 3.5b for convenient comparison.

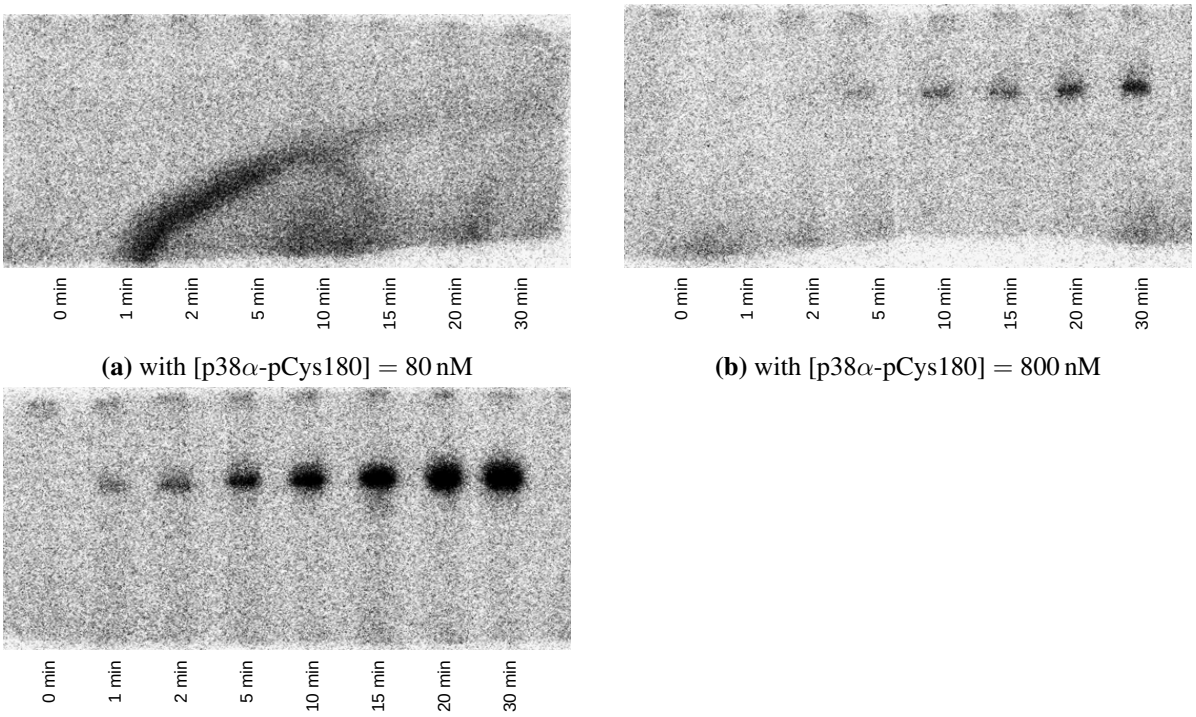
Michaelis-Menten kinetics is the most commonly used method in evaluating enzyme kinetics,[105] a point illustrated by the various estimates of Michaelis-Menten kinetic constants for p38α against ATF2 in the literature.[33,106,107] To mathematically simplify this model, the steady-state approximation is made. For this approximation to be valid, the substrate (ATF2) needs to be in vast excess relative to the enzyme (p38α kinase).[105,108] With the highest practically obtainable concentrations of ATF2 being in the µM range (maximum of 14 µM), the concentration of p38α would have to be in the low nM range (80 nM was chosen) in order to satisfy the condition of excess.

The enzyme and substrate concentration requirements had implications for both the assay detection method used, and the assay time required. In order for accurate measurements to be made, the assay signal must be well resolved and sufficient readings must be made over a range of signal strength, i.e. the noise-signal ratio must be low. For the example of substrate phosphorylation, this can be overcome in two ways:

1. The assay detection method would have to be sensitive enough to differentiate small variations in substrate phosphorylation, so that the endpoint signal can still be small. This scenario requires a sensitive detection method, or

2. A wide extent of substrate phosphorylation is used so that the sensitivity of the assay method is less important. The endpoint signal is also much larger in this case. This scenario lengthens the assay time.

Particular attention was paid to measurements performed with p38α-pCys180, since it was the slower of the two p38α active forms. ERA was investigated as to whether it would be a sufficiently sensitive detection method to give an acceptable signal-noise ratio. However, testing over a 30 min timecourse experiment showed that no discernible signals could be detected, when the p38α concentration used was that required of the kinetic assay (Fig. 3.9). Thus, if ERA were used as the assay method, despite being the most sensitive method, timepoints would still have to be collected with intervals of 30 min–1 h. Given the safety hazards involved with using radioactive materials, such an extended timecourse experiment using ERA would pose unnecessary risks. MS detection was therefore chosen as the detection method not only for the richer data it would provide, but also due to practical considerations.



(a) with [p38α-pCys180] = 80 nM

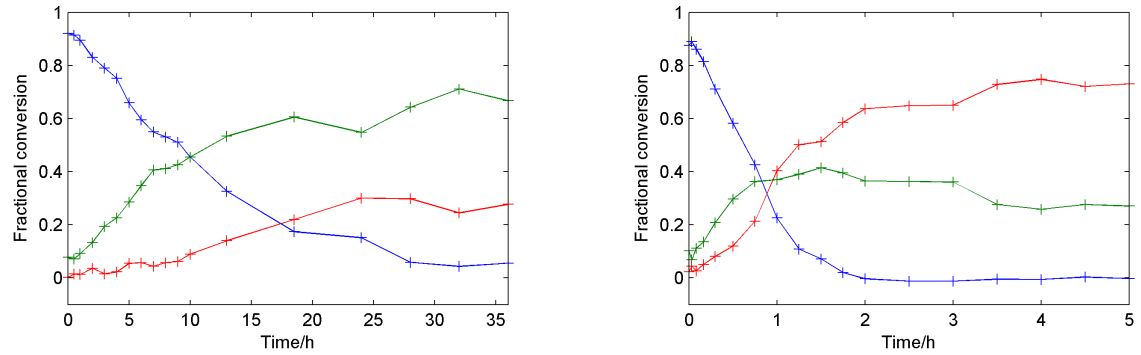
(b) with [p38α-pCys180] = 800 nM

(c) with [p38α-pCys180] = 4 µM

Fig. 3.9 Gel autoradiographs of timecourse experiments over 30 min, where ATF2 was treated with p38α-pCys180 at different concentrations. **(a)** No phosphorylation was observed over this time, while **(b)** a signal was visible after $\sim$15min. **(c)** Signals for all timepoints (except $t = 0$) were easily distinguished.

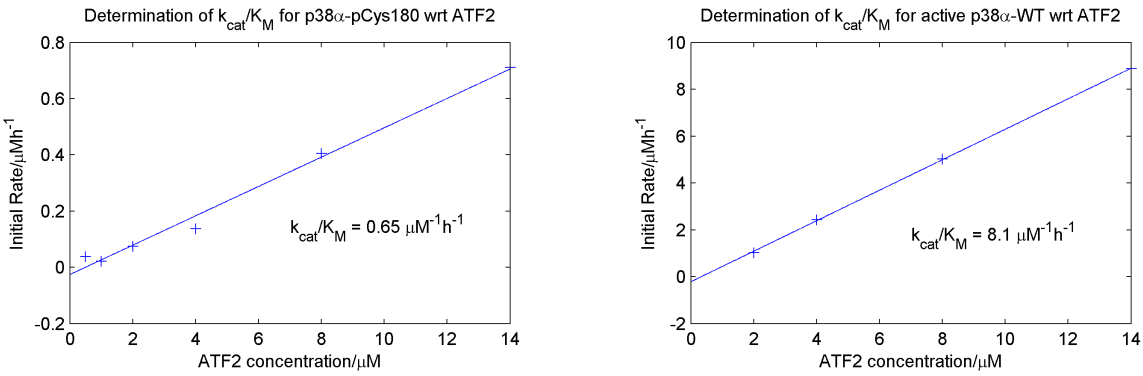
3.5.2 Analysis of Results from Kinetic Assay

To determine the kinetic parameters, initial rates were determined by running timecourses (Fig. 3.10) over a range of substrate concentrations. Plotting initial rate against initial substrate concentration showed that the practical limitation on ATF2 concentration had not allowed for substrate saturation to be reached, and that data had only been collected in the low substrate regime (Fig. 3.11). Thus, only the specificity constant ($k_{cat}/\mathrm{K_M}$) could be determined, and not the values of the component constants separately.



(a) with p38α-pCys180 **(b)** with active p38α-WT

Fig. 3.10 Representative timecourses from the kinetic assay ([ATF2] = 14 µM). The three phosphorylation states of ATF2 can be discerned separately: ATF2 (blue), ATF2-P (green), ATF2-PP (red).



(a) with p38α-pCys180 **(b)** with p38α-WT

Fig. 3.11 Plots of initial rate against initial substrate concentration from kinetic assay. As the plots did not reach horizontal asymptotes, substrate saturation was not reached.

With MS detection providing data not just for the consumption of ATF2 but also on the production of ATF2-P and ATF2-PP, the data were analysed to give specificity constants for all three levels of phosphorylation (Table 3.1). Comparison of global $k_{cat}/\mathrm{K_M}$ (the consumption of unphosphorylated ATF2) showed that active p38α-WT was 12.5$\times$ more active than p38α-pCys180. The values obtained were in good agreement with those quoted in a previous study where biologically derived p38α-pThrpTyr and

p38α-pThr180 had been investigated.[33] The order of magnitude of the values estimated in this study were also in good agreement with the ones quoted. This demonstrates that the technique of kinase activation by chemical modification yields species that are similarly active to their wild-type counterparts, despite the concerns expressed in Section 3.1, that is to say that the synthetic species p38α-pCys180 behaves as a mimic of p38α-pThr180, as the design intended. This observation of mimicry in kinases, replacing phosphothreonine with phosphocysteine has been further corroborated in the literature.[48] The possibility that the lower activity of p38α-pCys180 with respect to active p38α-WT could be due to loss of the cysteine phosphate[16] under reducing conditions was discredited by control experiments, which showed that there was no loss of the cysteine phosphate upon exposure to the conditions of the kinetic assay (done in collaboration with Sébastien Galan).

Table 3.1 $k_{cat}/\mathrm{K_M}$ values, determined without taking the $+40$ Da adduct into account.

	p38α-pTpY	**p38α-pC180**
(k_{cat}/$\mathrm{K_M}$ global)/μM^{-1} h^{-1}	8.14 ± 0.08	0.65 ± 0.03
(k_{cat}/$\mathrm{K_M}$ ATF-P)/μM^{-1} h^{-1}	5.30 ± 0.33	0.59 ± 0.03
(k_{cat}/$\mathrm{K_M}$ ATF-PP)/μM^{-1} h^{-1}	2.84 ± 0.29	0.07 ± 0.01
ratio ATF-P/ATF-PP	1.9 ± 0.2	8.9 ± 0.2

Interestingly, the two active forms of p38α had different ratios between the specificity constants for ATF2-P and ATF2-PP, which means that they produce ATF2-P and ATF2-PP to different extents. p38α-pCys180 is relatively less efficient at producing ATF2-PP than producing ATF2-P, when compared to active p38α-WT. The mechanism of ATF2 conversion to ATF2-PP by active p38α-WT has been shown to occur *via* a distributive mechanism, whereby in order to achieve the bis-phosphorylation of ATF2, p38α dissociates from ATF2-P (either ATF2-pThr69 or ATF2-pThr71) after the first phosphorylation, before re-associating for the second.[107] In the reported study, one of the pieces of "strong evidence" for the distributive mechanism in active p38α-WT was that ATF2-P was in significant excess early on in the timecourse. This observation has also been made in the current study for both active forms of p38α (Fig. 3.10) and furthermore, the timecourse with p38α-pCys180 shows ATF2-P concentration being higher than ATF2-PP concentration over the whole timecourse. Thus, not only does active p38α-WT work according to a distributive mechanism, it is highly likely that p38α-pCys180 functions according to this mechanism as well.

The timecourses also show that although active p38α-WT is more efficient at producing ATF2-PP than p38α-pCys180, some ATF2-P still remains at the timecourse endpoint. This is demonstrated in the time-

courses as a plateau in the curves for both active forms of p38α. This could be the result of enzyme inactivation due to the extended time and elevated temperatures used in the assay. However, this is unlikely to be a major factor, particularly in the case of p38α-pCys180 which has already endured 10 h at 37 °C during the chemical modification stage, yet still shows activity after this process. Instead, a biological explanation is more likely.

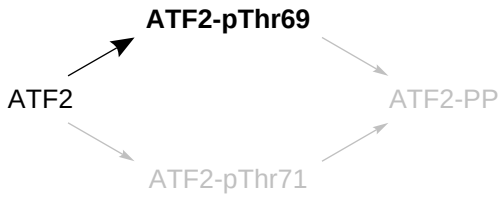
ATF2-P encompasses two distinct species, ATF2-pThr69 and ATF2-pThr71, both of which have very different behaviours in forming ATF2-PP when phosphorylated by p38α. It has been shown that for active p38α-WT, while this active form doesn't have a bias towards forming ATF2-pThr69 or ATF2-pThr71 from ATF2, ATF2-pThr69 is converted 40 times slower to ATF2-PP than ATF2-pThr71 is converted.[107] This therefore leads to a build-up of ATF2-pThr69 at the end of the reaction, which in this case, manifests in the plateau in the timecourse of active p38α-WT. A similar phenomenon must also be evident in the case of p38α-pCys180, to result in a similar plateau.

To determine the mechanistic action of p38α-pCys180 in more detail, the product profile after p38α-pCys180 treatment was determined in more detail. MS/MS analysis was used to do this. Analysis (in collaboration with Ritu Raj) of a sample taken near the p38α-pCys180 timecourse endpoint found that the ATF2-P was predominantly ATF2-pThr69, with no evidence of ATF2-pThr71 being present (Section 3.8.7). This suggested that like active p38α-WT, the plateau in the p38α-pCys180 timecourse is due to a lack in efficacy at converting ATF2-pThr69 to ATF2-PP, compared to the equivalent reaction with ATF2-pThr71.

Active p38α-WT and p38α-pCys180 therefore follow very similar mechanisms of action, since they both sequester ATF2-pThr69 towards the end of the reaction. However, they do so to different extents. This difference is what causes there to also be a difference in the specificity constant ratios between the two active forms, as mentioned previously (Table 3.1). For the ratios to manifest themselves as they are, when compared to the corresponding relative rates for active p38α-WT, either:

1. the conversion of ATF2 to ATF2-pThr69 is relatively faster than the conversion of ATF2 to ATF2-pThr71 for p38α-pCys180
2. the conversion of ATF2-pThr69 to ATF2-PP is relatively slower than the conversion of ATF2-pThr71 to ATF2-PP for p38α-pCys180

Putting the comparisons into context, the literature shows that active p38α-WT produces ATF2-P with the ATF2-pThr69:ATF2-pThr71 ratio being very close to unity and to reiterate, active p38α-WT converts ATF2-pThr71 to ATF2-PP 40 times faster than it converts ATF2-pThr69 to ATF2-PP.[107] p38α-pCys180 on the other hand has different values for these ratios, where both of these ratios would be higher than for active p38α-WT. A fitting of the timecourse to the proposed model of the distributive mechanism (Scheme 3.2) to determine these ratios was not attempted for p38α-pCys180, however, it is likely that the propensity of p38α-pCys180 to produce ATF2-P more readily than active p38α-WT comes mainly from the former aforementioned condition (that the conversion of ATF2 to ATF2-pThr69 is relatively faster than the conversion of ATF2 to ATF2-pThr71), since a small change in this ratio will have a large, proportional change in this parameter, and therefore dominate the enzyme's behaviour (i.e. adding 1 to 1 has a greater effect than adding 1 to 40).



Scheme 3.2 Schematic representation of the distributive mechanism for the phosphorylation of ATF2 by p38α. The arrows are drawn to represent the reactions as catalysed by p38α-pCys180, with the reactions that are slower having small grey arrows. These difference in rates result in the sequesterisation of ATF2-pThr69. Only the forward reactions are considered.

Closer examination of the original mass spectra from the timecourses showed that there was an additional satellite peak associated with the peak of each ATF2 phosphorylation state, in all spectra (Fig. 3.12). This peak was always at $\sim$+40 Da and was variable in porportion between spectra. It was found that MS corresponding to samples which had been left to stand for longer at room temperature before analysis had stronger satellite peaks, leading to the hypothesis that these satellites were due to cyanate adduct formation with the urea quenching solution.[109,110] Re-analysis of the data, taking these adducts into account by adding their intensities to the main peak, resulted in the same comparative conclusions being drawn from the re-calculated specificity constants (Table 3.2), thus suggesting that the hypothesis of adduct formation was justified, and did not affect the conclusions already drawn.

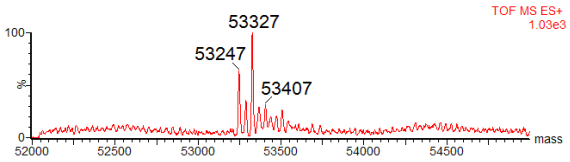


Fig. 3.12 Representative deconvoluted MS showing peaks corresponding to ATF (53 247 Da), ATF-P (53 327 Da) and ATF2-PP (53 407 Da). Associated with each peak is a satellite peak at $\sim +40$ Da, which corresponds to cyanate adduct formation with the urea quenching solution.[109,110]

Table 3.2 $k_{cat}/\mathrm{K_M}$ values, as determined when taking the $+40$ Da adduct into account.

	p38α-pTpY	**p38α-pC180**
($k_{cat}/\mathrm{K_M}$ global)/$\mu\mathrm{M}^{-1}$ h^{-1}	7.43 ± 0.21	0.60 ± 0.03
($k_{cat}/\mathrm{K_M}$ ATF-P)/$\mu\mathrm{M}^{-1}$ h^{-1}	5.13 ± 0.32	0.51 ± 0.02
($k_{cat}/\mathrm{K_M}$ ATF-PP)/$\mu\mathrm{M}^{-1}$ h^{-1}	2.30 ± 0.24	0.09 ± 0.02

3.6 Effect of Inhibitors on Active Forms of p38α

As an additional means of probing potential differences in the structure and function of the two p38α active forms, their behaviour in the presence of inhibitors was investigated. The purpose of this study was to explore the hypothesis that the effects of a partially active form of p38α on inhibitor action could be revealing as to the kinase's structure and function. Since kinases have two main conformations, two major strategies of inhibition have arisen, with each strategy targeting one of these conformations. Thus, Type I inhibitors bias the binding of the active or DFG-in conformation, while Type II inhibitors preferentially bind the inactive or DFG-out conformation.[6] A range of p38α-specific inhibitors were thus tested (Scheme 3.3), some being Type I (VX745,[111] TAK715,[112] SB202190,[106,113] and **20**[114]), some Type II (BIRB796[63] and **21**[114]) and one that was unclassified (JX401[115]).

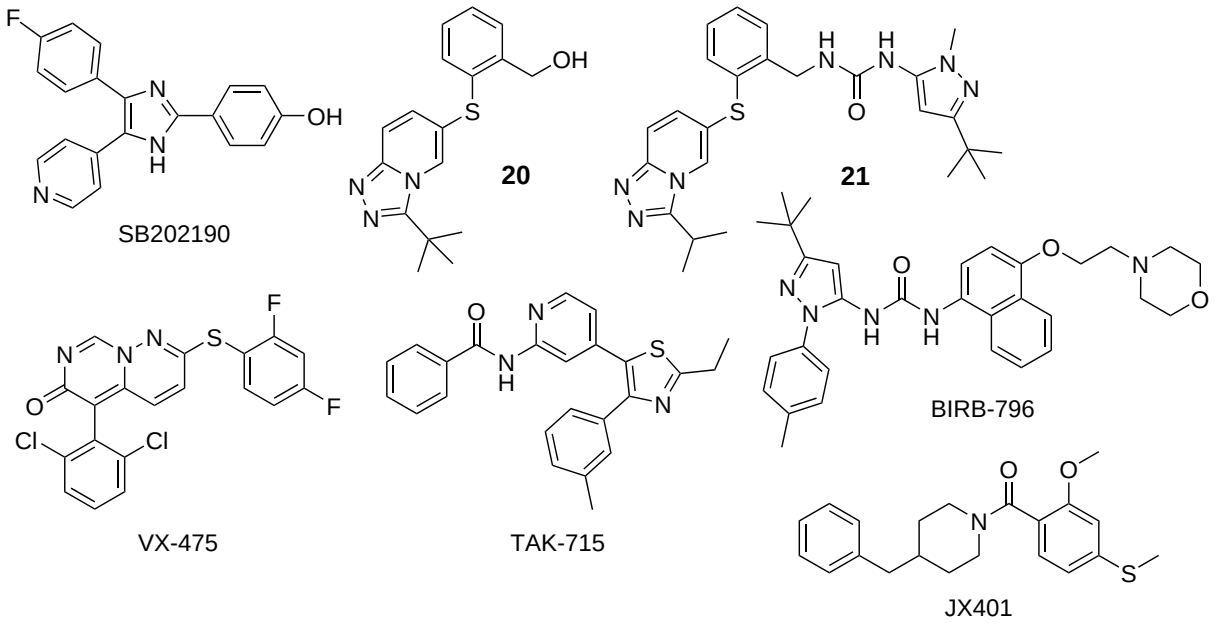
3.6.1 Inhibitor Assay Development

Before the inhibitors could be tested, a suitable assay needed to be developed, particularly with regards to assay signal detection. Ideally, the assay method chosen would be easily scalable for the large number of reaction conditions to be tested. For any given inhibitor tested, there would be $\sim$20 conditions to test:

$$(\sim 10 \text{ different inhibitor concentrations}) \times (2 \text{ active forms of p38}\alpha)$$

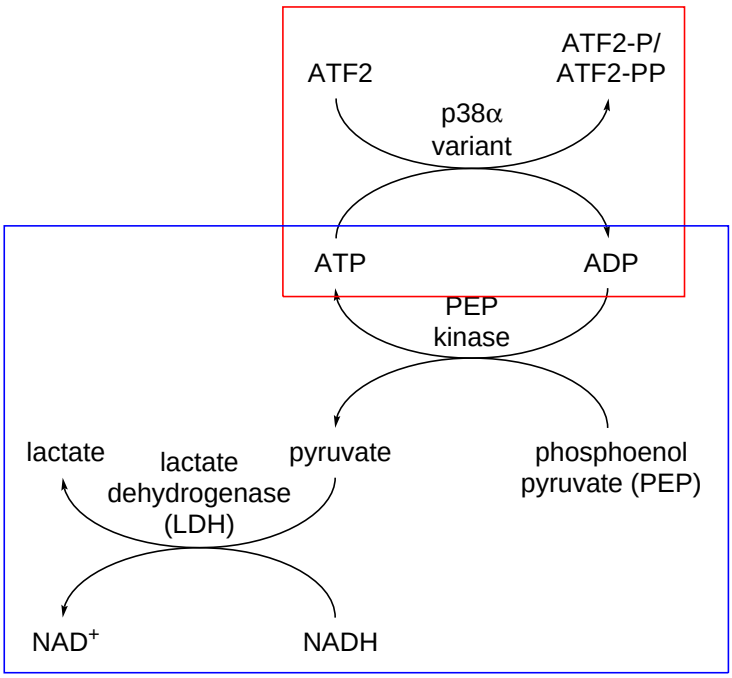
This estimate does not include the numerous replicates that would also be needed.

The assay detection methods employed so far were not easily scalable, having instead been chosen because they are able to make direct measurements of substrate species (Section 3.2). The main reason for the lack



Scheme 3.3 All the p38α-specific inhibitors tested in this study.

of scalability is because they are both stopped assays. Since measurement of species quantities in samples cannot be taken while the enzymatic reaction is still running, the reaction must be quenched when a sample is taken. This is the case for both the assay under MS detection, and in the ERA. Quenching was therefore done using urea or SDS/EDTA respectively. A large number of conditions therefore means a large number of samples to quench, which would be inconvenient and therefore not particularly scalable.



Scheme 3.4 Reaction pathway for the NADH-linked assay with the kinase reaction (red box) and NADH reporting reactions (blue box) highlighted.

Instead, a NADH-linked assay was attempted (Scheme 3.4, in collaboration with Sébastien Galan). In this assay regime, the ADP liberated on substrate phosphorylation is converted back to ATP by phosphoenol pyruvate kinase, producing pyruvate. This subsequently drives a cascade of enzymatic reactions to convert NADH to NAD^+. The disappearance of NADH, which can be monitored continuously at 340 nm in the UV, is then detected. Trials of this assay, using the same protein concentrations as for experiments described with the WT controls (Section 3.4), did not detect any signal above background. The low initial substrate (ATF2) concentration was suspected to have caused this. From a rough calculation of the maximum absorbance difference that could be achieved, i.e. assuming all the ATF2 was converted to ATF2-PP:

$$A = \epsilon c l$$

$$A = (6220\,\mathrm{M^{-1}cm^{-1}}) \times (6 \times 10^{-6}\,\mathrm{M}) \times (0.5\,\mathrm{cm})$$

$$A = 0.019 \text{ to } 2\,\mathrm{s.f.}$$

Considering that the change in UV absorbance was likely to be much less than 0.019, the spectrophotometer would not have sufficient resolution to detect even this maximum absorbance difference accurately. Further, as the concentration of substrate could not be practically increased to produce more NAD^+, the use of the NADH-linked assay for investigating p38α inhibition was abandoned. Instead, the already established ERA was used, since it was a more scalable assay than MS detection.

3.6.2 Inhibitor Curves

Inhibition curves were determined for the inhibitors against both active forms of p38α (p38α-pCys180 and active p38α-WT) (Fig. 3.13). For any given inhibitor, comparison of curves from the two active species showed that any deviation in the profile of the curves was insignificant. This was confirmed with IC_{50} values that were very similar between the active forms for each inhibitor (Table 3.3). The same observation was made for the Hill coefficients, an indicator of the steepness of the sigmoidal curve (Table 3.4), the consistent determination of which had been improved by increasing the resolution of the inhibitor dilution series used from a 5$\times$ dilution series to a 2$\times$ dilution series. Thus, although the activities of the two active forms were different, the effect of inhibitors was the same.

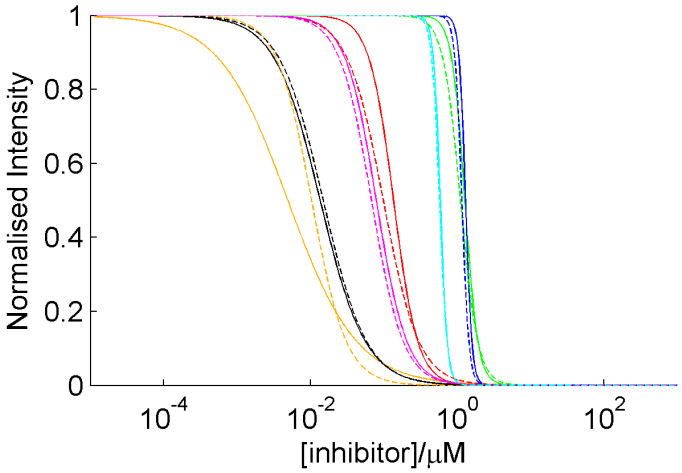


Fig. 3.13 Normalised IC_{50} curves for all the inhibitors investigated (Key: VX745 (red), TAK715 (green), BIRB796 (blue), JX401 (orange), SB202190 (black), **20** (magenta), **21** (cyan)) in the study for inhibition against both p38α-pCys180 (dashed lines) and active p38α-WT (solid lines). The general correspondence of the curves of the same inhibitor, between the two active forms of p38α was noted.

Table 3.3 IC_{50} values for all the inhibitors of the study, against both active forms of p38α.

Inhibitor	**p38α-pC180/nM**	**p38α-pThrpTyr/nM**
VX745	99 ± 8	136 ± 5
TAK715	1091 ± 64	1264 ± 24
BIRB796	1128 ± 68	1309 ± 40
JX401	11 ± 2	5 ± 2
SB202190	14 ± 1	13 ± 1
20	68 ± 6	78 ± 2
21	558 ± 24	586 ± 9

Table 3.4 Hill coefficient values for all the inhibitors of the study, against both active forms of p38α.

Inhibitor	**p38α-pCys180**	**p38α-pThrpTyr**
VX745	1.8 ± 0.1	2.6 ± 0.1
TAK715	3.8 ± 0.7	4.8 ± 0.4
BIRB796	14.0 ± 5.6	9.9 ± 1.1
JX401	2.0 ± 0.1	0.9 ± 0.04
SB202190	1.5 ± 0.3	1.4 ± 0.1
20	2.0 ± 0.1	2.1 ± 0.2
21	12.6 ± 6.2	9.3 ± 0.5

However, the range of Hill coefficients obtained required further attention. It was noted that the Type II inhibitors were associated with high Hill coefficients (8–15), TAK715, which has been shown to bind to both DFG-in and -out conformations in structural studies[116] displayed Hill coefficients of intermediate value, while all the known Type I inhibitors had lower Hill coefficients (1–2). This association suggested that the Hill coefficient could be used as a marker for Type II inhibitor behaviour, under which JX401 would most likely be a Type I inhibitor.

Attempts were made to rationalise the association of high Hill coefficient and Type II behaviour by investigating the possible causes. One of these potential causes is that the p38α species could form a dimer,[73] the aggregation of which could result in high Hill coefficients.[117] This hypothesis was unlikely for two reasons. Firstly, dimer formation in isolated p38α is favoured upon cysteine oxidation,[73] which was prevented either by the reducing conditions of the storage buffer used, or by the absence of oxidisable native cysteines[118] in the case of p38α-pCys180. Secondly, oligomers had been observed in the purification process, some of them presumably being dimers, but these had been separated from the monomers.

The purification process for p38α variants had involved a multistep procedure. After the initial affinity chromatography stages of GST affinity chromatography, GST-tag cleavage and GST-tag rebinding, variants under purification were subject to anion exchange. This separated the majority of the oligomers from the monomer, giving rise to two peaks in the FPLC UV trace (Fig. 3.14a), and subsequently two protein batches which contained different -meric forms of the target p38α variant (p38α-Cys180 in the case of the Figure displayed). The batch containing the monomers was used in the studies described. Confirmation that the two batches were indeed different -meric forms came from intact protein MS, where both batches had the expected molecular mass for the monomer, and from gel filtration, where the two batches were treated separately and had major peaks at different retention times (Fig. 3.14b and 3.14c). The major peak from the anion exchange chromatography gave rise to a peak with longer retention time in gel filtration, showing that this peak corresponded to a lighter-meric form, namely the monomer. The separation of the different -meric forms of the p38α variants demonstrates that under the experimental timescales and conditions, the monomer is stable against multimer formation, and thus likely to have remained a monomer in subsequent experiments.

It should however be noted that dimer formation in p38α has been proposed as part of the kinase's regulatory process. This observation is for a different active form of p38α though, where the reported mutant[119]

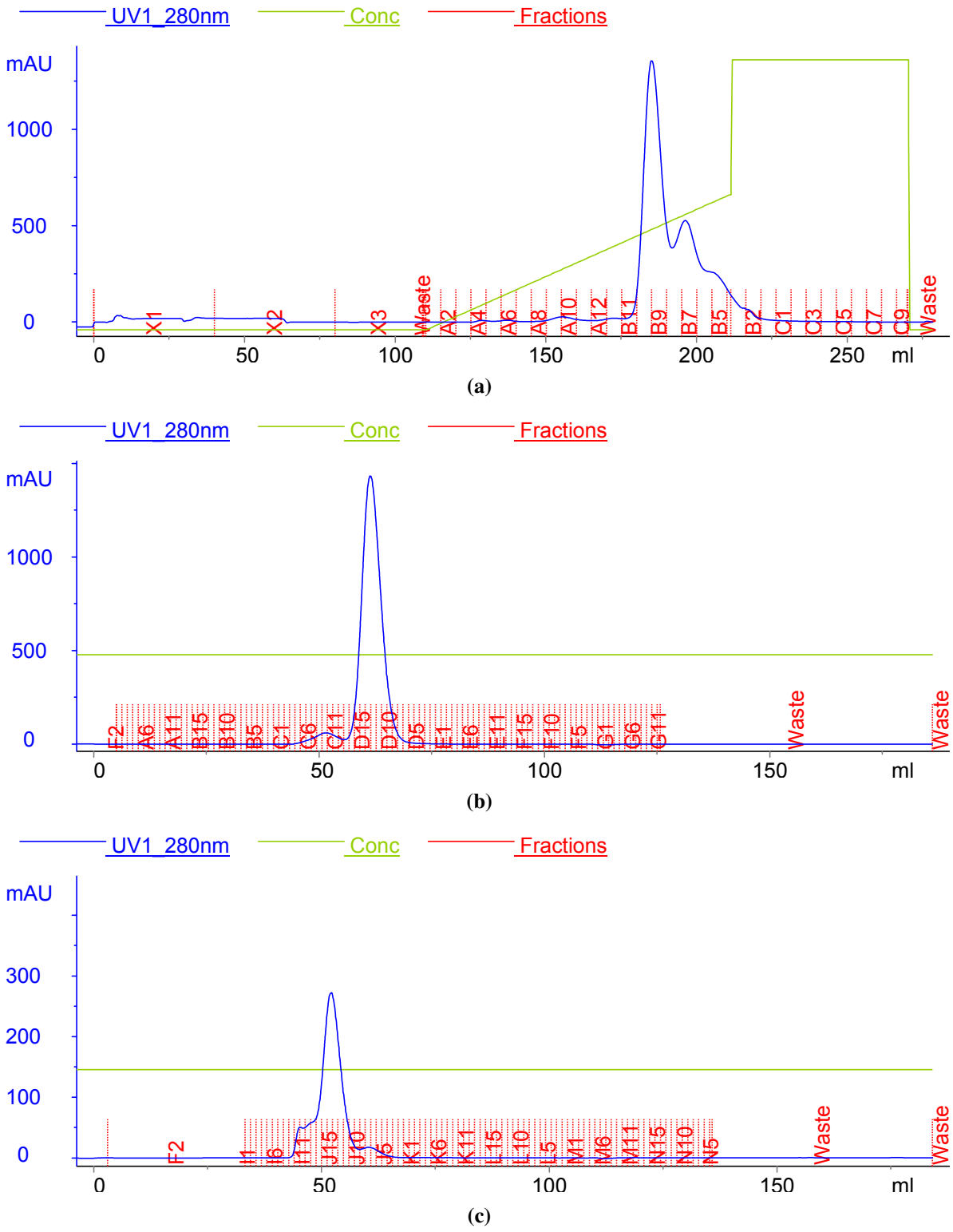


Fig. 3.14 FPLC UV traces from the purification of p38α-Cys180. **(a)** Anion exchange was used as an intermediate step in the purification. The two main peaks correspond to the monomer (around Fraction B9) and multimers (around Fraction B7) of p38α-Cys180. **(b)** The major peak from anion exchange gave a product with longer retention time in gel filtration than **(c)** the minor peak, thus demonstrating that the major peak was of a lighter -meric form, i.e. the monomer.

behaves as p38α-pThr323.[51] This active form is able to autophosphorylate at Thr180 and forms homodimers in order to do so, but it is also able to form heterodimers with unphosphorylated p38α. A full discussion of this alternative regulatory pathway of p38α is outside the scope of this study, but it bears little implication on this study since there are currently no known reports of the dimer remaining intact once phosphorylated at Thr180.

Regardless, to further reduce the likelihood of dimer formation, additional detergent was added to the buffer, ensuring that the detergent concentration was beyond the critical micelle concentration (CMC). This distinction of CMC is important since although detergent had been included in the study of the range of inhibitors, it was diluted to such a concentration that it would not have contributed to preventing aggregation. Comparison of inhibition curves collected with or without additional buffer detergent showed that there was no significant difference in the Hill coefficient between datasets collected with or without the extra detergent (data not shown).

Another hypothesis for high Hill coefficient was that the slow binding kinetics of BIRB796[63,120] and **21**[121] to the p38α variant was the cause. For this reason, the closely analogous inhibitors **20** and **21** had been used in the aggregation detergent study, since the variable of inhibitor binding position would have been minimised. In this hypothesis, it would follow that since **20** has faster binding kinetics than **21**,[121] it also has a lower Hill coefficient. Evidence for this hypothesis comes in the stoichiometric model:[117]

$$\frac{[\mathrm{I}]}{K_\mathrm{d}} = \frac{\text{inh }\%}{100 - \text{inh }\%} + \frac{[(\text{inh }\%)/100][\mathrm{E}]}{K_\mathrm{d}} \tag{3.2}$$

From Equation 3.2, it is found that when K_d is small, as in the case of BIRB796 and inhibitor **21**, the enzyme concentration in the second term in this equation becomes significant and the curve starts to exhibit a steeper behaviour.[117] Another analogous pair of inhibitors to inhibitors **20** and **21** was required,[122] to test the hypothesis, with the extra requirement that both these inhibitors had faster binding kinetics. However, due to the difficulty in obtaining samples of such analogous, propriety inhibitors, the study on these inhibitors could not be undertaken. Thus, any causal effect, resulting in an association between Type II inhibitors and high Hill coefficients could not be established.

3.7 Summary

After establishing robust methods for synthesising chemical variants of p38α, the chemical derivatives from p38α-Cys172 and p38α-Cys180, and the parent species themselves, were all investigated for their enzymatic activity. Only p38α-pCys180 was found to be active, which was confirmed by detecting the phosphorylated species of ATF2 using MS and radiolabel detection. The activity of p38α-pCys180 demonstrated that phosphorylation of p38α at the native phosphorylation site at 180 was sufficient for conformational change to an active state. Conversely, phosphorylation at a site not natively phosphorylated, position 172, did not confer activity, showing that not only is the event of phosphorylation important, the phosphorylation must also be in the correct position and not just simply in the activation loop.

The activity of p38α-pCys180 was compared in more detail to that of enzymatically activated p38α-WT. As expected, this second active form of p38α was more active than p38α-pCys180 due to its predominantly bis-phosphorylated state. By comparison to literature reported values, the activity of p38α-pCys180 is consistent with the mono-phosphorylated p38α-pThr180,[33] which is the naturally occurring species that p38α-pCys180 is a mimic of. Given this similarity in the enzyme kinetics, p38α-pCys180 is therefore a sufficient mimic of p38α-pThr180. It can further be claimed that phosphocysteine is a sufficient mimic of phosphothreonine in kinases more generally, since activation by chemical modification has also been demonstrated in other kinases requiring phosphorylation at threonine for activation.[48]

The development of a robust and practical MS detection method for the kinetic assay was instrumental in the collection of the large amounts of data required to investigate the mechanistic workings of the two active forms of p38α. Since ATF2 has two phosphorylation sites that are phosphorylated by p38α, this assay technique allowed the first and second phosphorylation events to be discerned independently. It was thus observed that p38α-pCys180 is not only a slower enzyme than active p38α-WT, its product profile is also different, since it has a higher tendency to retain the ATF2 at the mono-phosphorylated state.

Activity against inhibitors was used as a further discriminator for the extent of biological recognition capabilities of p38α-pCys180. Inhibitors behaved similarity against both p38α-pCys180 and active p38α-WT, suggesting that the binding modes of the inhibitors, and the ATP binding sites of the p38α active forms were similar between the two forms. At the same time, a chance observation was made that the steepness

of inhibition curves was not always the same, with Type II inhibitors generally having higher Hill coefficients than the Type I inhibitors. The exact cause of this phenomenon however could not be determined.

Having demonstrated the mimicry of phosphocysteine for phosphothreonine, the further scope of this method will be explored in the next chapter, when attention is turned to whether phosphocysteine can also be used as a mimic of phosphoserine in kinases.

3.8 Experimental

3.8.1 General Measures

Enzymatic Reactions and Activity Assay

ATF2 (Sigma Aldrich), ATP (Sigma Aldrich), active and inactive p38α-WT (Merck Millipore) and [γ -^{32}P]-ATP (Perkin Elmer) were all used as per manufacturer's instructions. Inhibitors VX745 (Key Organics), TAK715 (Tocris), BIRB796 (Selleckchem), JX401 (Tocris) and SB202190 (Sigma Aldrich) were made into DMSO stock solutions (2 mM) on arrival and stored at −20 °C. Samples of the remaining 2 inhibitors (**20** and **21**) were kindly sent by Dr. Lyn Jones and co-workers, Pfizer, Cambridge, MA, USA and similarly made into stock solutions. All other chemicals and buffer components were bought from common suppliers and used as purchased. LC-MS analysis was done on a Micromass LCT Premier instrument using a Chromolith® FastGradient RP-18 endcapped 50×2 mm monolithic HPLC column (Merck). A linear gradient was run from 5–95% of solvent mixture B (1% formic acid in acetonitrile) into solvent mixture A (1% formic acid in water). The gradient was run with a flow rate of 0.3 mL/min over 6 min for the Chromolith column. All the other general measures taken for whole protein MS were as previously described (Section 2.11.1, "Protein Modification Reactions" and "Quantification of MS Data").

3.8.2 Measurement of the Activity of All Chemical Variants of p38α-X172 and -X180 by MS Detection

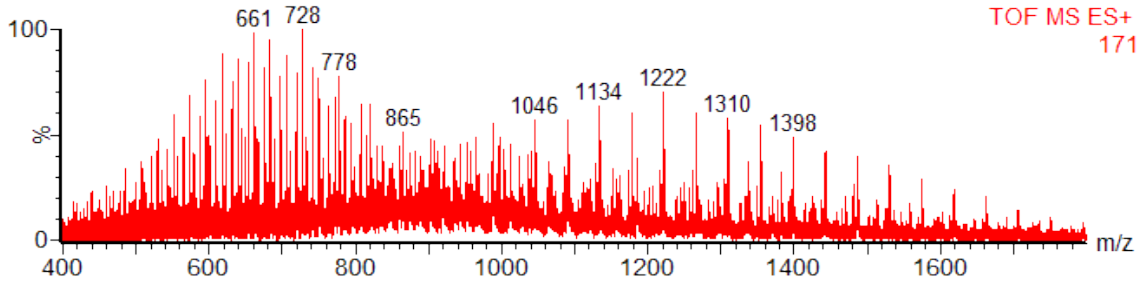
Kinase activity buffer stock (5.00 µL of a 10× stock containing 200 mM Tris pH 7.5, 20 mM DTT, 50 mM β-glycerophosphate, 2.0 mM Na_3VO_4, 100 mM $MgCl_2$, 50 mM NaF) was diluted with water (13.4 µL). A sample of p38α kinase chemical variant (10.0 µL of a 20 µM solution) in p38α reaction buffer (50 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol) was added to the buffer, followed by ATP solution (5.00 µL

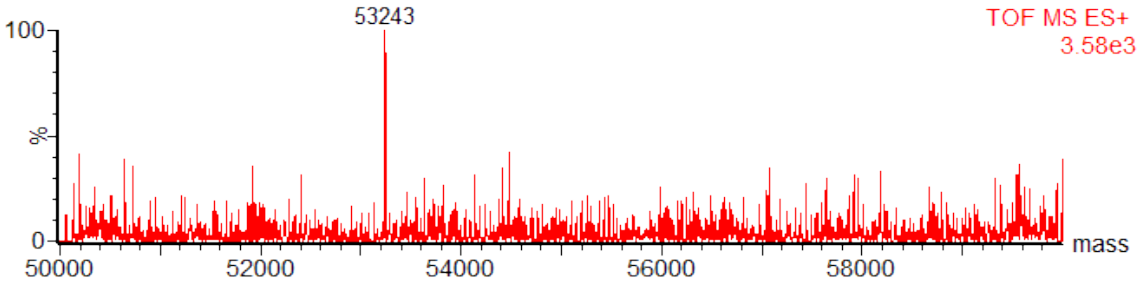
of a 5.0 mM solution in water, pH 7.0). A sample (3.34 µL) corresponding to t = 0 was taken before the substrate ATF2 (7.16 µL of a 18.9 µM solution) in ATF2 storage buffer (20 mM HEPES pH 7.5, 0.5 mM EGTA, 1 mM DTT, 5 µg/mL aprotinin, 10 mg/mL leupeptin, 0.25 mg/mL AEBSF, 0.03% Brij-35, 150 mM NaCl, trehalose 10%) pre-diluted with water (7.75 µL) was added. Samples (5 µL) were taken at given timepoints (1, 5, 10, 30 and 60 min, then 3 h and 16–18 h) and the samples quenched with urea (10 µL of a 10 M solution in water). The samples were then analysed by LC-MS for the mass of the substrate. For all the reactions with p38α-X180 species, samples taken after 150 min were also analysed by SDS-PAGE, which corroborated that the ATF2 in the assay reaction mixtures remained intact. MS displayed are for the timepoint >16 h: (Unmodified substrate: expected mass: 53 kDa (value provided from manufacturer's product specification sheet), observed range of masses: 53 242–53 248, Monophosphorylated substrate: expected mass: +80 from unmodified substrate, observed range of masses: 53 326–53 327, Diphosphorylated substrate: expected mass: +160 from unmodified substrate, observed mass: 53 405).

The final concentrations of all the reaction components:

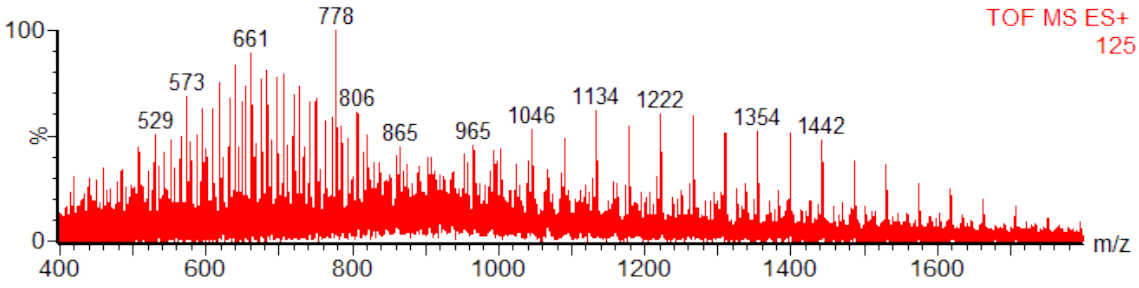
Component	Final Concentration
p38α kinase	4.0 µM
ATP	500 µM
kinase assay buffer:	
Tris	20 mM
DTT	2 mM
β-glycerophosphate	5 mM
Na_3VO_4	0.2 mM
$MgCl_2$	10 mM
NaF	5 mM
ATF2	3.0 µM

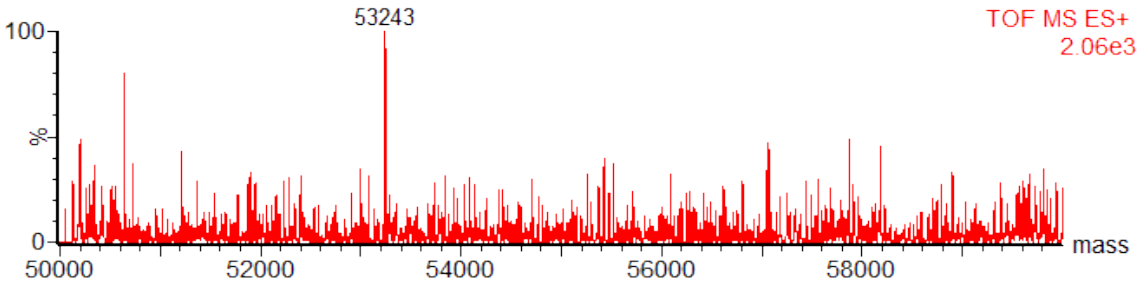
ATF2 after Treatment with p38α-Cys172:



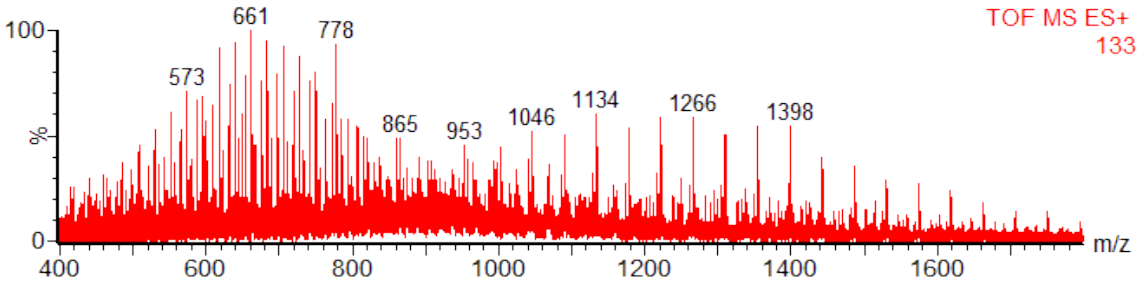


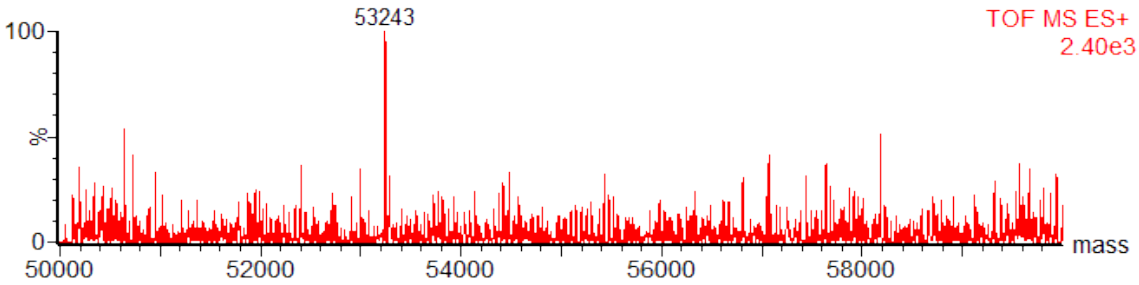
ATF2 after Treatment with p38α-Dha172:



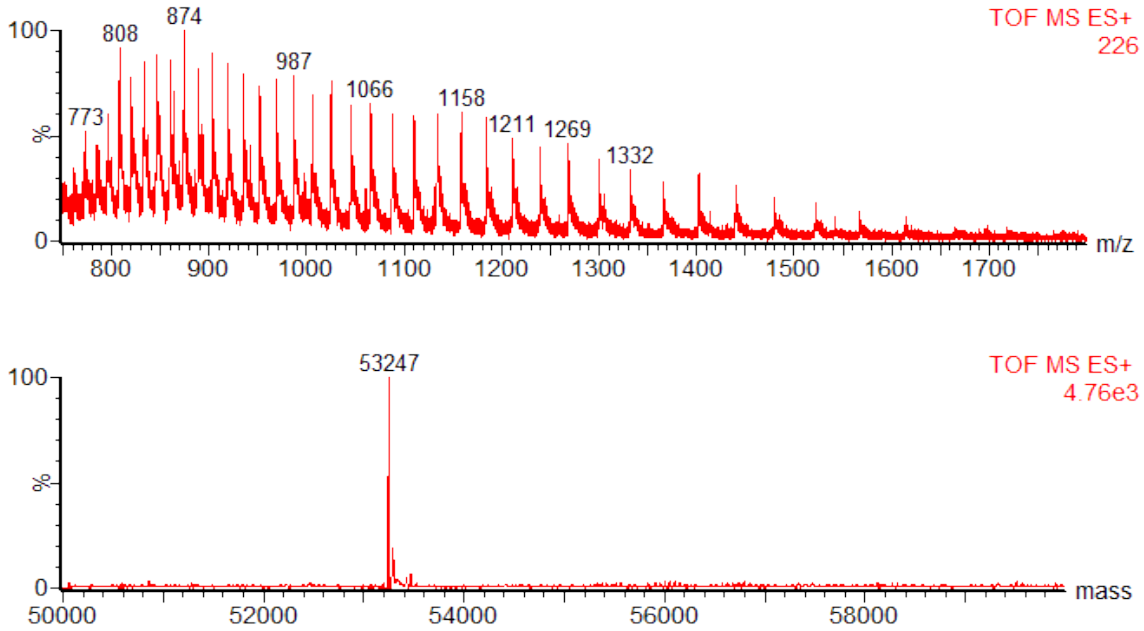


ATF2 after Treatment with p38α-pCys172:

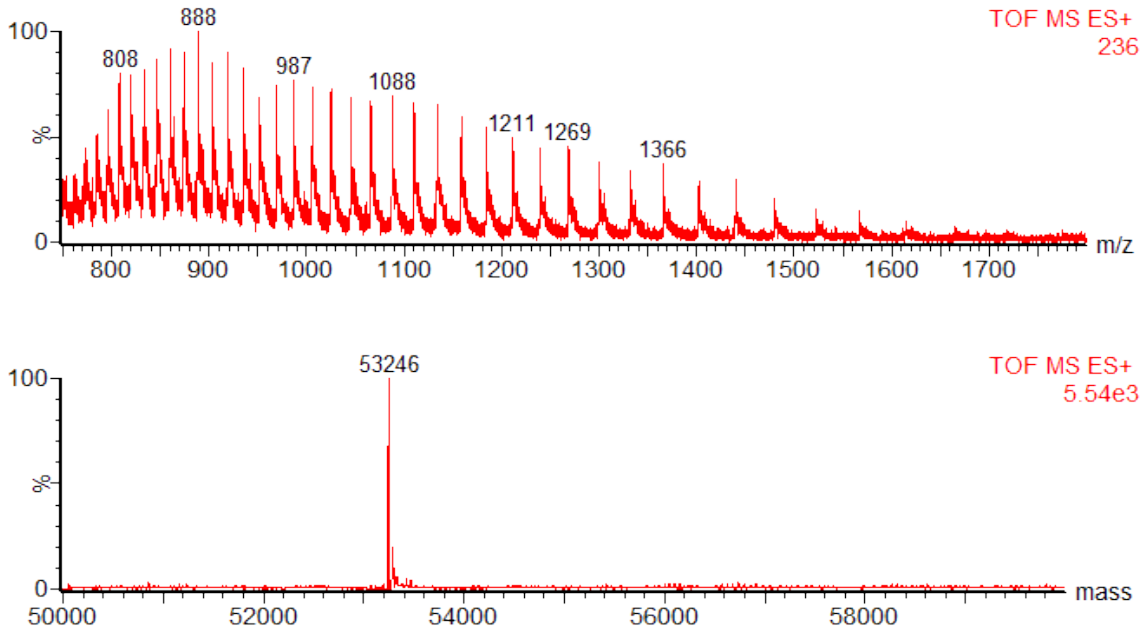




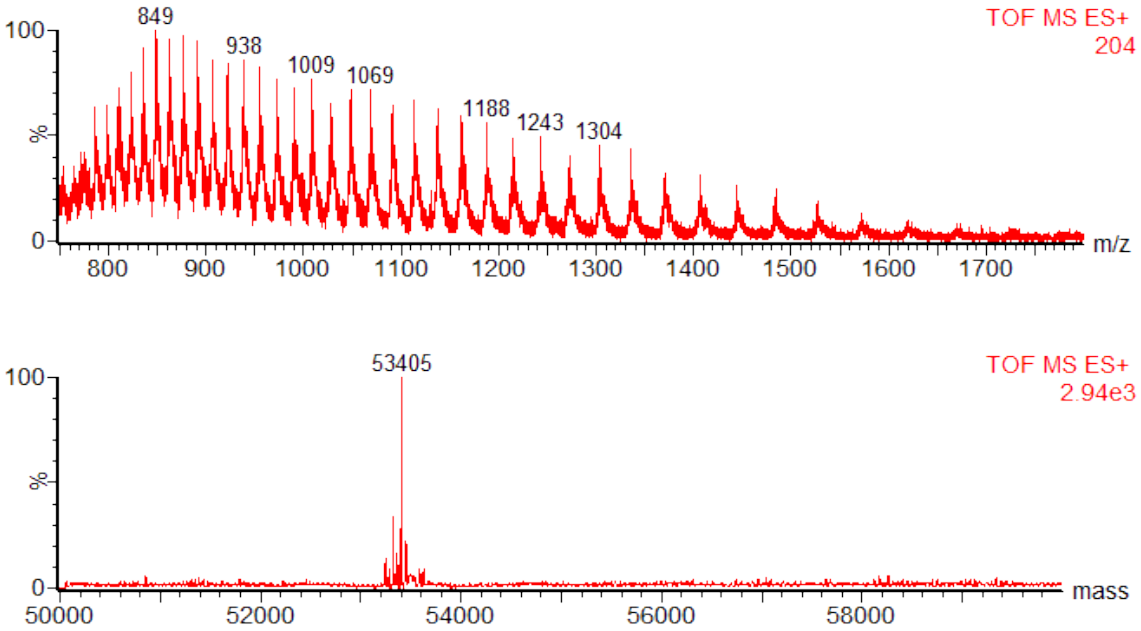
ATF2 after Treatment with p38α-Cys180:



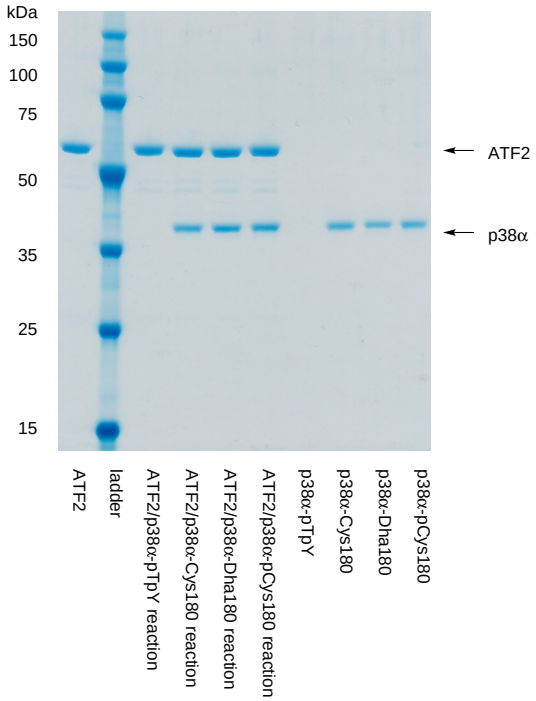
ATF2 after Treatment with p38α-Dha180:



ATF2 after Treatment with p38α-pCys180:



SDS-PAGE Analysis of Assay Reaction Samples with p38α-X180 Chemical Variants:



3.8.3 Measurement of the Activity of All Chemical Variants of p38α-X172 and -X180 by Electrophoretic Radioassay

Enzymatic Reaction

Kinase assay buffer stock (7.5 µL of a 10× stock containing 200 mM Tris pH 7.5, 20 mM DTT, 50 mM β-glycerophosphate, 2.0 mM Na_3VO_4, 100 mM $MgCl_2$, 50 mM NaF) was diluted with water (appropriate quantity for a final reaction volume of 75 µL) and ATP (7.5 µL of a 5 mM solution in water, pH 7.0) added. Kinase (modified or unmodified mutant, 15.0 µL of a 20 µM solution) in p38α reaction buffer and subsequently [γ-^{32}P]-ATP (8–13 µCi in 0.75–3 µL of solution) were added. A sample (3.34 µL) corresponding to t = 0 was taken before initiating the enzymatic reaction with the addition of ATF2 (12.06 µL of a 18.9 µM solution) in ATF2 storage buffer (20 mM HEPES pH 7.5, 0.5 mM EGTA, 1 mM DTT, 5 µg/mL aprotinin, 10 mg/mL leupeptin, 0.25 mg/mL AEBSF, 0.03% Brij-35, 150 mM NaCl, trehalose 10%), pre-diluted with water (11.13 µL). Samples of further timepoints (2, 5, 10, 18, 30, 45, 60, 75, 90, 105, 120 and 150 min) were taken and the samples (5 µL) quenched with 1.5× SDS/EDTA loading dye (10 µL) (18.75 mM Tris pH 6.8, 45 mM EDTA, 18.75% glycerol 3.75%, SDS, 1.5% β-mercaptoethanol, 0.15% bromophenol blue) for SDS-PAGE analysis. Similar reactions were run with active and inactive p38α-WT but with a different final kinase concentrations (80 nM and 800 nM respectively). The final concentrations of all the reaction components:

Component	Final Concentration
p38α kinase	4.0 µM
ATP	500 µM
[γ-^{32}P]-ATP	120–170 µCi/mL
kinase assay buffer:	
Tris	20 mM
DTT	2 mM
β-glycerophosphate	5 mM
Na_3VO_4	0.2 mM
$MgCl_2$	10 mM
NaF	5 mM
ATF2	3.0 µM

SDS-PAGE analysis

Gels were hand-cast using an appropriate casing block for the PROTEAN II xi Multi-Cell gel tank (Bio-rad). 12% gels were made from pouring the solution (for 6 gels):

Stock Solution	Quantity
30% acrylamide (37.5:1 acrylamide:bisacrylamide)	160 mL
1.0 M Tris pH 8.8	150 mL
water	82 mL
10% SDS	4 mL
10% APS	4 mL
TEMED	160 µL

The gels were topped with sat. butanol after pouring. Once the gels were set, 5% stacking gels were poured on top from the solution:

Stock Solution	Quantity
30% acrylamide	17 mL
1.0 M Tris pH 6.8	12.5 mL
water	68 mL
10% SDS	1 mL
10% APS	1 mL
TEMED	100 µL
bromophenol blue	

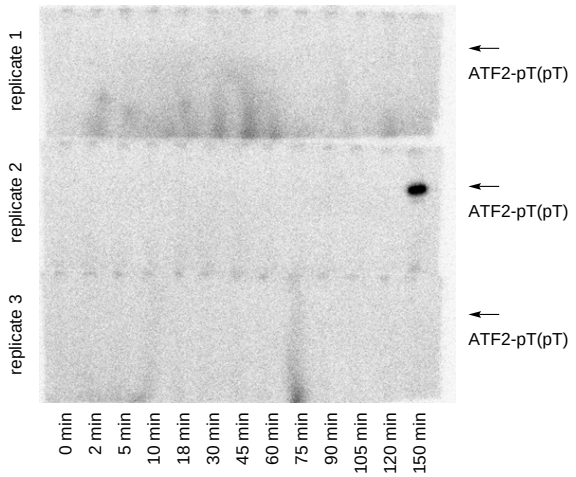
An appropriate comb was inserted after pouring the stacking gel. Samples (10 µL) were loaded in alternate lanes on the set gel with 1× blank SDS/EDTA loading dye (10 µL) (12.5 mM Tris pH 6.8, 30 mM EDTA, 12.5% glycerol 2.5% SDS, 0.1% bromophenol blue) in the otherwise unloaded lanes. The gels were run in Tris/glycine running buffer (25 mM Tris base, 25 mM glycine, 0.1% SDS) for 1 h at 600 V, 1000 mA, 250 W.

Gel Image Acquisition

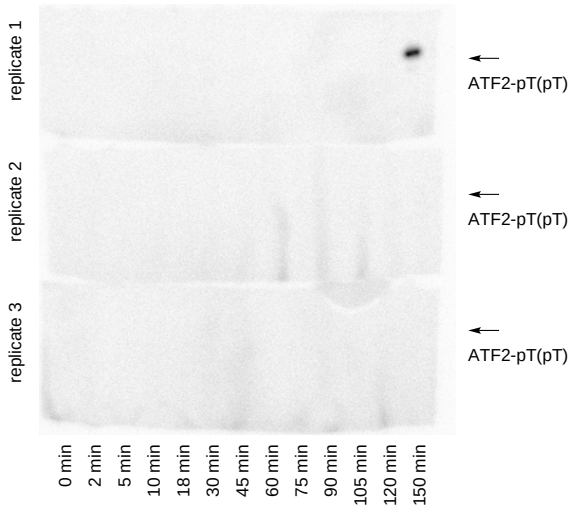
The dye front of the gels were cut off and the gels washed in EDTA solution (∼100 mM) before mounting between acetates and exposing to a phosphoscreen for 1.5 h.

To allow for quantification, a dilution series of the ATP (both ATP and [γ-^{32}P]-ATP) at concentrations as used in the reaction was also made. The dilution series was spotted on Whatman paper No. 1 and

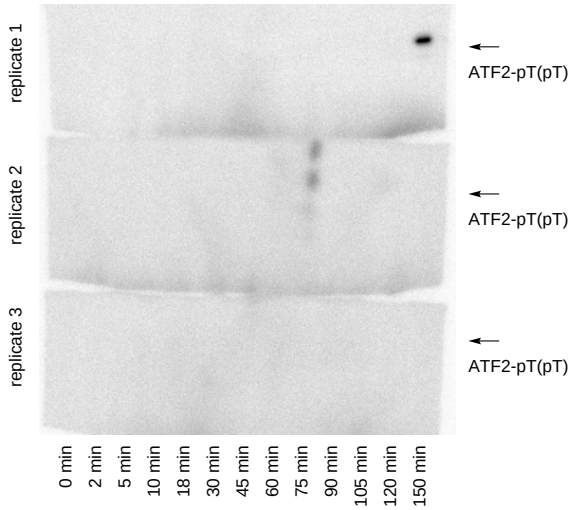
exposed to the phosphoscreen at the same time as the gels. Finally, the screens were scanned using a Storm Phosphoimager. The gels from all the variants showed timecourses of no conversion, except for the ones with p38α-pCys180 in **(f)**, active p38α-WT in **(g)**, and inactive p38α-WT in **(h)**. The gels are displayed as triplicates. The remaining bright spots in the other gels are positive controls (except for in **(e)** where the positive control was **(f)**).



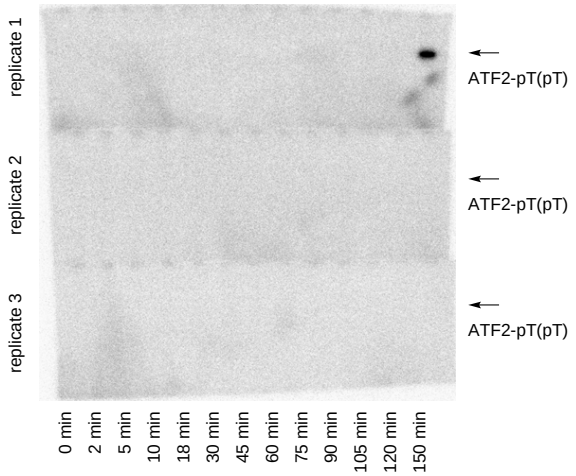
(a) Kinetic assay with p38α-Cys172. The only spot visible is a positive control where only one positive sample was run per triplicate.



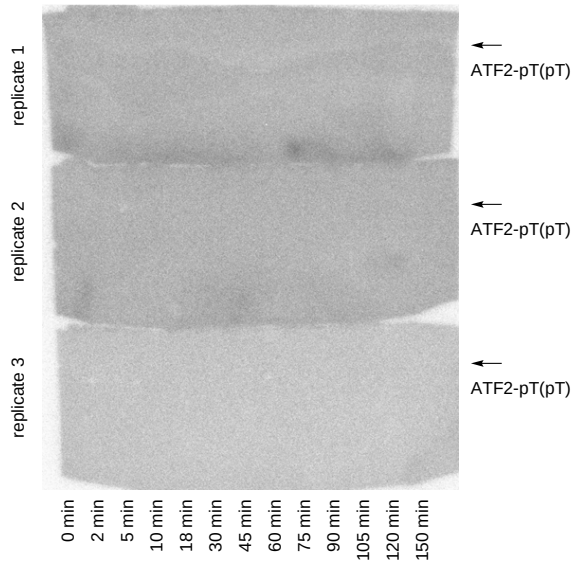
(b) Kinetic assay with p38α-Dha172. The only spot visible is a positive control where only one positive sample was run per triplicate.



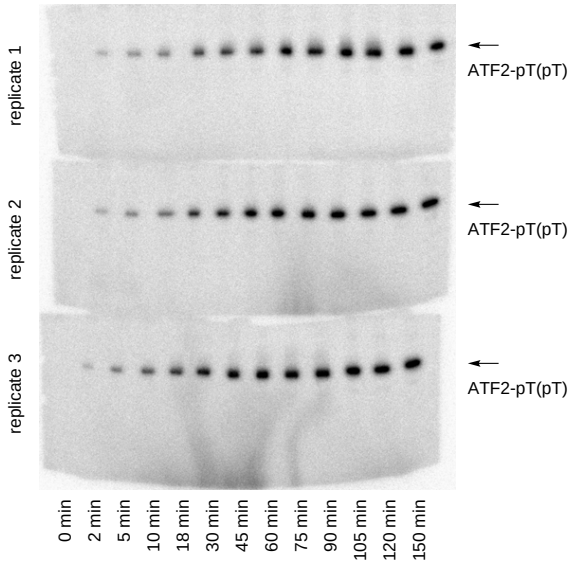
(c) Kinetic assay with p38α-pCys172. The only spot visible is a positive control where only one positive sample was run per triplicate.



(d) Kinetic assay with p38α-Cys180. The only spot visible is a positive control where only one positive sample was run per triplicate.

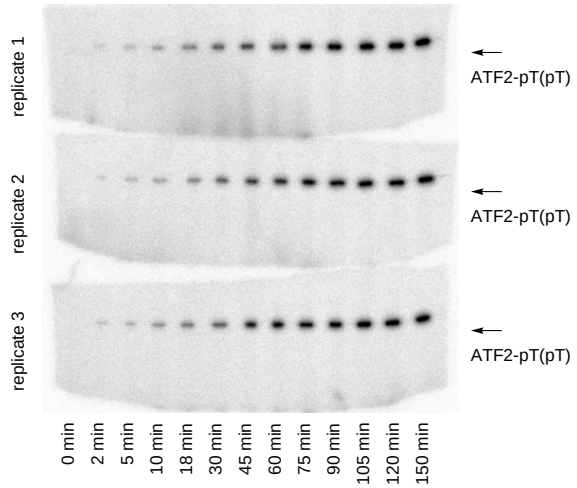


(e) Kinetic assay with p38α-Dha180. Gel **(f)** was run concurrently with this one and acts as the positive control.

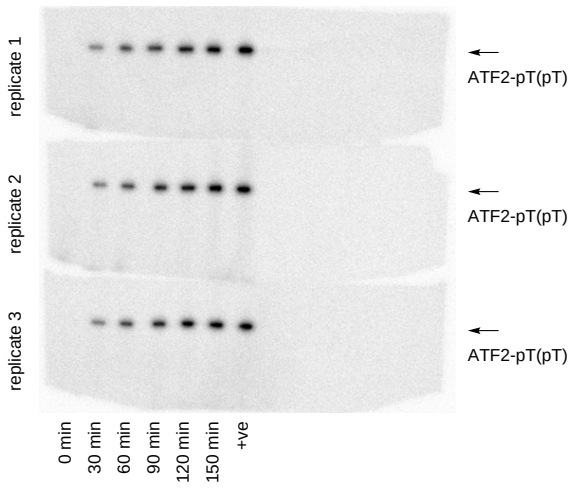


(f) Kinetic assay with p38α-pCys180.

p38α-WT Controls:



(g) Kinetic assay with active p38α-WT.

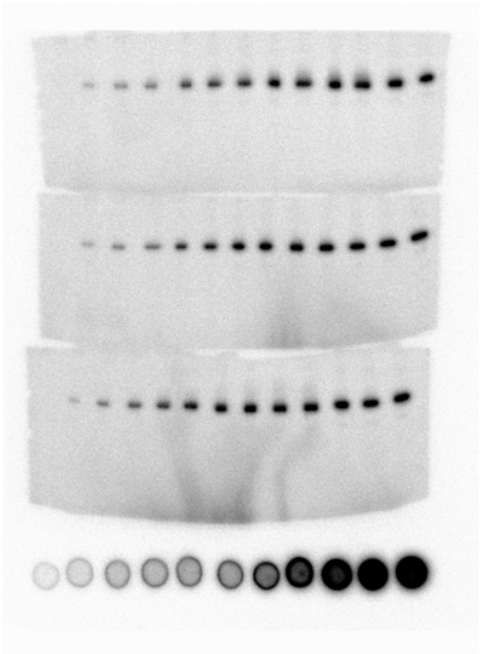


(h) Kinetic assay with inactive p38α-WT.

Data Analysis and Quantification

Gel densitometry was done using ImageQuant (GE Healthcare). In ImageQuant, square boxes of equal size were drawn around each gel spot, and around a representative blank area of the gel for the background reading. In a similar fashion, circles of equal size were drawn around spots corresponding to the ATP standard dilution series. The intensity of all the pixels within all the selected areas was counted and

the data were output into the appropriate programs (Excel or MATLAB) for further mathematical and statistical analysis. A typical Phosphoimager scan showing gels above and standard dilution series below:



For quantification, the measured intensities from the standard dilution series were plotted against to the known quantity of radiation put into each spot, giving rise to a standard curve. The gel spots containing unknown quantities of radiation could then be quantified against this curve.

3.8.4 Quantitative Measurement for Comparison of MS to Electrophoretic Radioassay

Enzymatic Reactions

10× kinase assay buffer (7.5 µL) was diluted with water (20.16 µL) and ATP (7.5 µL of a 5 mM solution in water) was added. ATF2 (12 µL) in ATF2 storage buffer was diluted with water (13 µL) and this solution was added (24.84 µL) to the buffer/ATP master mix. After taking a sample (4.0 µL) for t = 0, kinase (14.0 µL of a 20 µM solution) in p38α reaction buffer was added to initiate the reaction. Samples (5 µL) were taken at further timepoints (2, 5, 10, 18, 30, 45, 60, 75, 90, 105, 120, 150 and 180 min, times corresponding to the radioassay) and quenched in urea solution (10 µL of a 10 M in p38α reaction buffer). Samples were then analysed by LC-MS using a Chromolith® FastGradient RP-18 endcapped 50×2 mm monolithic HPLC column (Merck).

Data Analysis and Quantification

LC-MS datasets were processed using MassLynx as previously described (Section 2.11.1). Data from the deconvoluted spectra were output from MassLynx as a spectrum list for quantitative analysis using MATLAB R2012b.

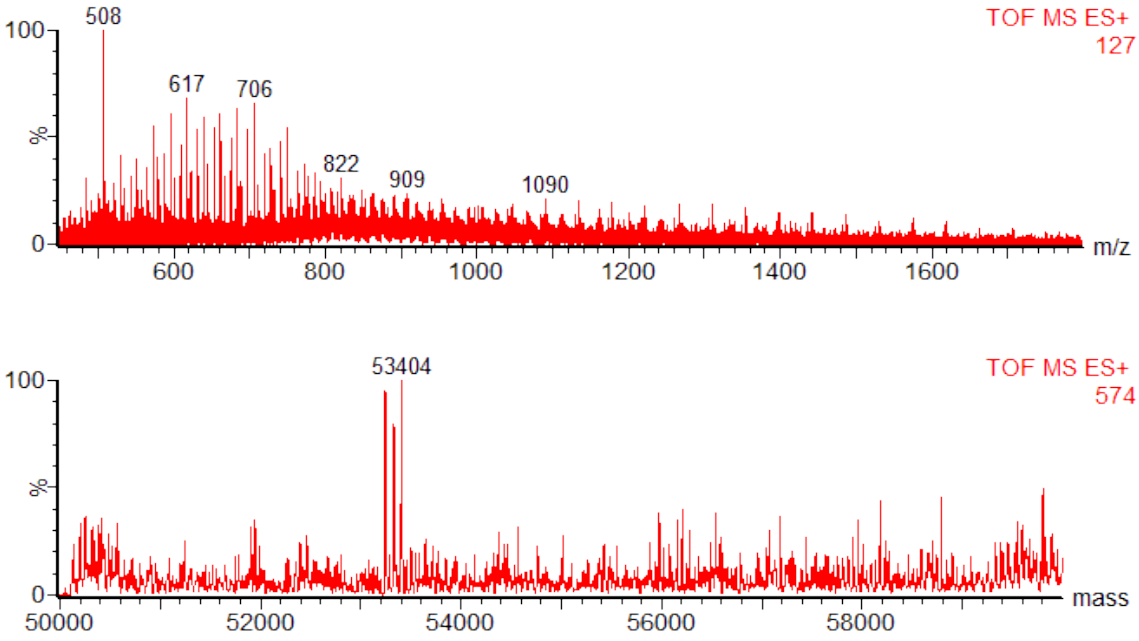
In MATLAB, quantities of the three substrate species (unphosphorylated, monophosphorylated and diphosphorylated) were estimated from relative peak intensity, setting the baseline at the median of all the values in the spectrum list. The peaks were picked by taking the maximum intensities from the three ranges where the peaks of these species were expected.

To convert the MS data so that it could be compared to the data from electrophoretic radioassay (ERA), the MS signals from the two phosphorylated species were added according to the equation $[\gamma\text{-}^{32}\text{P on ATF2}]_{\text{ERA}} = [\text{ATF2-P}]_{\text{MS}} + 2[\text{ATF2-PP}]_{\text{MS}}$. Both the curves from MS and ERA were normalised for the comparison.

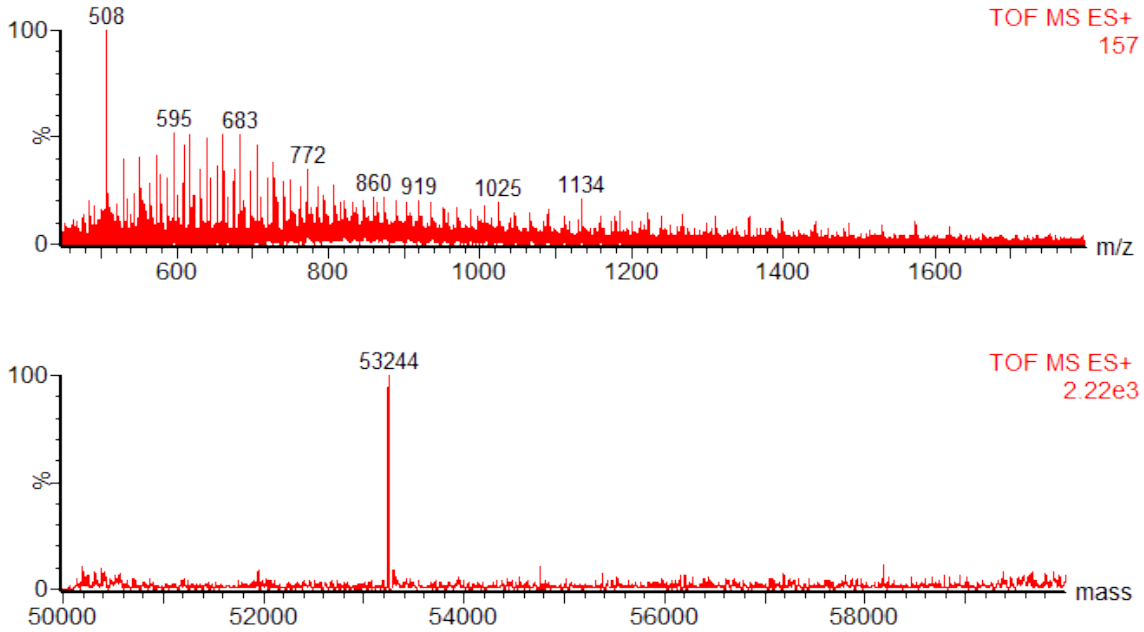
3.8.5 Measurement of Activity in Wild-type Controls for p38α

Assay reactions were set up in the same way as those for measuring the activity of the mutants except for changes in the p38α kinase concentration used (10 µL of a 80 nM solution) and the timepoints taken (0, 1, 2, 5, 10, 15, 20, 25 and 30 min).

Control: ATF2 after Treatment with Active p38α:



Control: ATF2 after Treatment with Unactive p38α:



3.8.6 Quantitative Measurement of the Kinetics of Enzymatically Active Forms of p38α

Enzymatic Reactions

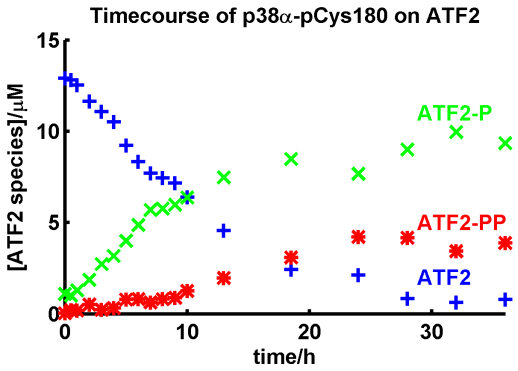
A modified procedure to the one above was used. 10× kinase assay buffer (60 µL), water (4.8 µL) and ATP (60 µL of a 5 mM solution in water) were combined to give a master mix. Aliquots (20 µL) were taken and ATF2 (74.2 µL, final concentrations: 14, 8, 4, 2, 1 and 0.5 µM) in ATF2 storage buffer added. After taking a sample (4.75 µL) for t = 0, kinase (4.75 µL of a 1.6 µM solution of either p38α-pCys180 or p38α-pThrpTyr, final concentration: 80 nM) in p38α reaction buffer was added to initiate the reaction. Samples (5 µL) were taken at further timepoints (0.5 h and every h for 10 h, then every 3–4 h for 36 h total time for p38α-pCys180, 2, 5, 10, 18, 30, 45, 60, 75, 90, 105, 120, then every 180 min for 5 h total time for p38α-pThrpTyr) and quenched in urea solution (10 µL of a 10 M in p38α reaction buffer). Samples were then analysed by LC-MS.

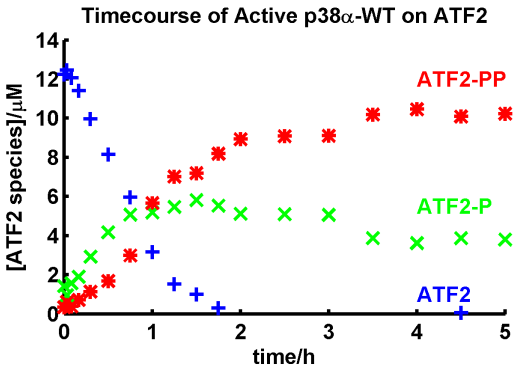
Data Analysis and Quantification

The MS data were analysed and quantified using the procedure described in Section 3.8.4. The peak proportions were then normalised to the starting ATF2 concentration. Determination of $k_{cat}/\mathrm{K_M}$ was done by taking the absolute value for the initial rates for each of the ATF2 species. This was plotted against substrate concentration and from the gradient, $k_{cat}/\mathrm{K_M}$ could then be calculated. Finally, the

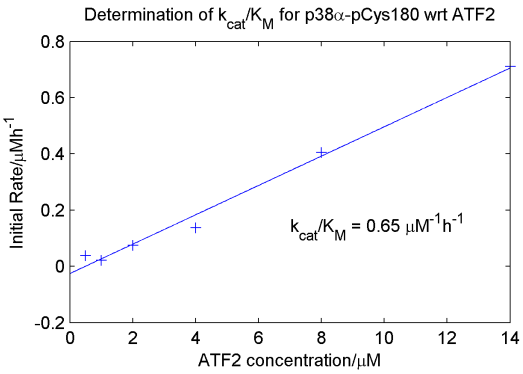
standard error in the value for the gradient was estimated by inputting the [S] vs. rate data into Graphpad PRISM v.5.01.

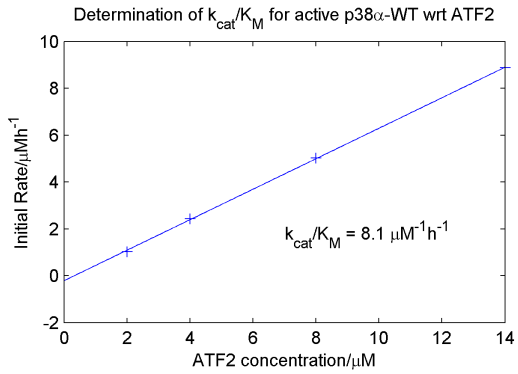
Example timecourses:



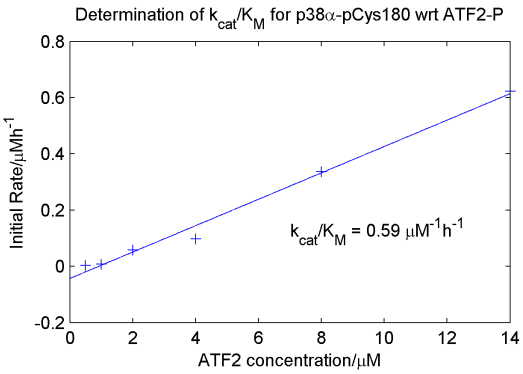


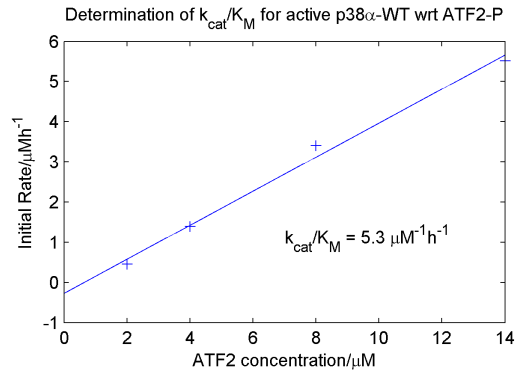
ATF2:



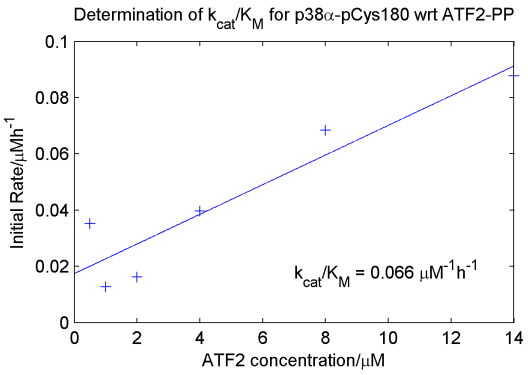


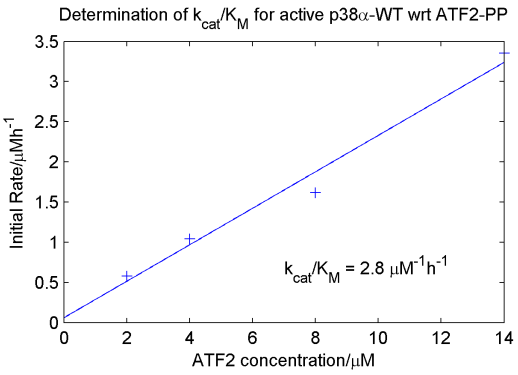
ATF2-P:





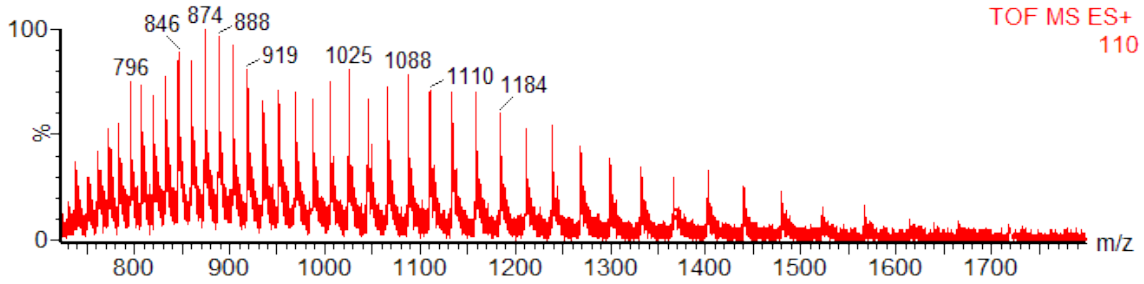
ATF2-PP:

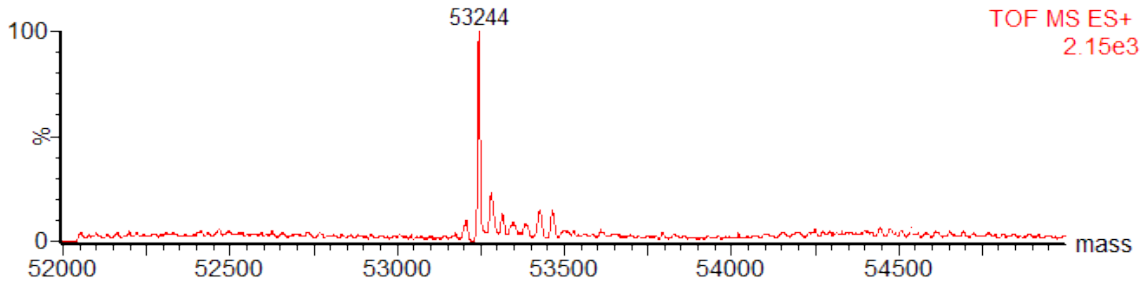




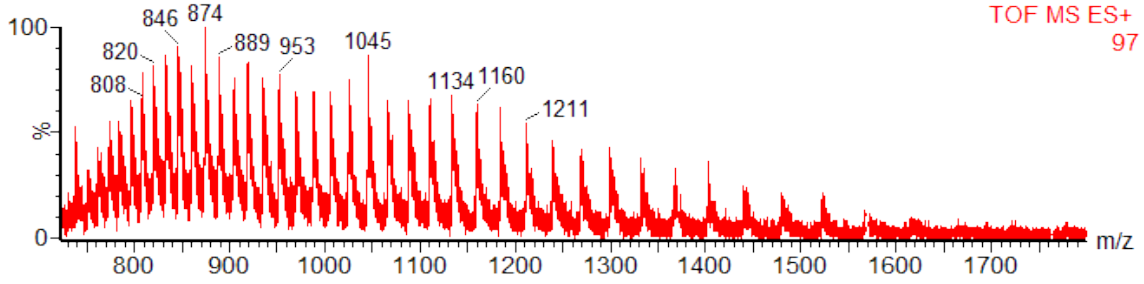
MS of a representative timecourse (for p38α-pCys180, [ATF2] = 14 μM):

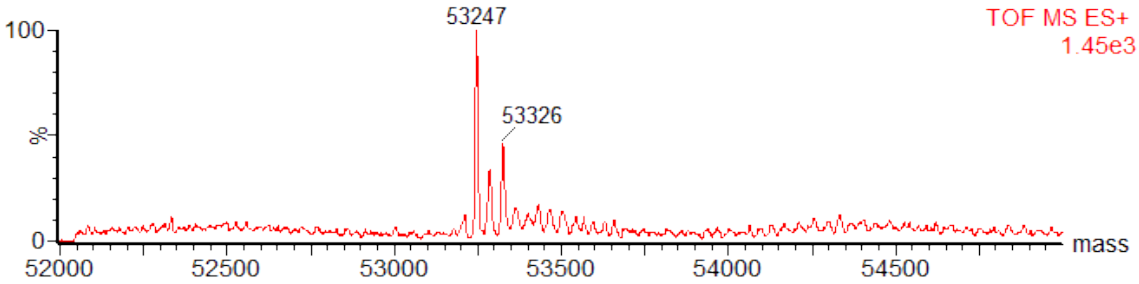
Start:



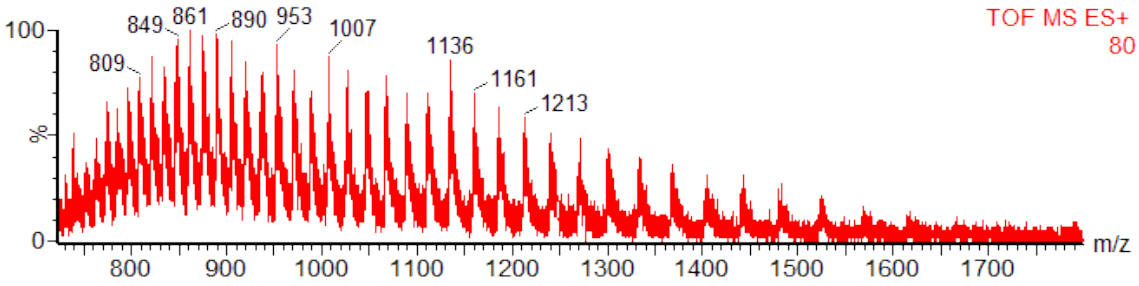


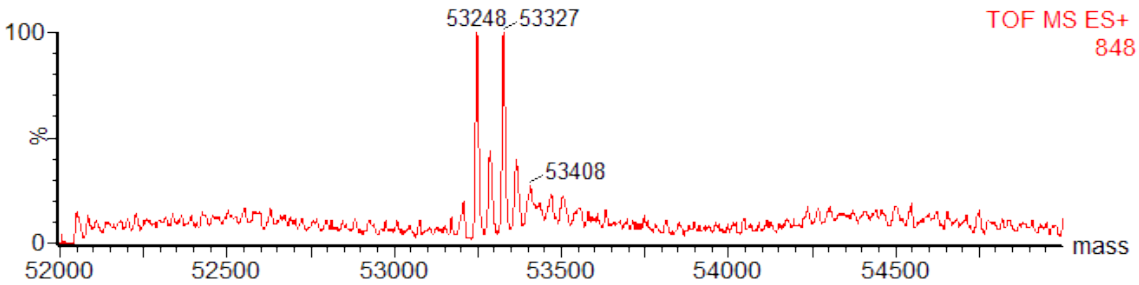
After 5 h:



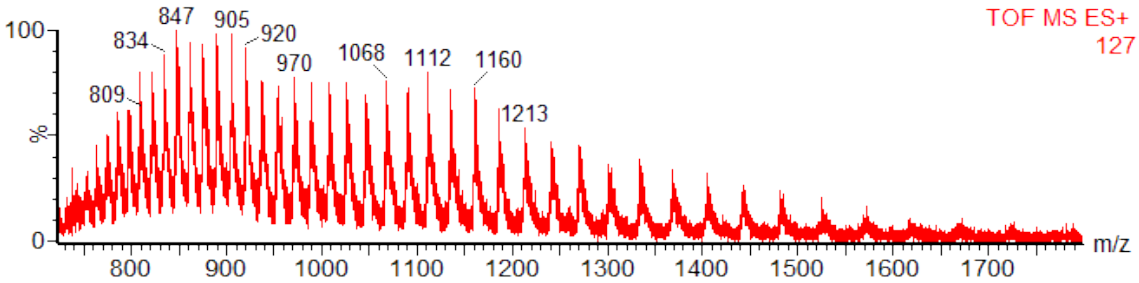


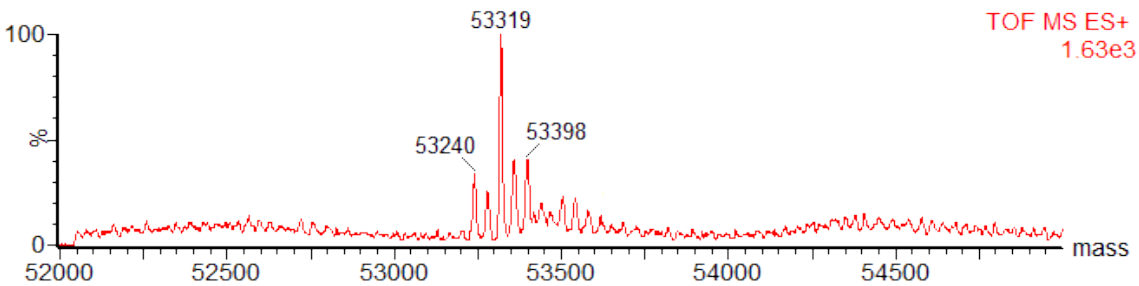
After 10 h:



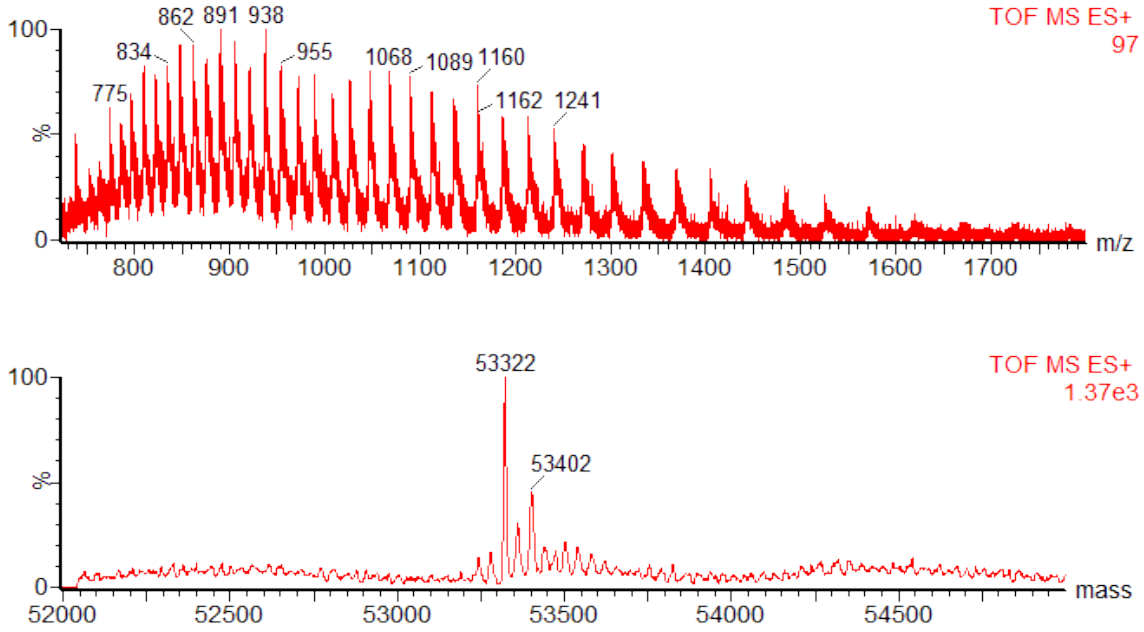


After 18.5 h:





After 36 h:



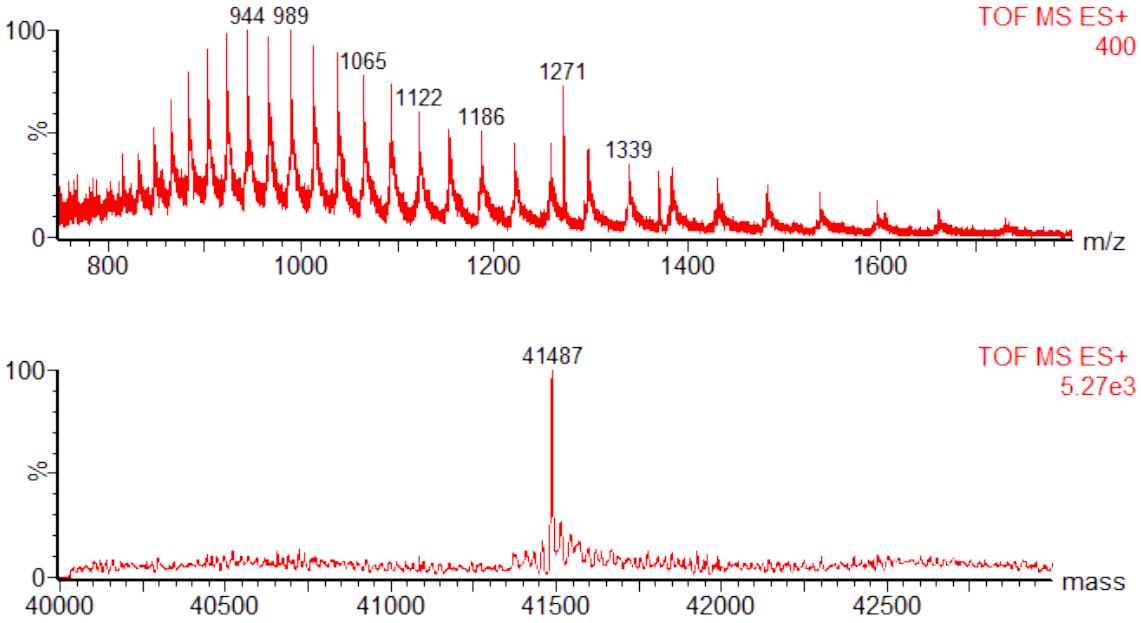
Data Reanalysis for Intermediate ATF2 Species Adducts

Close observation of the mass spectra revealed an adduct mass from the parent peak of +37–40 Da, which could be explained by cyanate adduct formation from the urea quench.[109,110] Reanalysis of the data were done to take this adduct into account, adding the mass of the adduct onto the corresponding parent peak and repeating the subsequent timecourse and kinetic analysis as described above. The two analyses revealed the same trends as observed in the previous analysis.

Control to Determine Phosphocysteine Stability against Assay Buffer Conditions

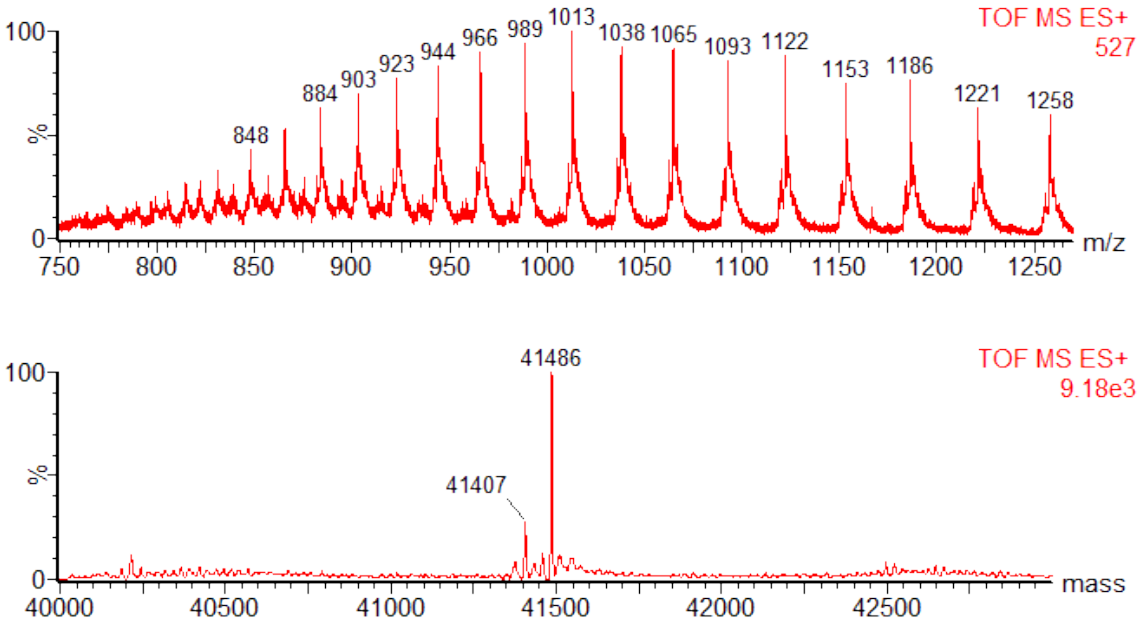
Due to the potential for phosphocysteine to be unstable towards reducing buffer conditions,[16] the stability of the chemically modified p38α-pCys180 was verified. In collaboration with Sébastien Galan, the same buffer conditions as described in Section 3.8.6 were used. 10$\times$ kinase assay buffer (0.5 mL), ATF2 storage buffer (3.7 mL), water (40 µL), ATP (0.5 mL of a 5 mM solution, pH 7.0) and p38α reaction buffer (217 µL) were combined. An aliquot (993 µL) was taken and p38α-pCys180 added (6.64 µL of a 1.0 mg/mL solution in p38α reaction buffer, final concentration: 80 nM). The protein was left to incubate at room temperature for 18 h. After this time, the protein was concentrated using Vivaspin 500 (MWCO 10 000) and the concentrated sample analysed by LC-MS. Analysis showed no detectable

amounts of dephosphorylation product (p38α-pCys180: calculated mass: 41 486, observed mass: 41 487, p38α-Cys180: calculated mass: 41 406).

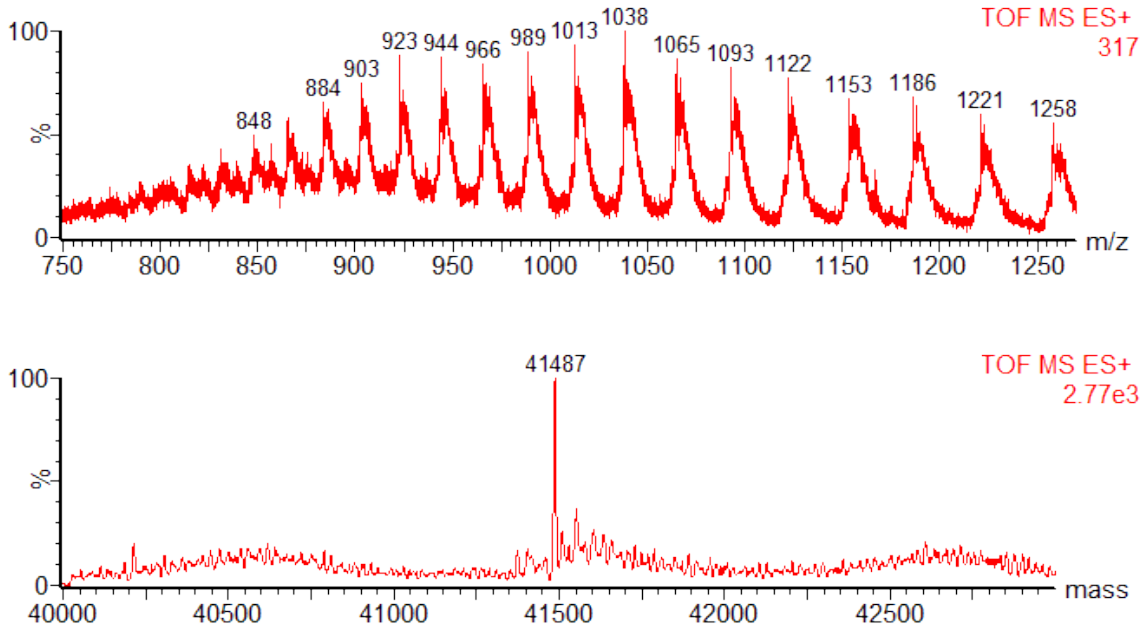


Control to Determine Thiophosphate Stability against DTT Treatment

In collaboration with Sébastien Galan, to p38α-pCys180 (10 µL of a 0.5 mg/mL solution) in p38α reaction buffer was added DTT (10 µL of a 414 mM solution, final DTT concentration: 207 mM) in reaction buffer. The mixture was incubated at room temperature for 18 h. LC-MS analysis after this time showed ~15% conversion to the dephosphorylated product (p38α-pCys180: calculated mass: 41 487, observed mass: 41 487, p38α-Cys180: calculated mass: 41 407, observed mass: 41 406).



In a similar experiment, a lower DTT concentration (final concentration: 2.75 mM) was used. LC-MS after 18 h showed no detectable dephosphorylation.



3.8.7 LC-MS/MS Analysis for Detection of Phosphopeptides after ATF2 Phosphorylation

Protein Digestion

In collaboration with Ritu Raj, proteolytic digestion of phosphorylated ATF2 was done using the FASP protocol[123] with 10 kDa Microcon filtration devices (Millipore). Protein sample (1–2 µg) was mixed with urea solution (200 µL of an 8 M solution in 50 mM ammonium bicarbonate buffer), loaded into the filtration devices and centrifuged (14 000 × g, 15 min). The concentrate was reduced in the filtration device by adding DTT (100 µL of a 50 mM solution in 50 mM ammonium bicarbonate buffer) at 50 °C for 20 min, followed by centrifugation (14 000 × g, 15 min). The sample was alkylated with iodoacetamide (100 µL of a 50 mM solution in 50 mM ammonium bicarbonate buffer) for 25 min in the dark. Excess alkylating agent was removed by centrifugation (14 000 × g, 15 min), followed by washes of the centrifugal unit with ammonium bicarbonate buffer (3 × 100 µL). The resulting concentrate was diluted with ammonium bicarbonate buffer (50 µL of a 50 mM solution) containing trypsin (final enzyme:protein ratio: 1:50 (w/w)). The sample was incubated overnight at 37 °C. Following overnight digestion, peptides were eluted to a fresh Eppendorf tube by centrifugation. Ammonium bicarbonate buffer (50 µL) was added to the centrifugal unit and centrifuged again to elute the remaining peptides. The digested protein was diluted (final

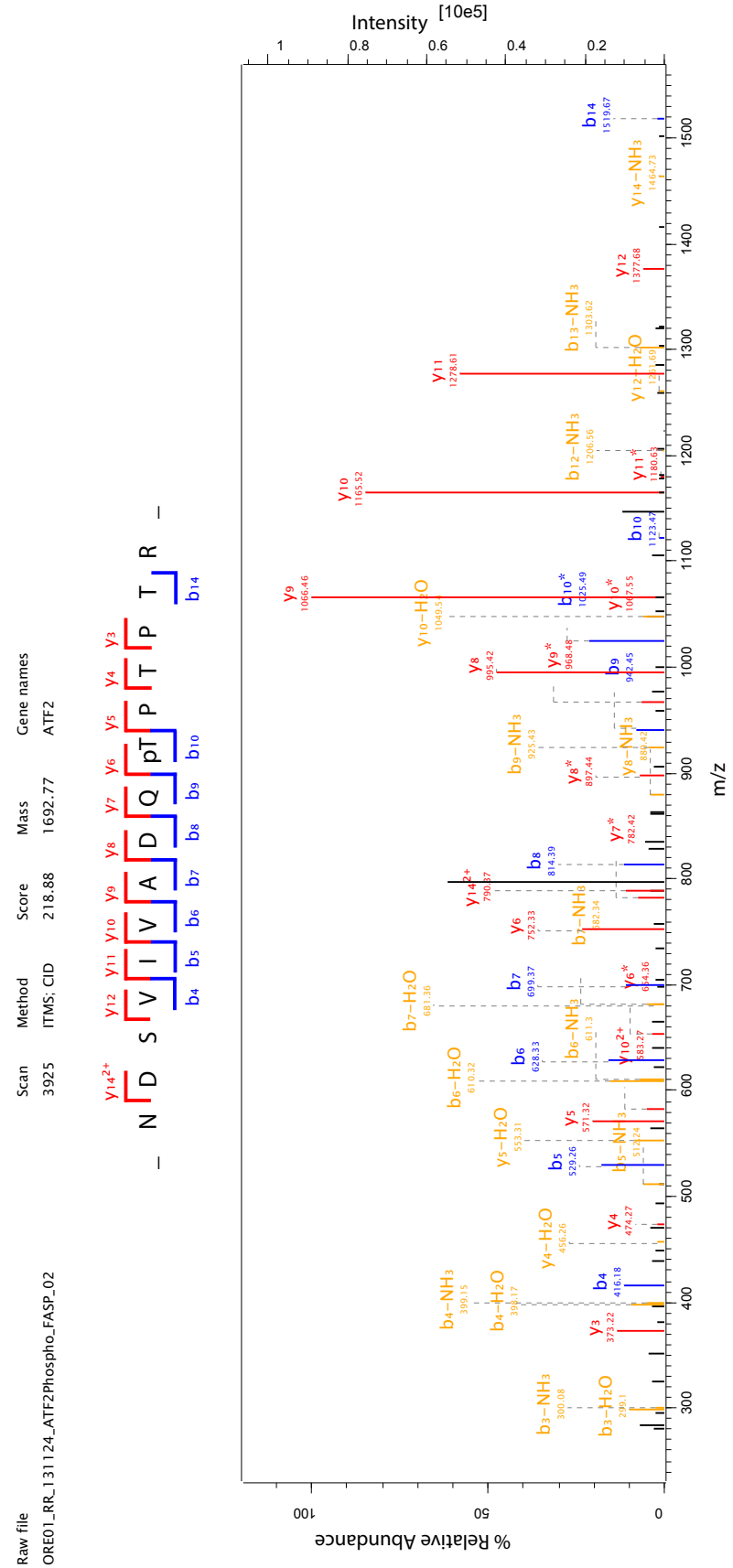
concentration: 100 fmol/µL) with 1% formic acid for analysis. LC-MS/MS analysis of the samples was then performed according to the conditions in Section 2.11.9 for phosphopeptides.

Data Analysis

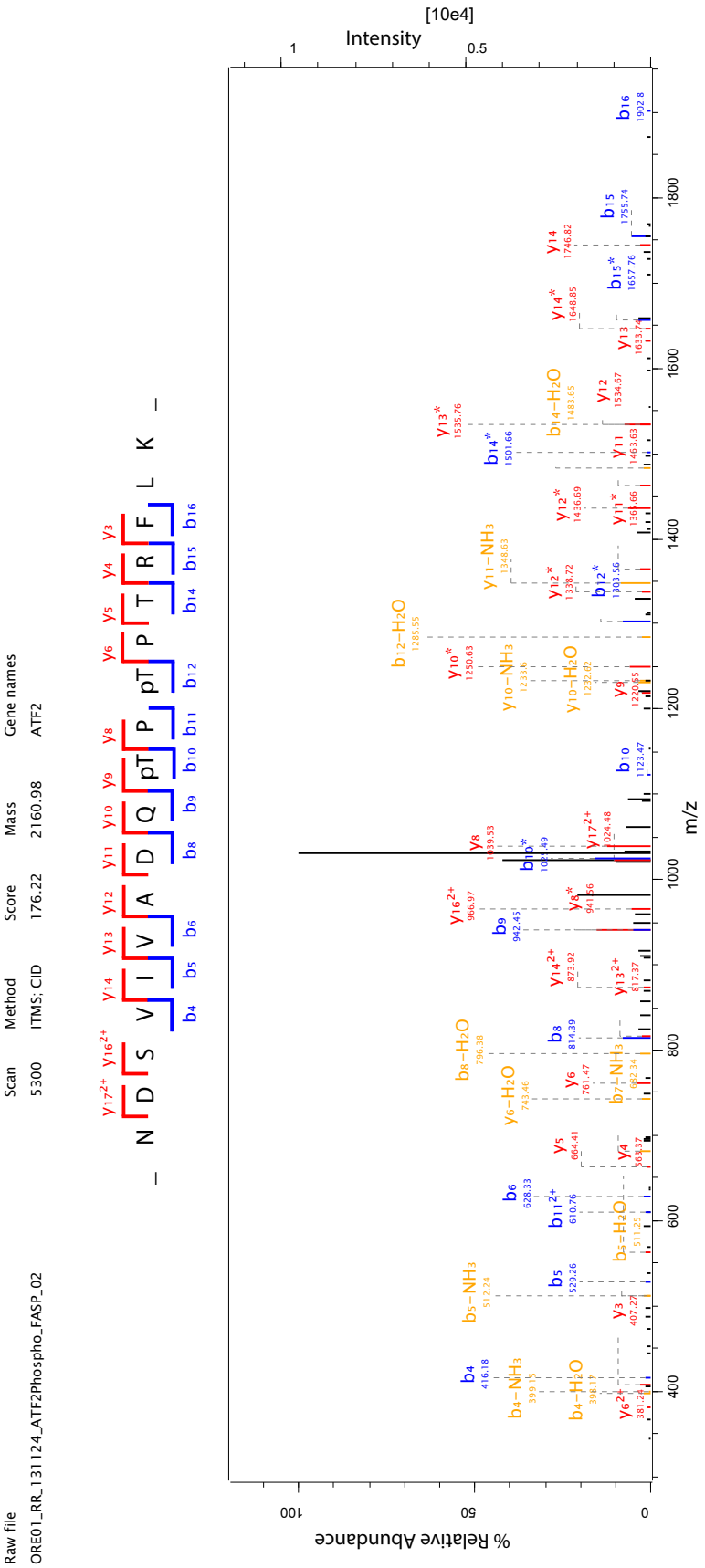
The data were processed similarly as for phosphopeptide data from the p38α variants (Section 2.11.9). The raw data files generated were processed using MaxQuant software (Version 1.4.1.2), integrated with Andromeda search engine as described elsewhere.[96,97] For identification of phosphorylated peptides, Andromeda searched peak lists against the human database (UniProt) as well as against a list of common contaminants. Trypsin was selected as specific digestion mode with maximum number of missed cleavages set at 2. Acetylation (N-term), oxidation (M) and phosphorylation (STY) were used as variable modifications and carbamidomethylation (C) as fixed modification. All spectra were manually validated.

The analysis showed that only two of the three possible phosphopeptides corresponding to Thr69 and Thr71 phosphorylation were detected: 1. mono-phosphorylation at Thr69 and 2. bis-phosphorylation at both sites. As was consistent with observations from whole protein ESI-MS, MS intensity for the mono-phosphorylated peptide was stronger than for the bis-phosphorylated one.

ATF2-pThr69Thr71:



ATF2-pThr69pThr71:



3.8.8 Determination of Inhibition Curves

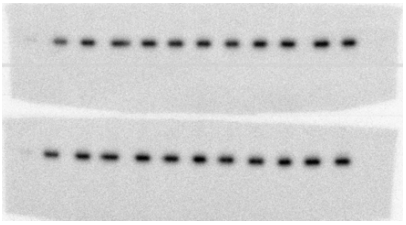
Enzymatic Reaction

Stock solutions (2.0 mM) of all inhibitors were made in DMSO. The stock was serially diluted with water (twice 10× dilution) to give a solution in 1% DMSO (20 µM). The stock was diluted to the first working concentration (8 µM) with 1% DMSO before a 2× serial dilution series was made with 1% DMSO (11–13 dilutions made).

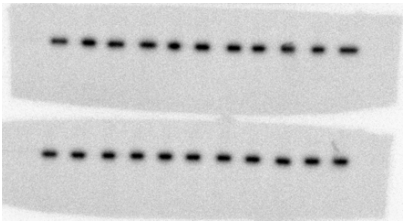
For enough samples corresponding to 10 IC_{50} curves, a master mix of 10× kinase assay buffer (75.0 µL), water (28.5 µL) and ATF2 (kinase substrate) (59.6 µL of a 1 mg/mL in ATF2 storage buffer) and kinase (either p38α-pCys180 or p38α-pThrpTyr, 45 µL of a 400 nM solution in p38α reaction buffer) was made. An ATP mixture of cold ATP (75.0 µL of a 2.5 mM solution in water) and [γ-^{32}P]-ATP (5.63 µL of a 11.6 mCi/mL solution) was made separately.

Test samples were made by adding master mix (4.01 µL) to inhibitor solution (1.875 µL) and incubated for 30 min at room temperature before initiation of the reaction by addition of ATP mixture (1.50 µL). Samples were then incubated at 19 °C for 30 min for p38α-pThrpTyr, or for 24 h for p38α-pCys180. After this time, the reactions were quenched using SDS/EDTA loading dye, the samples loaded onto 12% gels for SDS-PAGE analysis and the gels imaged by autoradiography as previously described (Section 3.8.3). Final reaction concentrations were also the same as previously described, except for kinase concentration (80 nM final concentration).

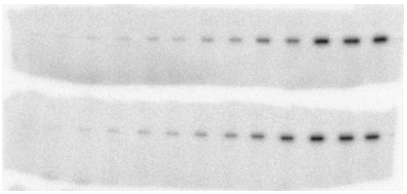
with p38α-pCys180, VX745



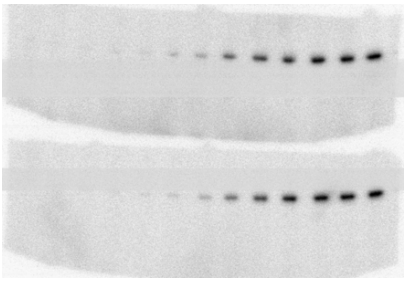
with p38α-pCys180, TAK715



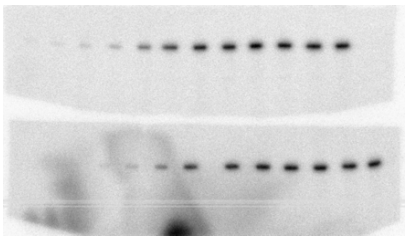
with p38α-pCys180, BIRB796



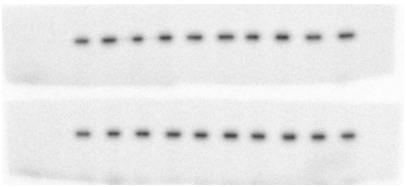
with p38α-pCys180, JX401



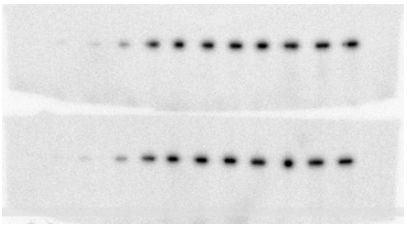
with p38α-pCys180, SB202190



with p38α-pCys180, inhibitor **20**

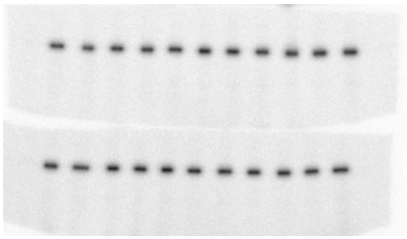


with p38α-pCys180, inhibitor **21**

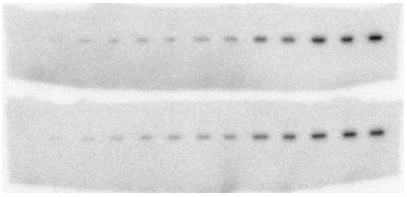


with active p38α-WT, VX745

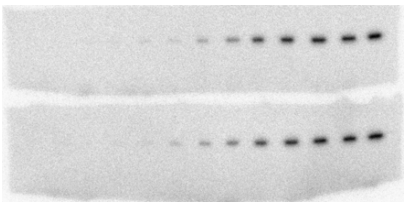
with active p38α-WT, TAK715



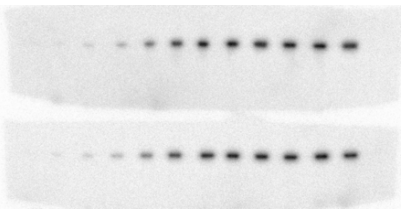
with active p38α-WT, BIRB796



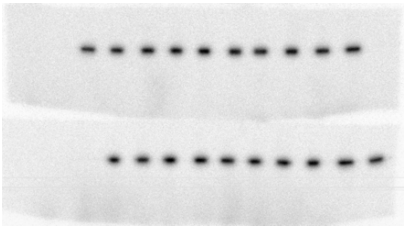
with active p38α-WT, JX401



with active p38α-WT, SB202190



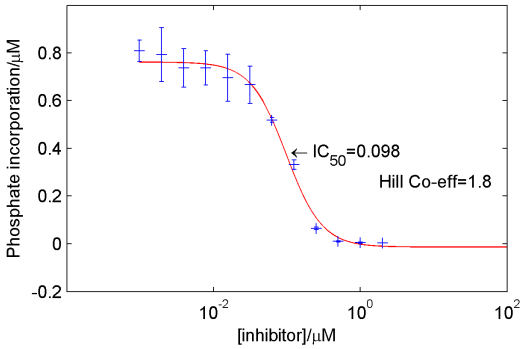
with active p38α-WT, inhibitor **20**



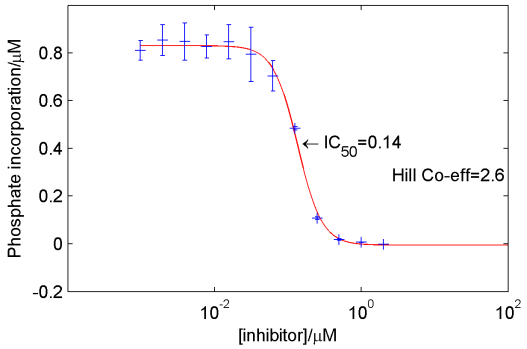
with active p38α-WT, inhibitor **21**

Data Analysis and Quantification

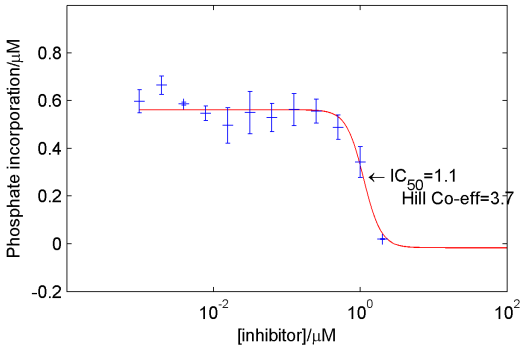
Gel densitometry was performed using ImageQuant (GE Healthcare) as described above (Section 3.8.3). The data from the gel as determined from ImageQuant was input into MATLAB for fitting to the 4 parameter logistic function.[124]



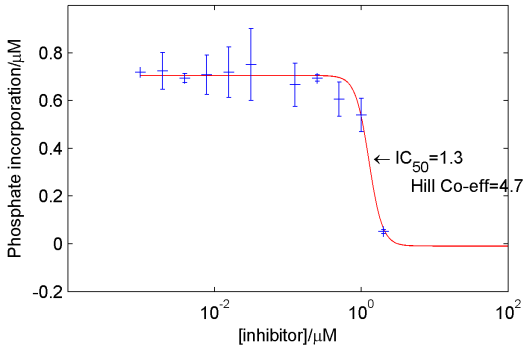
with p38α-pCys180, VX745



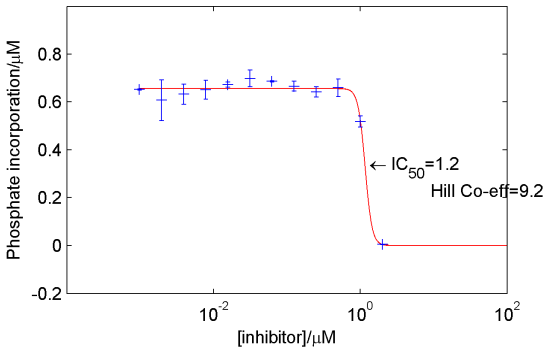
with active p38α-WT, VX745



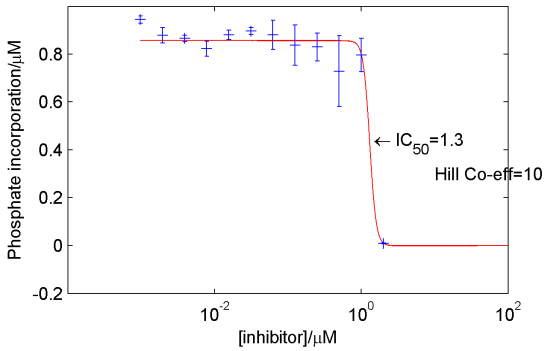
with p38α-pCys180, TAK715



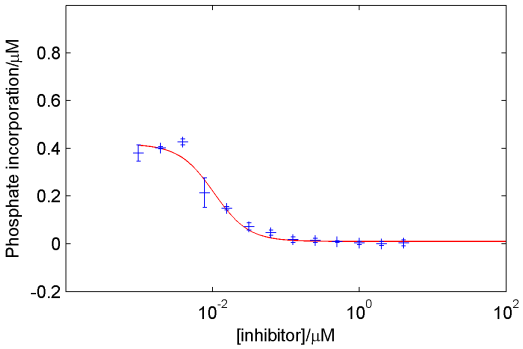
with active p38α-WT, TAK715



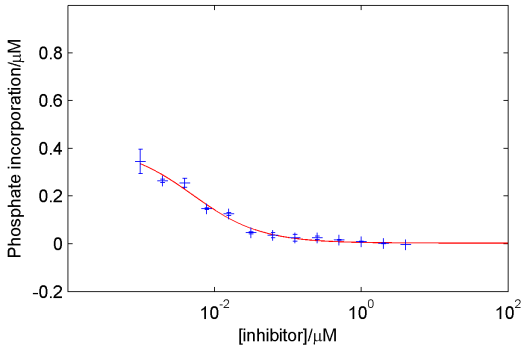
with p38α-pCys180, BIRB796



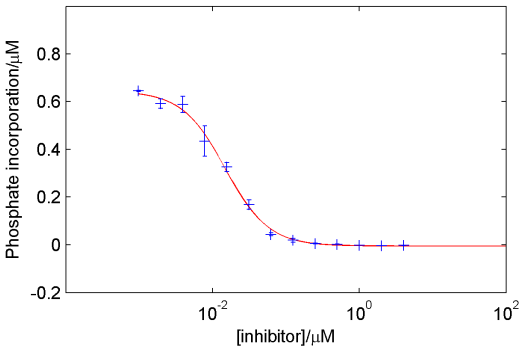
with active p38α-WT, BIRB796



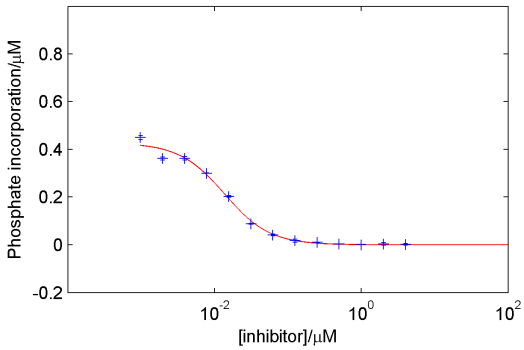
with p38α-pCys180, JX401



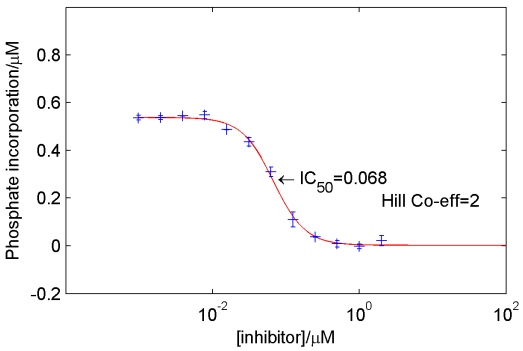
with active p38α-WT, JX401



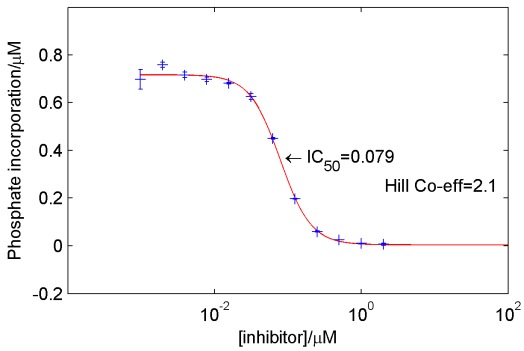
with p38α-pCys180, SB202190



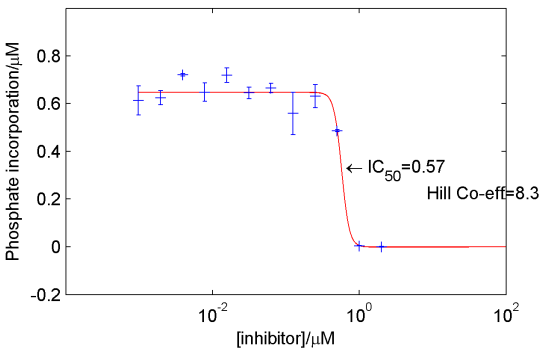
with active p38α-WT, SB202190



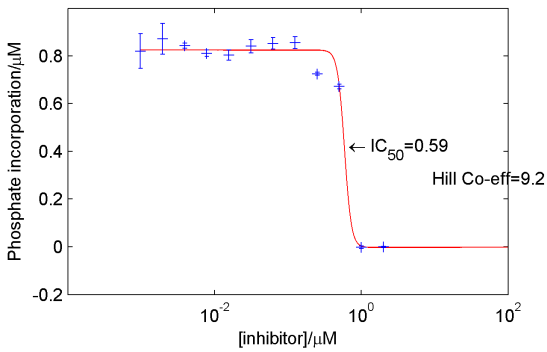
with p38α-pCys180, inhibitor **20**



with active p38α-WT, inhibitor **20**



with p38α-pCys180, inhibitor **21**



with active p38α-WT, inhibitor **21**

Chapter 4

Chemical Modification of MEK1 Kinase

Following the success in demonstrating the enzymatic activity and biological recognition capabilities of p38α-pCys180, a synthetic mimic of the naturally occurring p38α-pThr180, MEK1 kinase was selected for further studies, in the pursuit of broadening the scope for studying kinases by chemical modification. Phosphocysteine had thus been shown to be a sufficient mimic of phosphothreonine in the p38α study, and a similar link between phosphocysteine and phosphoserine was to be investigated with MEK1. Suitable MEK1 variants therefore were to be designed for this further study.

4.1 Construct Design

As for the chemical modification of p38α, the reliance on cysteine as the reactive residue for chemical modification meant that cysteine needed to be introduced at the desired site for subsequent modification. At the same time, potentially competing native cysteines of MEK1 had to be mutated out. With no existing plasmid containing the *MEK1* gene in-hand, the required *MEK1* mutant was designed and synthesised *de novo*. Close scrutiny of the MEK1 native cysteines in crystal structure (Fig. 4.1) showed two cysteines, Cys277 and Cys376, to be likely candidates for any off-target modification. These were both promptly mutated to serine. A further two cysteines, Cys142 and Cys341, appeared to be solvent inaccessible, and were thus left intact. A final two cysteines, Cys121 and Cys207 were considered to be potentially reactive but given their proximity to the ATP binding cleft, mutation away from cysteine could prove to be detrimental to the enzymatic activity of MEK1 further down the pipeline (Fig. 4.1). These two cysteines were therefore also left intact.

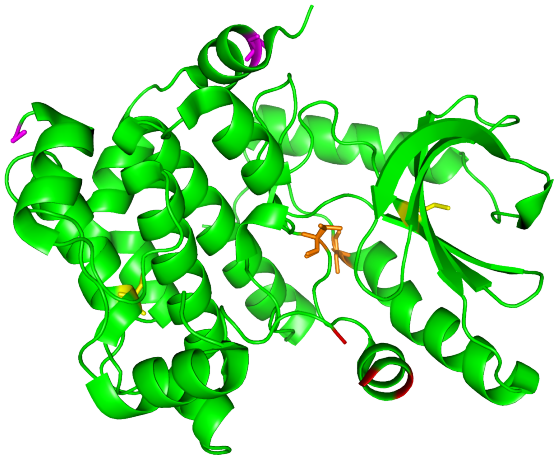


Fig. 4.1 Crystal structure of MEK1 kinase highlighting the surface accessibility of native cysteine residues. The cysteine are of 3 categories: highly solvent accessible Cys277 and Cys376 (magenta), hindered Cys121 and Cys207 (orange), and buried Cys142 and Cys341 (yellow). The sites of native phosphorylation and therefore the target sites of chemical modification, Ser218 (red, in the α-helix) and Ser222 (red, spanning the gap adjacent to the α-helix) are also highlighted (red). Drawn from 3EQC.pdb, a structure with MEK1 in its DFG-in conformation.[5]

MEK1 is natively phosphorylated at two positions for full activation, Ser218 and Ser222, and thus unlike with p38α, investigation of phosphorylation at both sites could be benefited by the use of chemical modification techniques. However, for the purposes of making a model system for chemical modification on MEK1, and for simplicity, only one of these sites was chosen to begin with. Turning to the crystal structure once again, close investigation of the MEK1 activation loop showed that there were some residues missing from the structure. These residues corresponded to the region containing Ser222, suggesting that this region was too disordered to obtain structural information in this instance. Conversely, as part of an α-helix, Ser218 is found in the crystal structure. Thus, it was hypothesised that Ser222 is in a more flexible region than Ser218, and is therefore more likely to be solvent accessible for chemical modification reactions. Position 222 was therefore selected as the position of choice for chemical modification studies, resulting in the final *MEK1* variant design, MEK1-S222C/C277S/C376S (**22**, also referred to as MEK1-Cys222) and a MEK1 variant that contained 5 cysteines overall.

Upon deciding on a vector to house the *MEK1* gene, the affinity tag and its location had to be considered. A His_6-tag was chosen for the affinity tag, due to its small size so that it wouldn't interfere with the protein's function if left uncleaved, while the *N*-terminal was chosen as the location for the tag, as this was further away from the ATP binding site than the *C*-terminus. The pET28a vector was chosen to fit

these criteria, resulting in the final plasmid pET28a-MEK1-S222C/C277S/C376S, which was synthesised by a commercial source.

4.2 Variant Library Expansion and Protein Gluconylation

In preparation for the biological studies that would follow later, a library of MEK1 variants closely related to MEK1-Cys222 was built up around the existing variant. The two constructs of note were: wild-type MEK1 (MEK1-WT) as a negative biological control, and MEK1-S222E as the positive control. The introduction of an acidic residue in MEK1-S222E is a well precedented way of MEK1 activation,[10,12,23,31,125] and therefore represents one of the commonly used methods of kinase activation available. Unlike p38α, which cannot be activated using this substitution, the study of MEK1 has benefited from this technique since the position of negative charge on aspartate or glutamate is similar to that of phosphoserine, or indeed phosphocysteine. Both of these mutants for the controls were obtained by site-directed mutagenesis of MEK1-Cys222 and all the variants in-hand subsequently expressed.

Two of the variants, MEK1-Cys222 required for chemical modification, and MEK1-S222E required as a positive control, were subsequently purified to a single band in SDS-PAGE analysis. Upon MS analysis however, both batches of the MEK1 variant showed two product peaks: one corresponding to the predicted molecular mass and a second, smaller peak at $+178$ Da (Fig. 4.2a and 4.2b). The presence of this second peak that occurs in the presence of an *N*-terminal His_6-tag is well precedented and represents the post-expression, *N*-terminal modification of glycine in the tag leader sequence by gluconic acid.[104] Although gluconylation was unlikely to affect the enzymatic activity of the His_6-tagged protein, particularly as the charge and size of this modification is not significant compared to the tag itself, the generation of an additional peak in the MS would make analysis of chemical modification reactions more complicated. It was thus decided to eliminate the possibility of *N*-terminal gluconylation, especially for variants undergoing chemical modification.

Replacement of the *N*-terminal glycine with either proline or phenylalanine are reported to stop gluconylation.[126] With the ready availability of the primer required for mutation to phenylalanine, an arbitrary choice was made to employ this mutation. Thus, Gly2 in the His_6-tag leader sequence was mutated accordingly. Expression, purification and subsequent MS analysis of this new variant confirmed that the variant was no longer being gluconylated, but unlike for variants without the G$\rightarrow$F mutation, the initiator

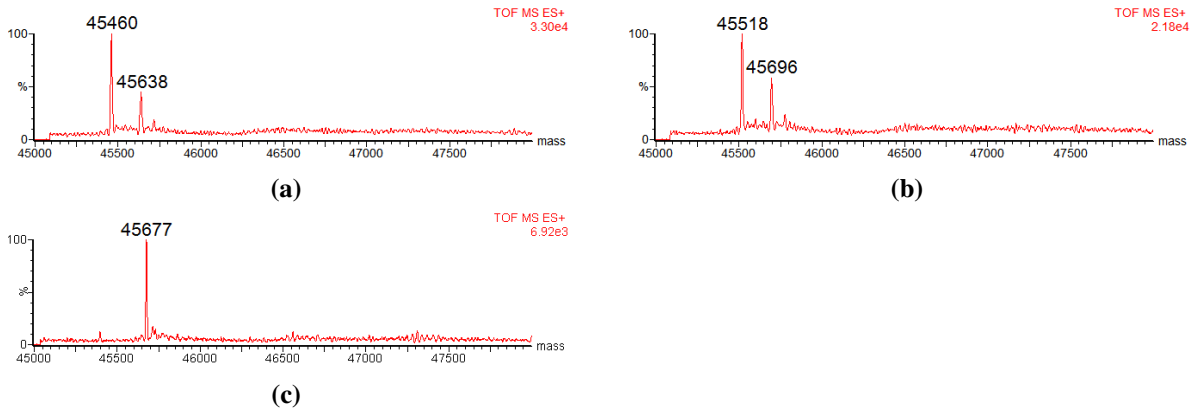


Fig. 4.2 Deconvoluted MS showing *N*-terminal gluconylation on **(a)** MEK1-Cys222 and **(b)** MEK1-S222E before the G(−19)F mutation was made. Both spectra display a smaller additional peak corresponding to the gluconylated product. Expected mass change upon gluconylation: +178 Da. **(c)** After G(−19)F mutation, the resultant variant of MEK1-Cys222 no longer suffered from gluconylation.

Met was retained, as precedented (Fig. 4.2c).[126] In keeping with the numbering of MEK1, this mutation gave rise to the MEK1 variant MEK1-G(−19)F/S222C/C277S/C376S (**23**), referred to as MEK1-Cys222 from here on in place of MEK1-S222C/C277S/C376S mentioned previously. Except for MEK1-S222E where the partially gluconylated material with the *N*-terminal glycine continued to be used and unless otherwise stated, this G(−19)F mutation was applied to all other variants of MEK1 used in this and the subsequent studies.

4.3 Increasing the Expression Yield of MEK1

Expression of the MEK1 variants produced thus far had always yielded the protein of interest in a soluble form, but the yield was prohibitively small towards large-scale chemical modification studies. Typical yields were ∼1 mg/L of culture. Given that a typical protein chemical modification reaction would require ∼0.2 mg/reaction and chemical reaction optimisation would likely require enough protein for tens of reactions, a total quantity of ∼10 mg would be needed for each variant. The protein expression yield therefore had to be increased in order to be practically viable.

The common parameters to change in an expression screen include the strain of *E. coli* used, expression temperature, expression time, the concentration of the inductant IPTG, and the nutrient medium used. The screening of expression strains was deemed unnecessary since the gene sequence had been optimised during synthesis for *E. coli* expression. Screening expression temperature and time were likewise not considered further since it is common practice to express eukaryotic proteins such as MEK1 at lower

temperatures (20 °C) overnight. Thus, only nutrient medium composition and IPTG concentration were screened.

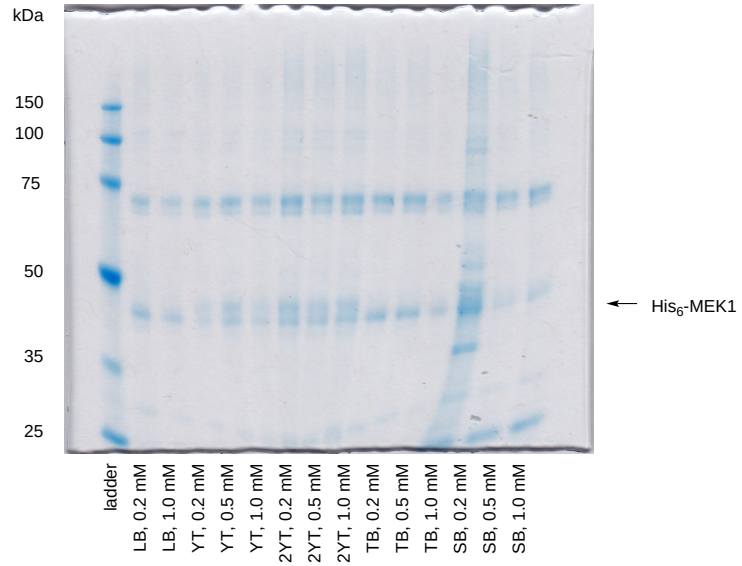
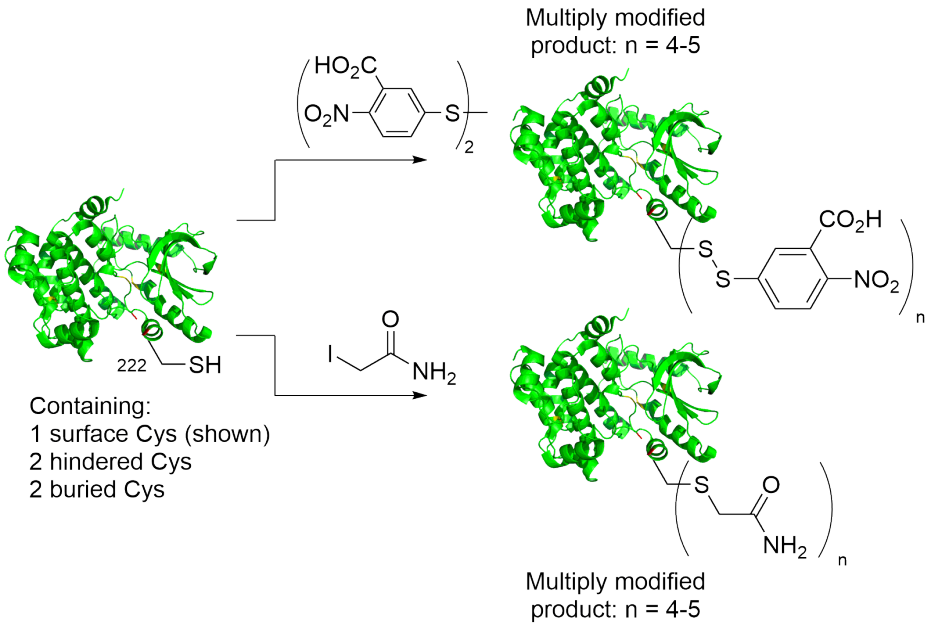


Fig. 4.3 SDS-PAGE analysis of purified samples from expression trials. The lane labels denote nutrient medium and IPTG concentration used for expressing the sample.

The range of media trialled were of increasing nutriment from LB, and thus able to support growing *E. coli* more adeptly. Thus, YT, 2YT, TB and SB media were tested and the range of IPTG concentrations varied within the typically used range. The trial showed that 2YT and TB media were consistent in giving higher levels of expression than LB medium over the range of IPTG concentrations used. Further, the IPTG concentration didn't have as much influence on expression levels as the type of medium used (Fig. 4.3). Since 2YT medium gave the highest level of expression given the number of cells grown (cells grown in TB medium need to be grown to $OD_{600} = 1.0$ for induction, compared to $OD_{600} = 0.6$ in unbuffered media such as 2YT), it was decided to use 2YT medium for further protein production. Deployment of 2YT over LB medium for the expression of MEK1-Cys222 resulted in an increase in expression levels of soluble protein from $\sim$1 mg/L to $\sim$3 mg/L.

4.4 Cysteine Accessibility Reactions

With an optimised method for producing MEK1 variants, attention could finally be turned to the chemical modification of MEK1-Cys222. However, before applying the tag-and-modify strategy of Dha installation and subsequent phosphate installation on MEK1 (see Section 1.7), a brief investigation was undertaken into how the remaining cysteines in MEK1-Cys222, which from the crystal structure were in different



Scheme 4.1 Reaction of MEK1-Cys222 with cysteine probes of Ellman's reagent and iodoacetamide. NB: Although both the reactions are drawn here schematically from the same MEK1 variant, the actual reactions themselves were undertaken with the G(−19)F negative and positive variants of MEK1-Cys222 respectively (see Fig. 4.4 for more details).

steric environments, would react with Ellman's reagent and iodoacetamide (Scheme 4.1). Somewhat unexpectedly, both reagents caused near stoichiometric modification of the 5 cysteines of MEK1-Cys222 (Fig. 4.4a and 4.4b).

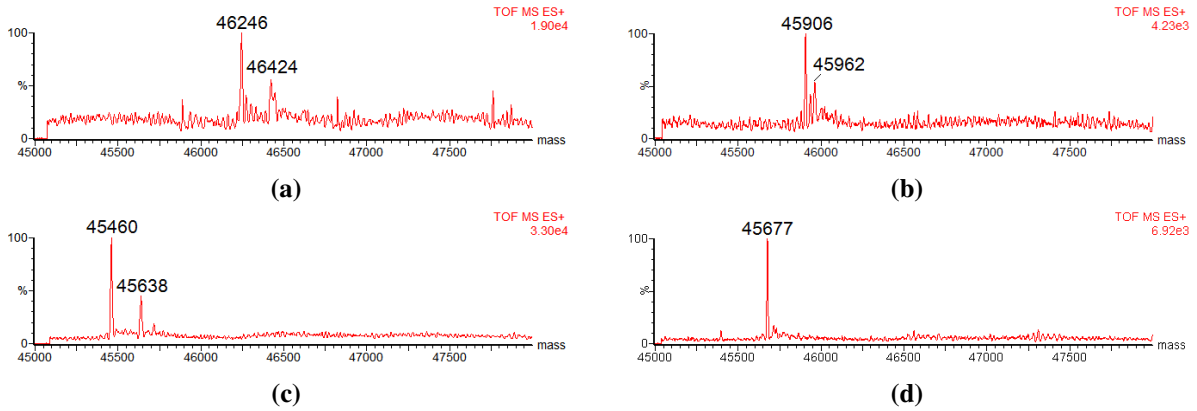
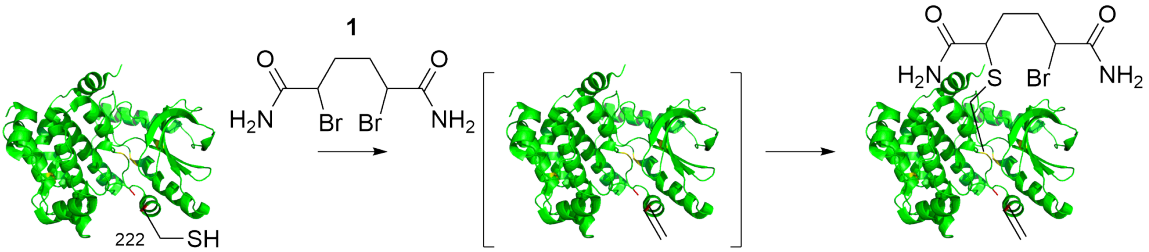


Fig. 4.4 Deconvoluted MS showing MEK1-Cys222 after reaction: **(a)** with Ellman's reagent. MEK1-Cys222 without the G(−19)F mutation was used in this reaction. The two peaks correspond to 4- and 5-adduct formation ($n = 4$ or 5). The peak of the 5-adduct is observed as a shoulder to the right of the gluconylated 4-adduct peak. **(b)** with iodoacetamide. MEK1-Cys222 with the G(−19)F mutation was used in this reaction. The two peaks correspond to 4- and 5-adduct formation ($n = 4$ or 5). **(c)** Unmodified MEK1-Cys222 without the G(−19)F mutation shown for comparison to **(a)** (reproduced from Fig. 4.2a for convenience). **(d)** Unmodified MEK1-Cys222 with the G(−19)F mutation shown for comparison to **(b)** (reproduced from Fig. 4.2c for convenience). Expected mass change from unmodified MEK1 variant upon: Ellman's adduct formation: +197 Da, Carbamidomethylation: +57 Da.

As had also been in the case with the use of cysteine probe reagents in the chemical modification studies on p38α (see Section 2.5), further use of these reactions had been intended either as a competitive chemical test for remaining Cys222 after installation of Dha at this site, (provided that the probe reagent also exclusively reacted with Cys222[43]) or more generally as an indicator for cysteine accessibility.[78,79,127] The greater extent of modification undergone by MEK1 compared to p38α meant that whereas these probe reagents saw limited use in the aforementioned capacities for the study on p38α, their use on MEK1 would be limited even further. It was therefore decided that these probe reactions would add little value to the study of chemical modification on MEK1 and were thus abandoned.

A potential explanation of the protein's behaviour against the cysteine probe reagents is that as the protein progressively reacts with the probe, it also causes the protein to progressively denature. In doing so, further cysteines are exposed, allowing them to also react.

4.5 Optimisation of the Installation of Dehydroalanine on MEK1-Cys222



23 **24** **25**

Scheme 4.2 Dha formation with dibromide **1** on MEK1-Cys222 **23** resulted in over-alkylation of the kinase (**25**).

Treatment of MEK1-Cys222 with dibromide **1** did not result in the installation of Dha at a single site, as *per* design (Scheme 4.2). Instead, further alkylation was observed (Fig. 4.5a), giving rise to the postulated product **25**. MS analysis taken at intermediate timepoints showed that prevention of this over-alkylation by terminating the reaction after the addition stage of the bis-alkylation–elimination reaction would not have been possible: the over-alkylated product was concurrent to unreacted MEK1-Cys222 (Fig. 4.5b). The already slow nature of Dha incorporation, even with super-stoichiometric quantities of dibromide **1** (150 eq.) meant that attempting to prevent over-alkylation by using nearer-to-stoichiometric quantities of **1** (10 eq. instead of 150 eq.) was not practically viable, since the reaction would take too long and the protein was unlikely to be stable enough to endure the required long reaction time.

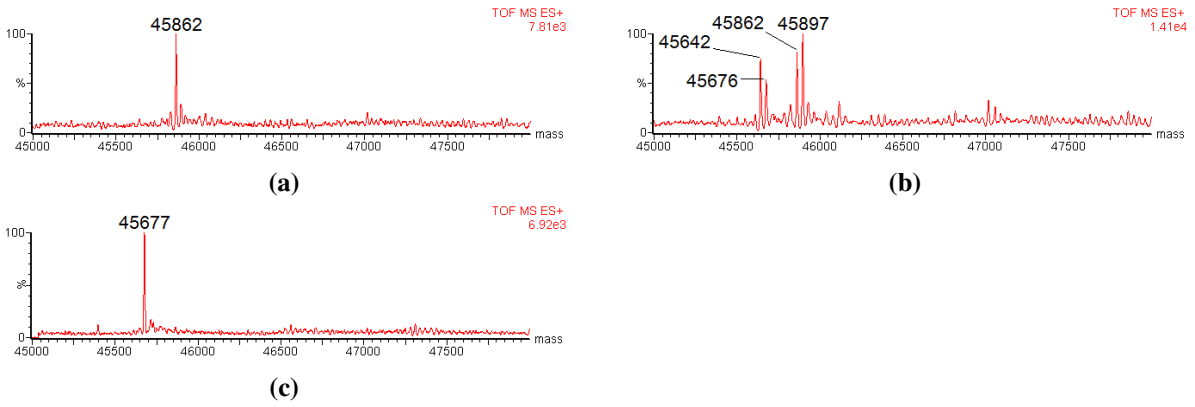


Fig. 4.5 Deconvoluted MS showing MEK1-Cys222 after treatment with dibromide **1** after: **(a)** 6 h. Only the peak corresponding to single Dha product with single bromide adduct remains. **(b)** 2 h. Peaks from all intermediate species and products were observed. In order of increasing mass: i. single Dha product, ii. unmodified MEK1-Cys222, iii. single Dha product with single bromide adduct, iv. single bromide adduct. **(c)** Unmodified MEK1-Cys222 (reproduced from Fig. 4.2c for convenience of comparison). Expected mass change from unmodified MEK1 variant upon: Dha formation: −34 Da, dibromide **1** mono-alkylation: +221 Da.

Regarding protein stability, this initial trial of Dha incorporation also showed MEK1 to be less tolerant to the reaction conditions (37 °C, 600 rpm mixing speed, 150 eq. dibromide **1**, 4–6 h reaction time) than p38α kinase. Whereas there were no visible signs of protein loss in the reactions with p38α, MEK1 showed progressive formation of filament-like precipitate in the reaction solution, which was postulated to be caused by protein aggregation. The hypothesis of protein precipitation was confirmed by measuring protein concentration in the solution phase of the reaction mixture before and after the reaction by UV absorbance, where it was found that roughly half the protein had been lost from solution upon reaction.

A number of major obstacles thus needed to be overcome for successful Dha formation at Cys222:

1. Over-alkylation beyond single Dha formation had to be prevented so that the desired chemical variant would be as pure as possible.

2. Protein precipitation had to be prevented to a level which ensured that there was sufficient material at the end of the synthesis for functional and enzymatic studies.

3. Reaction rate had to be increased.

These became the focal points of efforts for reaction optimisation of Dha formation on MEK1-Cys222.

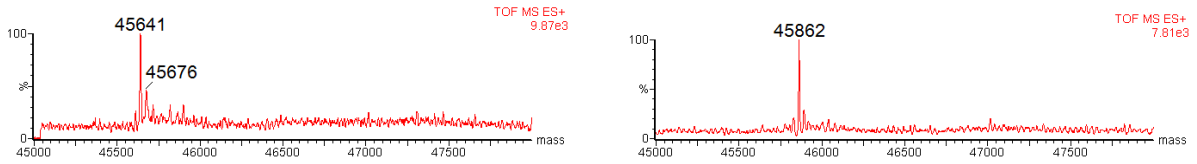
4.5.1 Substrate Protection

Monitoring of the initial trial of Dha formation by LC-MS showed that the molecular mass of products present did not evolve further from **25** (Scheme 4.2) over time, which in turn showed that the second, mono-alkylation (referred to as the "over-alkylation") was stable. If the over-alkylated state were not to be stable, it would have been expected that the over-alkylation would bis-alkylate the already attacked cysteine thiol to form the sulphonium, or eliminate completely to give another Dha, neither of which were observed. Given the close proximity of the remaining bromide to the cysteine thiol after over-alkylation, subsequent bis-alkylation and elimination would be expected to be a facile process. Thus, the stability of **25** suggested the over-alkylation was being actively stabilised by the protein's structure, which could occur if the reaction had occurred at a hindered cysteine. The hindrance of this cysteine would have to be enough to prevent the over-alkylation product from cyclising or eliminating, but would not be able to prevent over-alkylation completely. The presence of cysteines in MEK1-Cys222 of this steric profile correlated well with the observations, with the most likely candidate being Cys121 or Cys207.

The proximity of Cys121 and Cys207 to the bottom of the ATP binding cleft raised the possibility that if access to the cleft could be further restricted, over-alkylation also might be restricted. Thus, the principle of an enzyme inhibitor was tested. An inhibitor in a pharmaceutical capacity would normally block an enzyme's substrate from binding the active site, thus preventing enzymatic activity. In this case, the principle would be tested for whether access to that active site was blocked in a more general sense.

Being widely available, ATP was therefore used as the "inhibitor". Its inclusion in the reaction buffer did indeed block over-alkylation of MEK1-Cys222 (Fig. 4.6a). This modulation of reaction behaviour further supported the hypothesis that the over-alkylation occurred on one of the hindered cysteines in proximity to the ATP binding cleft. However, the selectivity modulation also slowed the rate of chemical reaction overall, such that at the arbitrary endpoint of 6 h, full conversion to the final product (over-alkylated MEK1-Cys222) was obtained when ATP was not present, but only ∼70% conversion to the final product (MEK1-Dha222) was achieved in the presence of ATP (Fig. 4.6). Although the design principles of using ATP as a "substrate protector" had been a logical step from observations of the crystal structure and preliminary chemical modification, the technique had also been used previously in other studies.[38] The literature study had also used magnesium salt in the reaction buffer, presumably further stabilising the

binding of the ATP onto MEK1, but the current study showed that this precaution was unnecessary, since protection still occurred without it.



(a) **(b)**

Fig. 4.6 Deconvoluted MS showing MEK1-Cys222 after treatment with dibromide **1** for 6 h: **(a)** in the presence of ATP. The MS shows incomplete conversion (∼70%) to the desired Dha product (**24**). **(b)** without ATP present. Full conversion to over-alkylated MEK1-Dha222 (**25**). (MS reproduced from Fig. 4.5a for convenience of comparison). Expected mass change from unmodified MEK1-Cys222 upon: Dha incorporation: −34 Da, dibromide **1** mono-alkylation: +221 Da.

4.5.2 Varying Physiochemical Conditions

With a viable solution for preventing over-alkylation on MEK1-Cys222, attention was turned to increasing the rate of the bis-addition–elimination reaction, while still looking at any side-effects on reaction selectivity. The simplest solution is to increase the amount of dibromide **1** used. This led to an increase in reaction rate (Fig. 4.7), but also increased the amount of precipitation in the reaction mixture. Most of this precipitate, with its granular appearance looked to be due to the crystallisation of excess **1**, but its presence in the mixture was not desired, since it could seed the aggregation of the MEK1 species and cause increased occurrence of protein precipitation. The solubility level of **1** was therefore estimated to determine the range of dibromide **1** concentrations within which precipitate would remain. Monitoring OD_{600} at different dibromide **1** concentrations showed that **1** was soluble to ≈1.0 mg/mL: around the concentration used in a reaction with 150 eq. of **1** (Fig. 4.8). Thus, given the narrow scope in raising the concentration of **1** without causing its precipitation, the increase in **1** was not more widely adopted.

Another common physiochemical parameter to change is the buffer. MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol) had been used up to this point, which like in the case of p38α reaction buffer had been chosen to help with protein stability (Section 2.3). Changes in buffer composition had already been shown not to make significant changes to the reaction rate (Section 2.8.3) and the use of phosphate buffer[43] also did not make a large change in rate. In this case, the reaction was slower in phosphate buffer than in MEK1 reaction buffer (Fig. 4.9). As MEK1 is able to bind ATP, which contains phosphates, the kinase might also be expected to interact with inorganic phosphate. Any such

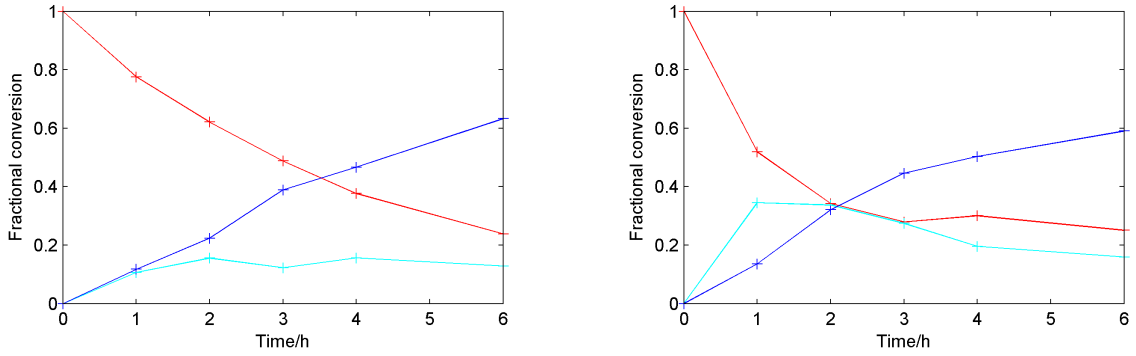


Fig. 4.7 Kinetic analysis of conversion of unmodified MEK1-Cys222 (red) to Dha product (blue) upon treatment with different quantities of dibromide **1**: **(a)** with 150 eq. and **(b)** with 500 eq. At the test end-point of 6 h, there was a higher conversion to Dha product in **(b)** than in **(a)**. The reaction goes *via* a mono-alkylated bromide adduct (cyan), which is a long-lasting intermediate. Substrate protection with ATP was used in these reactions.

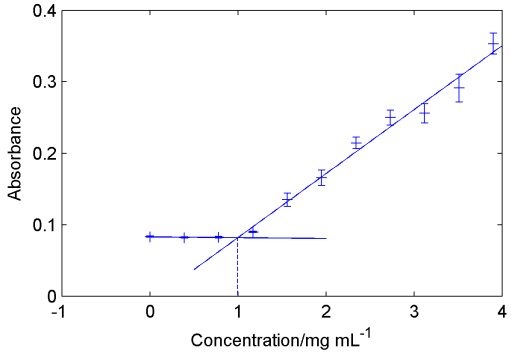


Fig. 4.8 Plot of optical density measurements (OD_{600}) made to determine the solubility of dibromide **1** in MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol). At low dibromide **1** concentration, a background OD_{600} reading is given, resulting in the horizontal line in the plot. Upon increasing concentration past the saturation point, the increasing amounts of precipitate cause a directly proportional increase in OD_{600}. The intersection of the two lines therefore represents the solubility level of dibromide **1** in the buffer.

interaction did not modulate the outcome of the Dha incorporation; over-alkylation still occurred when phosphate buffer without added ATP was used (Fig. 4.10).

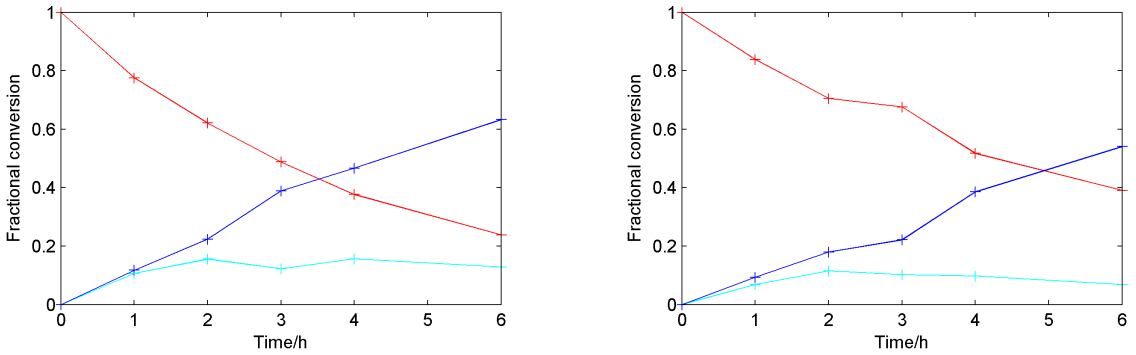


Fig. 4.9 Kinetic analysis of conversion of unmodified MEK1-Cys222 (red) to Dha product (blue) upon treatment dibromide **1** under different buffer conditions: **(a)** MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol) (reproduced from Fig. 4.7a for convenience of comparison) and **(b)** phosphate buffer (50 mM sodium phosphate pH 8.0). At the test endpoint of 6 h, there was a higher conversion to Dha product in **(a)** than in **(b)**. The reaction goes *via* a mono-alkylated bromide adduct intermediate (cyan). Substrate protection with ATP was used in these reactions.

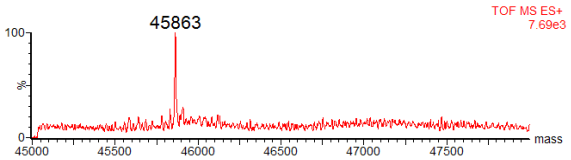


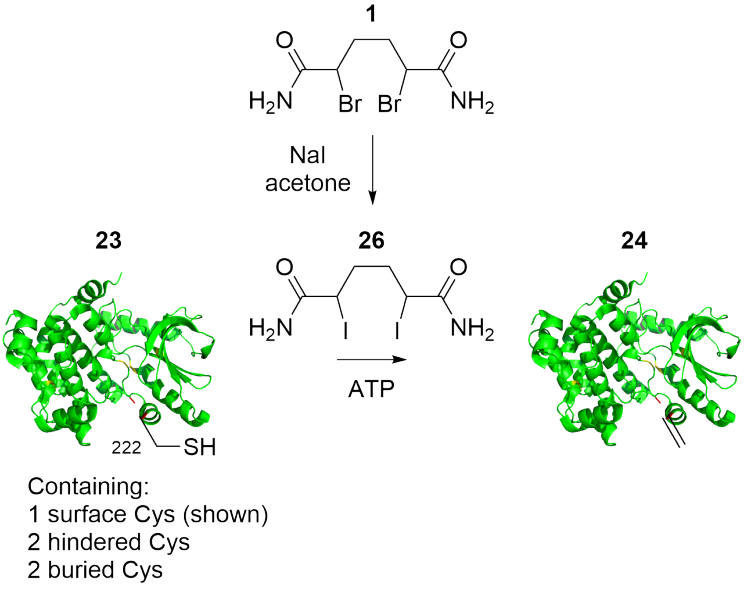
Fig. 4.10 Deconvoluted MS of final product after treatment of MEK1-Cys222 with dibromide **1** in phosphate buffer with no ATP. Only the over-alkylated MEK1-Dha222 **25** was present (calculated mass: 45 863, observed mass: 45 863).

4.5.3 Iodide Reactions

With the intention still to increase the rate of Dha formation, the chemistry of Dha formation was next scrutinised. Observations in the MS of the bis-addition–elimination reaction suggest the elimination step is fast on MEK1-Cys222 kinase, since the sulphonium intermediate is rarely detected. The notation that the addition reaction is slower and rate-determining was supported by the routine detection of the mono-alkylated intermediate. To make the addition step faster by modifying the reagents, either the thiol of cysteine would have to be made a stronger nucleophile, or the bromide of dibromide **1** would have to be changed to a better leaving group.

Following the periodic trend down the chalcogens, cysteine could be made into a better nucleophile if selenocysteine (or derivative thereof) were to be genetically incorporated instead.[128,129] As genetic incorpo-

ration would require the use of codon reassignment technology,[130] use of selenocysteine was not adopted since it would negate the advantage of relatively high expression levels achieved by using only natural amino acids. Following a similar periodic trend down the halogens, the bromides of dibromide **1** could be made into better leaving groups by changing them to iodides. Thus, modification of the bis-alkylation–elimination reagent was adopted as this was also a less laborious undertaking than re-expression with codon reassignment, and therefore was more preferable.



Scheme 4.3 Dha formation on MEK1-Cys222 using diiodide **26** as the bis-alkylation–elimination reagent.

Halogen exchange on dibromide **1** would furnish the desired diiodide (**26**, Scheme 4.3). Two methods of halogen exchange were tried:

1. encourage the formation of diiodide **26** from dibromide **1** *in situ*, with the protein present in its aqueous environment.

2. isolate the synthesised diiodide **26** before reaction with MEK1-Cys222.

in situ formation of diiodide **26** was attempted by the inclusion of sodium iodide into the reaction buffer (Fig. 4.11a), with the trial being undertaken on p38α-Cys172 as a model protein. The associated control reaction (Fig. 4.11b) had additional sodium chloride added to the standard p38α reaction buffer to keep the ionic strength consistent between the two experiments. No significant difference in reaction rate was observed between the two reactions.

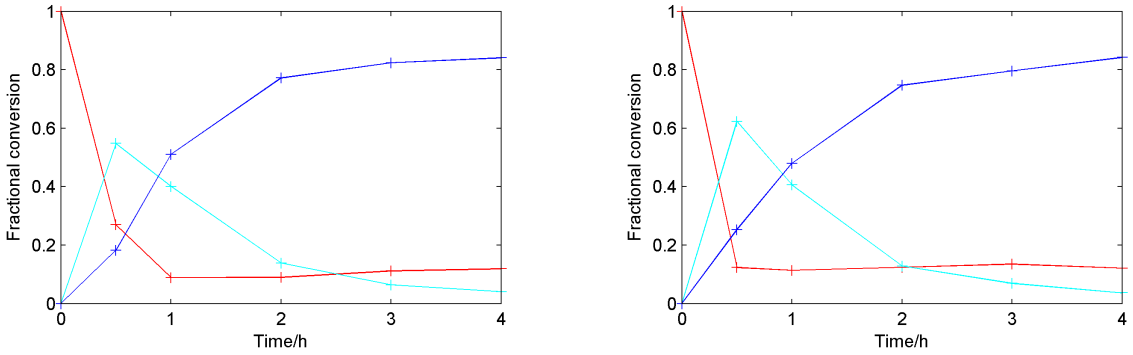


Fig. 4.11 Kinetic analysis of conversion of unmodified p38α-Cys172 (red) to Dha product (blue) upon treatment dibromide **1**: **(a)** with NaI (100 mM) in the buffer **(b)** with additional NaCl (100 mM) in the buffer. The reaction goes *via* a mono-alkylated bromide adduct intermediate (cyan).

Use of isolated diiodide **26** also did not improve on the rate of Dha incorporation already achieved. It has also been postulated that the increased size of iodide over bromide might make **26** more susceptible to regioselective reaction, so MEK1-Cys222 was used in this trial. However, treatment of MEK1-Cys222 without the presence of ATP still resulted in over-alkylation (Fig. 4.12). As in the case with dibromide **1** when no ATP was used (Section 4.5), the over-alkylation occurred when unreacted MEK1-Cys222 was still present, preventing the possibility of stopping the reaction after the complete mono-addition of MEK1-Cys222. With ATP included in the reaction buffer again, the reaction with diiodide **26** was in fact much slower than with dibromide **1** (Fig. 4.13). In comparison to dibromide **1**, the slower reaction was likely to have been due to a combination of poorer aqueous solubility of diiodide **26**, and the extra steric hindrance between the ATP and the dihalogeno-species, due to the larger halogen atoms.

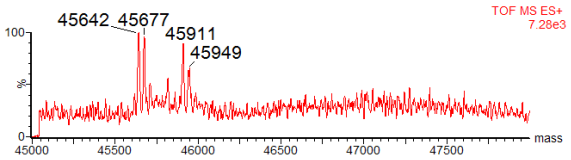
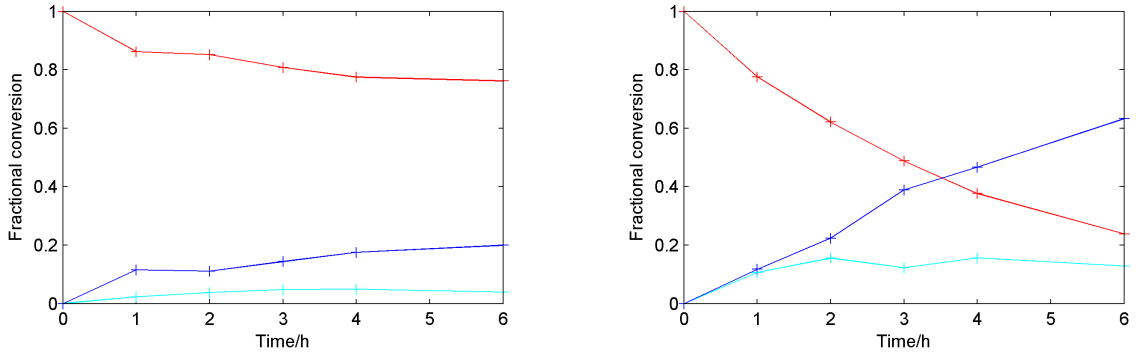


Fig. 4.12 Deconvoluted MS after 4 h of treatment of MEK1-Cys222 with diiodide **27** and no ATP present. MS peaks in order of increasing mass: i. single Dha product, ii. unmodified MEK1-Cys222 (no mass label above), iii. single Dha product with single iodide adduct, iv. single iodide adduct. Expected mass change from unmodified MEK1 variant upon: Dha formation: -34 Da, diiodide **26** mono-alkylation: $+268$ Da.

4.5.4 Smaller Substrate Protecting Agents

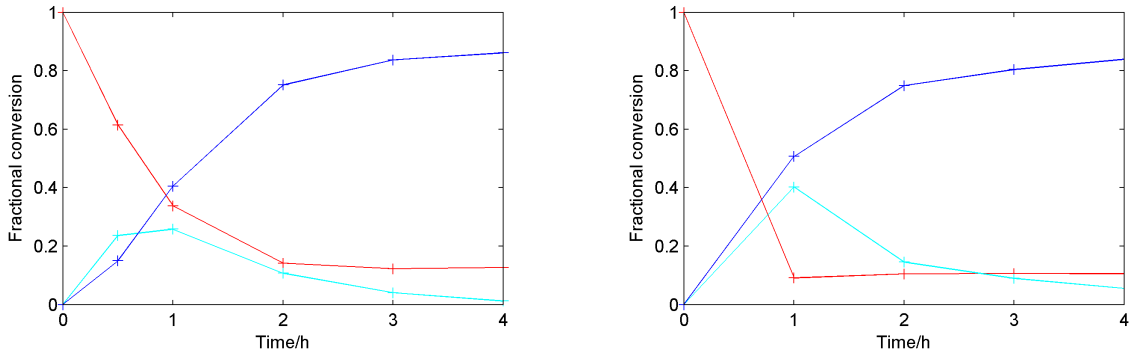
The slowness of diiodide **26** in Dha formation on MEK1-Cys222 was rather unexpected, particularly as no such significant decrease in rate had been observed for the corresponding comparison of dibromide



(a) **(b)**

Fig. 4.13 Kinetic analysis of conversion of unmodified MEK1-Cys222 (red) to Dha product (blue) upon treatment with: **(a)** diiodide **26** **(b)** dibromide **1** (reproduced from Fig. 4.7a for convenience of comparison). The reaction goes *via* a mono-alkylated adduct intermediate (cyan).

1 against diiodide **26** on p38α-Cys172. In the case of p38α-Cys172, both reagents were observed to have roughly the same rate (Fig. 4.14). Given that dibromide **1** and diiodide **26** only differ by the halide leaving group and therefore do not vary greatly in polarity, and given the similarity in reaction behaviour against MEK1-Cys222 with and without ATP being present, it is likely that both **1** and **26** have the same selectivity for Cys in the different steric and electronic environments around the protein. This therefore suggested that although the low solubility of the **26** may have contributed to a slower reaction rate on MEK1-Cys222, its interference with ATP was likely to be the major factor in the dramatic rate decrease.



(a) **(b)**

Fig. 4.14 Kinetic analysis of conversion of unmodified p38α-Cys172 (red) to Dha product (blue) upon treatment with: **(a)** diiodide **26** **(b)** dibromide **1** (reproduced from Fig. 2.11b for convenience of comparison). The reaction goes *via* a mono-alkylated adduct intermediate (cyan).

The additional interference between the ATP substrate and the bis-addition–elimination reagent diiodide **26** highlighted the possibility that this interaction could be influencing the rate of Dha incorporation using the corresponding dibromide **1**, but to a lesser extent. Thus, if the interference could be reduced by reducing the size of the ATP substrate to a fragment, an increase in Dha incorporation rate might be observed.

Looking at ATP analogues bound into the crystal structure[5] showed that it was the phosphate of the ATP that was in closest proximity to the activation loop, and therefore to Cys222 (Fig. 4.15). It was thus postulated that if a substrate protector with fewer phosphates were used, the rate of Dha incorporation might be increased. ADP and AMP were therefore tried and although they were both successful in preventing over-alkylation, the rate of Dha incorporation using these shorter substrate protectors was similar to that of ATP. There was therefore no advantage to reaction rate by using these truncated adenosine phosphates. However, it did show that AMP is sufficient for having the substrate protection effect.

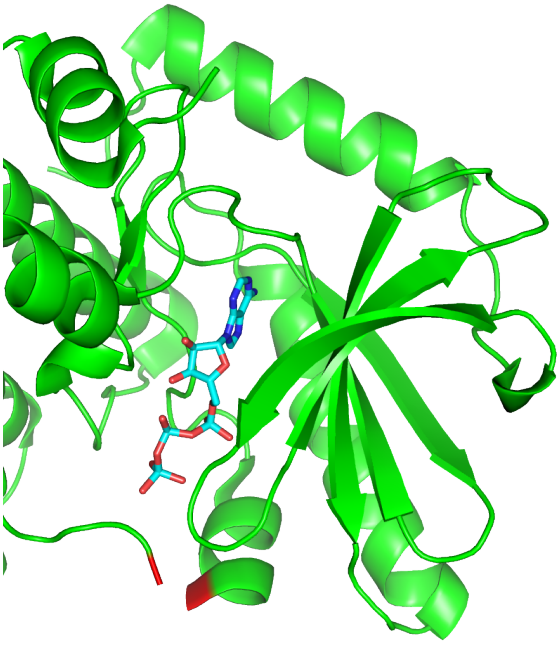


Fig. 4.15 Crystal structure of MEK1 kinase with ATP (cyan, coloured by element with phosphate oxygen in red) bound showing the close proximity of the phosphates to Cys222 (red, spanning the gap adjacent to the α-helix).

4.5.5 Protein Stabilising Reagents

Failing to achieve a workable solution to increase the rate of Dha formation, it was decided instead to try and find conditions that would better stabilise the protein, as set out earlier in Section 4.5. A previous chemical modification study on MEK1[38] had explicitly stated the importance of using stabilising agents to prevent protein aggregation, with the non-ionic detergent Brij-35 being used. A range of non-ionic detergents (Brij-35, Triton X-100 and TWEEN-20) were therefore screened, along with other common protein stabilising agents such as PEG,[131] glycerol[76] and arginine.[132]

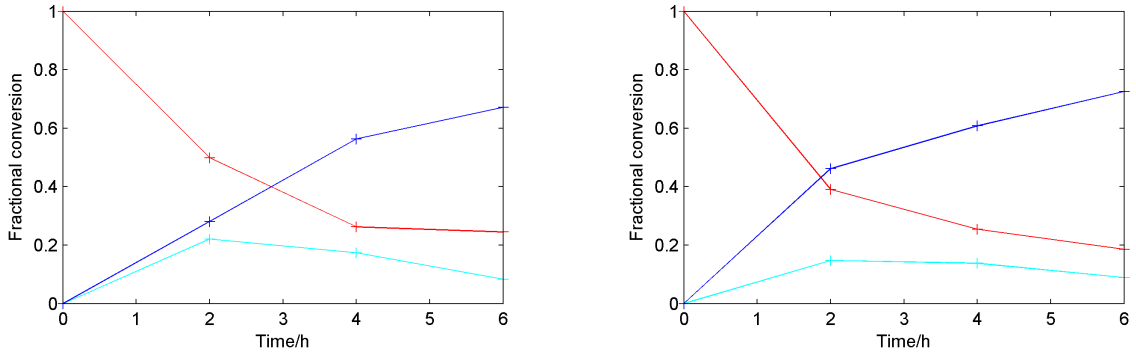


Fig. 4.16 Kinetic analysis of conversion of unmodified MEK1-Cys222 (red) to Dha product (blue) upon treatment with dibromide **1** in the presence of varying amounts of protein stabilising agents: **(a)** with MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol) and **(b)** with additional glycerol in the buffer (final concentration: 10%). The reaction goes *via* a mono-alkylated bromide adduct (cyan), which is a long-lasting intermediate. Substrate protection with ATP was used in these reactions.

Taking the effects of the buffer additives on reaction conversion and protein precipitation into consideration, it was found that there was little difference in the reaction conversion between the use of these agents (Fig. 4.16), but there were marked differences in the quality of the spectra. This difference in quality was reflected in SDS-PAGE analysis (Fig. 4.17). In this analysis, the supernatant liquid and precipitate were separated and analysed separately, where it was observed that samples which gave poorer quality spectra also generally gave smaller spot intensities for the soluble protein fraction. In quantifying this data by gel densitometry, ratios between spot intensities were taken between corresponding supernatant and precipitate spots, which revelled that the addition of extra glycerol resulted in the least protein precipitation (Table 4.1). Surprisingly, the addition of Brij-35 to the reaction buffer caused more protein precipitation than without it during these experiments. This result from gel densitometry was confirmed quantitatively by analogous measurements made using BSA assay protein concentration determination, as well as from visual inspection of the samples, where the reaction containing extra buffer glycerol gave a smaller pellet upon centrifugation.

Allied to this buffer additive study were modifications made to the LC-MS analysis technique. Improvements made to the LC-MS method during the chemical modification studies on p38α had been optimised for speed (Section 2.8.1) so that reactions could be monitored better in real time, thus sacrificing mixture separation. The inclusion of detergents in the reaction buffer caused these to co-elute with the protein species upon liquid chromatography. As the detergents form micelles at the working concentrations, this suggests that the micelles have similar size and charge properties to the protein. Their co-elution made

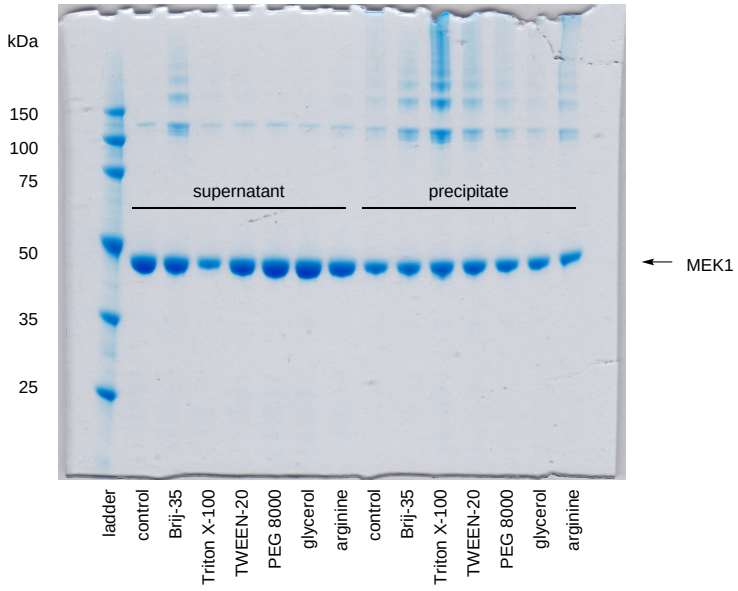


Fig. 4.17 SDS-PAGE gel of supernatant and precipitate fractions of reaction mixtures from a trial of protein stabilisation additives.

Table 4.1 Protein concentration ratios from protein stabilising additive screen.

	Supernatant:Precipitate Ratio	
Additive	**Densitometry**	**BSA Assay**
Control	1.90	1.45
Brij-35	1.58	1.33
Triton X-100	0.72	0.20
TWEEN-20	1.44	0.88
PEG 8000	2.08	1.26
Extra glycerol (10%)	2.32	1.70
Arginine	2.04	0.78

the analysis of the resultant spectra difficult, thus re-optimisation of the LC-MS method was necessary. A change in the column chemistry (from a polystyrene-divinylbenzene polymer monolith in the Dionex column to a C-18 silica monolith in Chromolith® FastGradient RP-18 endcapped 50×2 mm monolithic HPLC column from Merck) and a lengthening of the linear gradient allowed for better peak definition and consistent separation of protein from detergent. The use of this new LC method allowed for better quality spectra to be collected, and subsequently more consistent MS analysis. The chromatographic separation method described here was also the method of choice for the enzyme kinetic studies described in Chapters 3 and 5, since the improved separation also allowed for some separation of substrate and enzyme protein species, on top of any detergent buffer components.

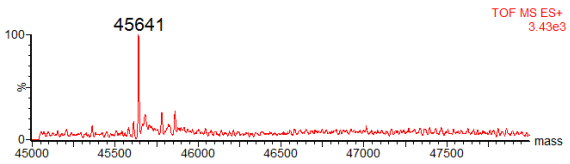
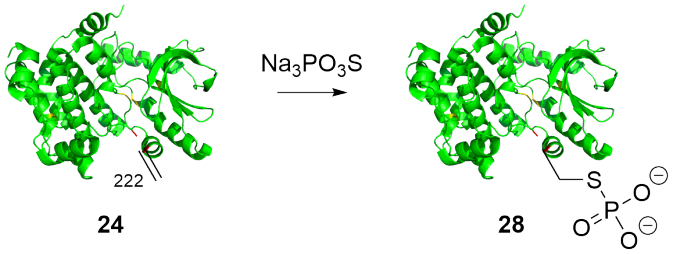


Fig. 4.18 Deconvoluted MS showing >85% conversion to the desired product MEK1-Dha222 (calculated mass: 45 642, observed mass: 45 641).

Since glycerol reduced the precipitation, it was duly employed for general use in Dha formation on MEK1. However, it had not completely eliminated protein precipitation. It had previously been assumed that the presence of precipitate of non-protein origin could seed the aggregation of the protein in solution (Section 4.5.2). It was also observed that upon buffer exchange and desalting of the protein before the reaction, fine particles, assumed to be resin from the desalting column used, were present in the sample. Putting the assumption and observations together, the protein was therefore filtered again, after the buffer exchange step, but before the addition of other reagents for the chemical reaction (Dha formation). Taking this precaution resulted in no observable filamentous precipitate from being formed, and had thus further prevented the protein from precipitating. Having sufficiently stabilised the protein in solution, increase in the reaction time could be afforded (to 8 h), finally resulting in >85% conversion of MEK1-Cys222 to MEK1-Dha222 (Fig. 4.18).

4.6 Installation of Thiophosphate on MEK1-Cys222

With the achievement of high conversion and selectivity for Dha formation on MEK1-Cys222, the subsequent step of installing phosphocysteine to generate the final phosphate mimic could be investigated (Scheme 4.4). Similar conditions to those used on p38α were initially tried (37 °C, 600 rpm mixing



Scheme 4.4 Addition of thiophosphate onto MEK1-Dha222.

speed, 30 000 eq. sodium thiophosphate, 4-1–4 h reaction time, Section 2.8 and 2.11.4). Analysis of samples by MS could not detect the presence of protein, despite having a signal in the total ion current (TIC) chromatogram. Degradation of the protein was suspected, particularly as the reaction mixture became increasing viscoelastic over time. Decreasing the amount of sodium thiophosphate used in the reaction did not significantly alter this outcome of the reaction mixture, suggesting that the degradation effect was only partially due to the amount of thiophosphate used.

The other likely candidate parameter to cause degradation was temperature. Further trials were therefore undertaken at room temperature with varying quantities of thiophosphate (in collaboration with Sébastien Galan). Under these conditions, it was found that the protein degraded less quickly, thus allowing the protein to be detected by MS. It could then be found that 20 000 eq. was the optimum amount of sodium thiophosphate required for the reaction, with any more than this causing degradation, characterised by a lack of MS signal. Being the first time that the formation of MEK1-pCys222 had been observed, this represented the starting point for further optimisation. From this starting point, two parameters were further tested:

1. Reaction temperature: The studies around temperature had been rather crude up until this point. A more rigorous analysis was therefore required. 37 °C and room temperature were again used.

2. Substrate protection: Given that the reaction itself or the reaction conditions as a whole seemed to be causing the protein degradation, substrate protection could be used as a scaffold for the protein, in a similar way that inhibitor binding generally causes an increase in melting temperature of an enzyme. Protein degradation was to be avoided as it could manifest in denaturation, leading to enzymatic inactivity. As in the Dha incorporation step, ATP was used.

The scaffolding effect of ATP had inadvertently been demonstrated earlier, during the trials of smaller substrate protecting agents for use in the Dha forming reaction (Section 4.5.4). The earlier trials had

focused on preventing dibromide **1** from causing over-alkylation on MEK1-Cys222, overlooking the other effects these substrate derivatives might have. Running samples of the supernatant liquid and precipitate fractions from reactions with different substrate protectors, and with no protector showed that more protein was retained in solution when protector was present (Fig. 4.19). The experiment without protector had a smaller spot for the supernatant liquid compared to the other conditions, an observation which was supported by densitometry (Table 4.2).

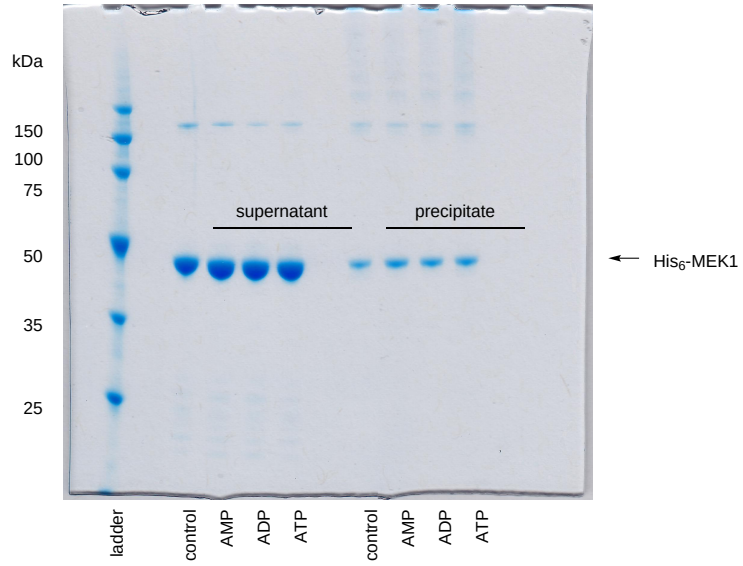


Fig. 4.19 SDS-PAGE gel of supernatant and precipitate fractions from a trial for MEK1-Dha222 formation. A range of substrate protectors were used.

Table 4.2 Protein concentration ratios from a trial for MEK1-Dha222 formation. A range of substrate protectors were used.

Additive	SN:ppt Ratio
Control	1.90
AMP	2.04
ADP	2.10
ATP	2.05

Success of the reactions conditions under trial for thiophosphate installation was determined by conversion rate and protein stability, once again optimising both the chemistry (reaction conversion) and the biology (protein stability)of the reaction, as with previous optimisation exercises. Using 20 000 eq. sodium thiophosphate in all reactions, the reactions run at 37 °C showed that the protein was stable up to ∼2 h, after which the total MS signal started to weaken and was thus taken as a sign of protein degradation. The weakening of the MS signal was characterised by an increase in the spectrum baseline level, with the spectrum becoming more noisy with a rise in the baseline. Associated with this signal weakening was the

physical observation that the consistency of the reaction mixture became more viscoelastic over time. In comparing the reaction with and without ATP, it was observed that the MS signal weakened more slowly when ATP was in the reaction mixture, but with the side-effect that reaction conversion was lower with ATP than without it at all the timepoints measured, resulting in the lower overall conversion (Table 4.3, Entries 1 and 2).

Table 4.3 Summary of the conditions investigated in the optimisation of thiophosphate incorporation on MEK1-Dha222.

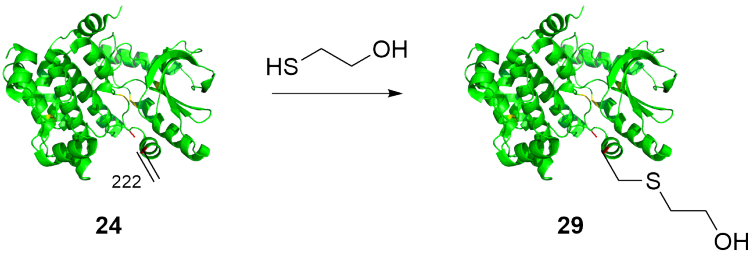
Entry	ATP	Time/h	Temp.	Conversion
1	✗	2	37 °C	>65%
2	✓	2	37 °C	55%
3	✗	8	RT	70%
4	✓	8	RT	50%
5	✗	18	RT	>70%
6	✓	18	RT	>85%

As expected, the rate of reaction decreased upon lowering the reaction temperature (Table 4.3, Entries 3–6) but surprisingly, the protein appeared disproportionately more stable under these conditions, such that even after 24 h, the protein MS signal remained clearly distinguishable and the viscoelasticity of the reaction mixture remained similar to the beginning, in all cases. The MS signal was weaker after ∼18 h, suggesting that the protein was starting to degrade at this stage, but not to a level visible on the macroscopic scale to cause a change in solution viscoelasticity. As before at 37 °C, the reaction with ATP was slower than without, but the MS were generally less noisy for the reaction with ATP, which again showed the stabilising effect of substrate scaffolding. The final, optimised conditions were chosen as 18 h reaction time at room temperature using ATP in the reaction buffer, which gave >85% conversion to MEK1-pCys222, thus producing the first MEK1 phospho-mimic, as determined by whole protein MS.

4.7 Characterisation of Chemical Variants Derived from MEK1-Cys222

As in the case of p38α (Section 2.9), further evidence of the chemical functionalisation of MEK1-Cys222 at position 222, and of the structural integrity of the chemical variants were needed. To do this, MS/MS analysis and CD spectroscopy were again used respectively. MS/MS analysis (in collaboration with Ritu Raj) was able to identify the Dha-containing peptide in MEK1-Dha222, which showed that Dha had indeed been installed at position 222 (Section 4.11.4). In order to increase the chances of peptide detection, the data-dependent decision tree (DDDT) methods used previously for p38α (Section 2.9.5) were also

used here. Installation of Dha at 222 had already been suspected, given the levels of interference to Dha incorporation when diiodide **26** was used, compared to dibromide **1**. This observation of interference would only be likely if both reagents reacted with the cysteine that was both the closest in proximity to the ATP binding site, and yet not fully blocked. Cys222 was the most likely candidate that fits both of these descriptions.



24 **29**

Scheme 4.5 Testing the reactivity of the incorporated Dha with β-mercaptoethanol.

Conversely, the installation of phosphocysteine at 222 could not be directly shown through finding the corresponding cysteine-containing peptide in MEK1-pCys222, where no such peptide could be detected. Despite the precautions taken with the MS/MS fragmentation mode with DDDT being used, the labile nature of the phosphate during MS/MS is likely to have still contributed to the lack of peptide detection. Instead, the reactivity of the installed Dha was independently challenged using β-mercaptoethanol (β-ME), where it was found that the single β-ME adduct (**29**) was formed (Scheme 4.5, in collaboration with Sébastien Galan). MS/MS analysis of **29** showed that the reaction had occurred at position 222, where the Dha had previously been installed (Section 4.11.4), demonstrating that the installed Dha was the most reactive group towards nucleophilic thiols. As thiophosphate is also such a nucleophile, it is therefore likely that phosphocysteine was also installed at position 222 to give MEK1-pCys222. This statement is further supported by the studies on enzymatic activity in Chapter 5.

Retention of protein structure is also key to enzymatic viability. This was investigated through CD spectroscopy, which found that all chemical variants derived from MEK1-Cys222 gave similar spectra (Fig. 4.20). However, further clues to structural retention could already be discerned prior to making the CD measurements, particularly from observations made during the sample preparation for MEK1-pCys222.

Before measuring the CD spectrum for a protein, the buffer should be well equilibrated.[133] In the context of the study described here, equilibration was achieved by extensive buffer exchange of all protein

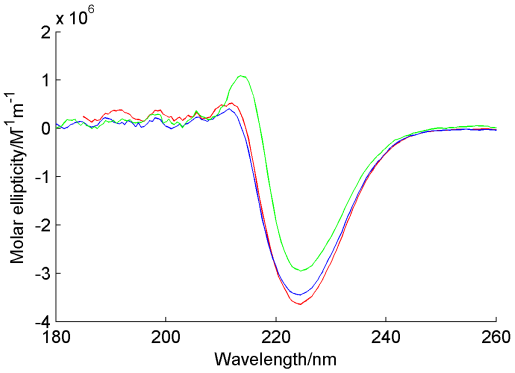


Fig. 4.20 CD spectra of all chemical variants derived from MEK1-Cys222. Key: MEK1-Cys222 (red), MEK1-Dha222 (blue), MEK1-pCys222 (green).

samples into buffer of the same composition. Buffer exchange could most easily done by repeated ultrafiltration/dilution, as was the case for MEK1-Cys222 and MEK1-Dha222. With MEK1-pCys222 however, ultrafiltration resulted in total loss of the protein, probably due to the protein adhering to the filtration membrane and was characterised by a lack of UV spectroscopic signal. Instead, buffer exchange had to be done by dialysis, which didn't result in this loss. UV spectroscopy of the dialysed material however did not reveal the spectrum typical of a MEK1 chemical variant such as MEK1-Cys222 or MEK1-Dha222, or of proteins in general with absorbance maximum ($\lambda_{\max}$) at 280 nm (Fig. 4.21a). Instead, the spectrum of the MEK1-pCys222 sample had $\lambda_{\max}$ at 260 nm (Fig. 4.21b). Another telling sign was that the absorbance intensity of the MEK1-pCys222 sample was suspiciously high, since the typical protein concentration used during chemical modification was ∼0.5–1 mg/mL. The sample of MEK1-pCys222 was also of this concentration and besides, this MEK1 chemical variant could not be concentrated by ultrafiltration to a level suggested by the sample's spectrum, since doing so would result in the loss of the protein. The high UV signal therefore had to be due to a contaminant from the reaction components used to form MEK1-pCys222, of which sodium thiophosphate and ATP were the components not common to the dialysis buffer. UV spectroscopy found ATP to have a spectrum matching that of the MEK1-pCys222 sample, again with $\lambda_{\max}$ at 260 nm (Fig. 4.21c).

A reason for the residual ATP was sort after. Lack of sufficient dialysis was ruled out since the sample of MEK1-pCys222 had been extensively dialysed, as determined from the lack of sodium thiophosphate. The lack of sodium thiophosphate in the sample had in turn been determined from the sample reacting with BCA reagent in the expected fashion; thiophosphate would quickly reduce the BCA reagent otherwise. As dialysis had been extensive, any ATP free in solution would also have been dialysed, along

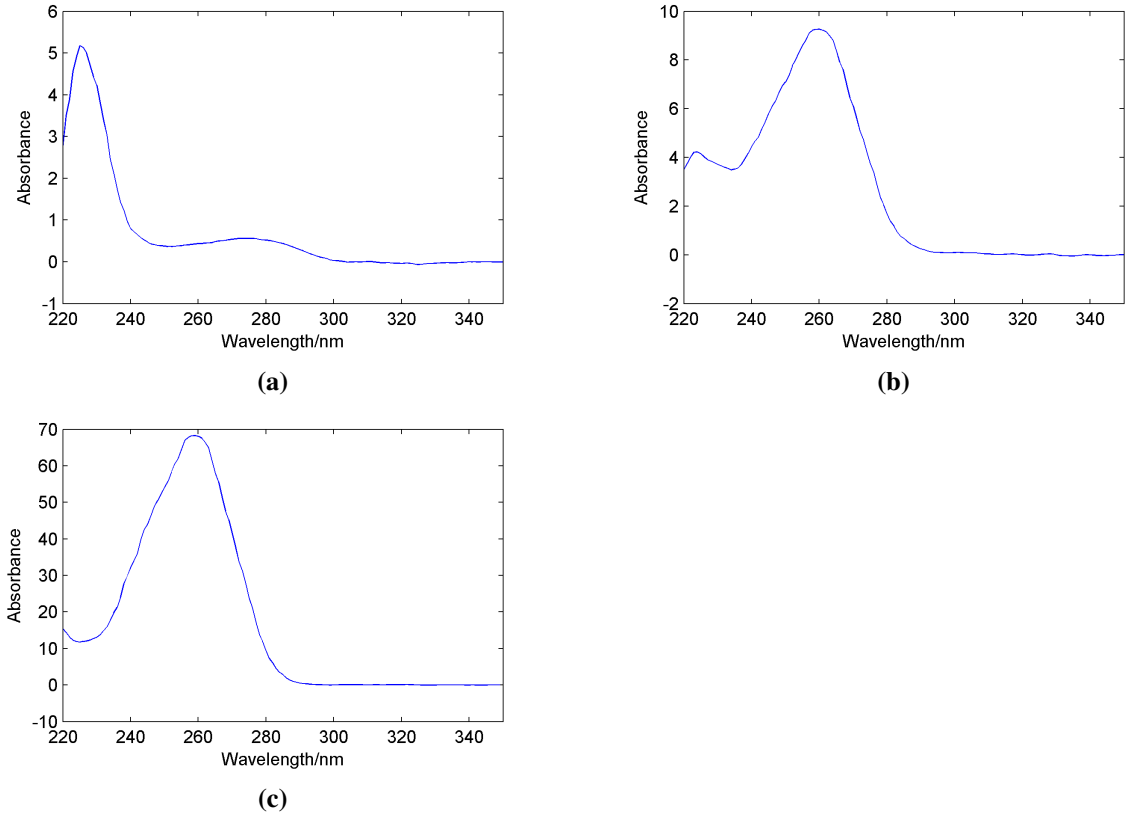


Fig. 4.21 **(a)** Typical UV spectrum of a MEK1 chemical variant, in this case MEK1-Dha222 ([MEK1-Dha222] = 1.00 mg/mL). **(b)** UV spectrum of MEK1-pCys222 sample ([MEK1-pCys222] = 0.46 mg/mL) showing the residual presence of ATP. **(c)** UV spectrum of ATP (5 mM).

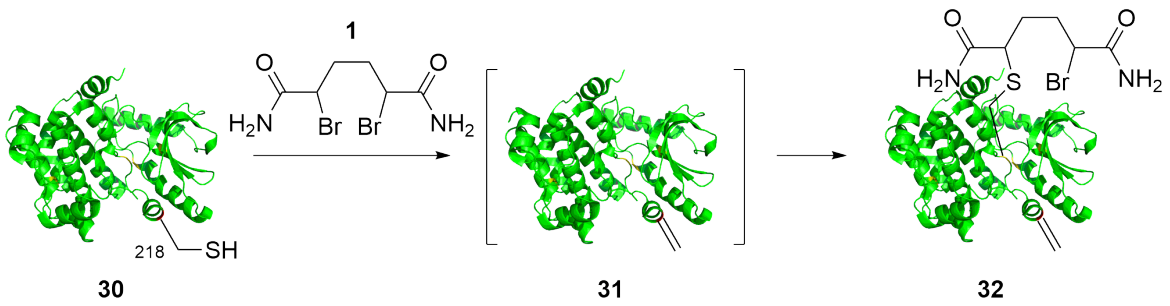
with the thiophosphate, so its appearance in the UV spectrum of the MEK1-pCys222 sample was surprising. The most likely mechanism for the ATP to be carried over into the sample is if it were bound to MEK1-pCys222. ATP is not known to non-specifically bind to protein. The MEK1-pCys222 structure must therefore have been intact, so that it retained its ability to bind ATP in the binding cleft. However, the reason for the apparently super-stoichiometric amount of ATP present, relative to the protein concentration in the MEK1-pCys222 sample is unclear.

With regards to CD experimental design, the temperature melt procedure used on p38α (Section 2.9.1) was not used in this case, as it had been of limited value in determining melting temperature. The melting temperature was however measured by differential scanning fluorimetry (DSF) in the original MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, i.e. no added glycerol) and found to begin at 34.0 °C. This would explain the apparent thermal intolerance of MEK1-Cys222 and its requirement for glycerol and ATP as stabilisers to keep it in solution.

In summary, MS/MS analysis and CD spectroscopy gave further evidence that MEK1-pCys222 had been formed as intended as a mimic of MEK1-pSer222, and was ready to be tested for its biological function.

4.8 Dha Incorporation Studies on MEK1-Cys218

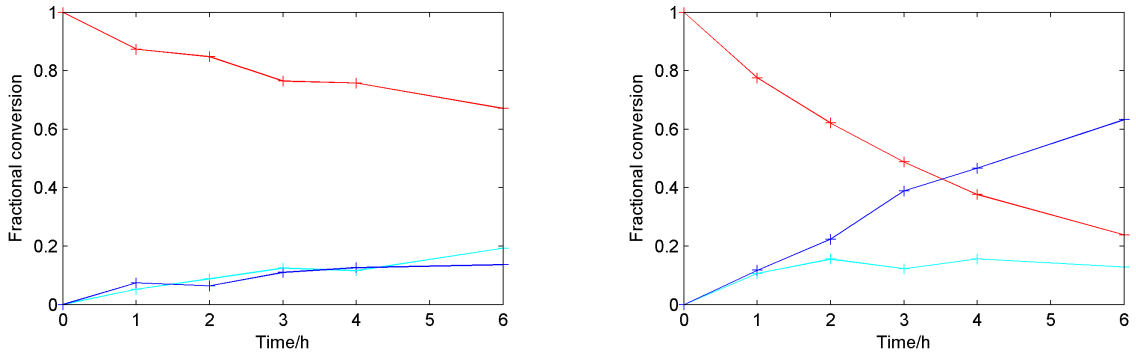
As had been mentioned at the beginning of this chapter, MEK1 has two native phosphorylation sites where phosphorylation contributes to the kinase's activation (Section 4.1), the other site being Ser218. In investigating this site by chemical modification, the plasmid construct containing the gene for the MEK1 variant MEK1-G(−19)F/S218C/C277S/C376S (referred to as MEK1-Cys218) was also generated, and the associated protein expressed and purified using the already optimised procedures. The protein was then subjected to trials of Dha formation, similar to the studies described for MEK1-Cys222.



Scheme 4.6 Dha formation with dibromide **1** on MEK1-Cys218 **30** resulted in over-alkylation of the kinase.

Like MEK1-Cys222, MEK1-Cys218 was initially subjected to dibromide **1** for attempted Dha formation and like with MEK1-Cys222, over-alkylated MEK1-Dha218 (**32**) was observed (Scheme 4.6 and Section 4.11.4). However, rather than just suppressing the over-alkylation when ATP was introduced to the reaction conditions, as was the case with MEK1-Cys222, the introduction of ATP strongly inhibited any reaction of dibromide **1** from occurring with MEK1-Cys218. Neither MEK1-Dha218 (**31**), nor its mono-alkylated precursor intermediate were observed in significant quantities in the presence of ATP (Fig. 4.22). Although obviously an undesired outcome, the lack of reaction on MEK1-Cys218 under these conditions was not completely unexpected, since it had already been observed from the crystal structure that the protein surface around position 222 was more accessible and flexible than around position 218 (Section 4.1). As explained previously for p38α (Section 2.8.2), the lack of reaction on MEK1-Cys218 also provided further evidence that phosphocysteine had indeed been installed at position 222 on MEK1-Cys222, since the only difference between the two MEK1 variants is in the designated reaction site. The change in

reaction rate must therefore be associated with this change in reaction position. By extension, this would mean that the Dha in over-alkylated MEK1-Dha218 was indeed being installed at position 218.



(a) **(b)**

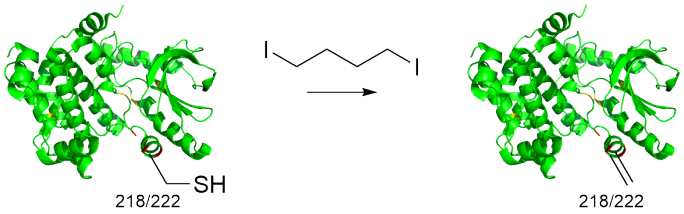
Fig. 4.22 Kinetic analysis of conversion of unmodified MEK1 variant (red) to its corresponding Dha product (blue) upon treatment with dibromide **1** in the presence of ATP (with substrate protection). Different MEK1 variants were subjected to this treatment: **(a)** MEK1-Cys218 and **(b)** MEK1-Cys222 (reproduced from Fig. 4.7a for convenience of comparison). The reaction goes *via* a mono-alkylated adduct intermediate (cyan).

The preparation of MEK1-Dha218 was obviously the desired outcome for further studies on MEK1-Cys218 chemical variants. To achieve this, the strategies that had previously been tried on MEK1-Cys222 in attempts to increase the reaction rate were applied to MEK1-Cys218. The aim with MEK1-Cys218 was to try and produce MEK1-Dha218 without any off-target modification, and without having to resort to mutating out any further native cysteines, since the use of ATP had failed to deliver this. The strategies that were tried were:

1. use of other bis-addition–elimination reagents.
2. use of lower concentrations of substrate protector.
3. use of substrate protectors other than ATP.

4.8.1 Use of other bis-Addition–Elimination reagents

In addition to dibromide **1** that has been routinely used throughout this study, and diiodide **26** which had been evaluated against MEK1-Cys222 (Section 4.5.3), 1,4-diiodobutane was also evaluated as a potential bis-addition–elimination reagent on MEK1 variants (Scheme 4.7). Interest in this compound was due to its lipophilicity, which in turn might cause 1,4-diiodobutane to have a different response to the steric



Scheme 4.7 Dha formation with 1,4-diiodobutane on MEK1 variants MEK1-Cys218 and MEK1-Cys222.

and electrostatic characteristics of a reaction site than the two dihalogeno-compounds already used. This could then result in a different regioselectivity.

Given its lipophilicity, 1,4-diiodobutane might not be expected to react with a protein in an aqueous environment, particularly as the neat liquid is immiscible with water. However, this expectation was challenged when it was found that the transcriptional regulator SarZ (staphylococcal accessory regulator Z) from *Staphylococcus aureus* reacted with 1,4-diiodobutane to form Dha with the sole native cysteine available (Section 4.11.5). The installation of Dha was shown by chemical tests, with the SarZ-Dha13 reacting with β-ME to give the β-ME adduct, thus demonstrating the reactivity of Dha, but it did not react with Ellman's reagent, demonstrating that the original cysteine had been consumed (unpublished data, done in collaboration with Marcel Scheepstra).[134] Thus, there was a good precedent for Dha formation occurring with 1,4-diiodobutane.

The treatment of MEK1-Cys218 with 1,4-diiodobutane yielded no significant levels of any product within a practical timeframe (Fig. 4.23a). Using 1,4-diiodobutane as the bis-addition–elimination reagent therefore was not a viable method of obtaining MEK1-Dha218. By way of comparison, MEK1-Cys222 was also subjected to the same conditions, where there was some conversion to the Dha product (Fig. 4.23b). Again, since the only difference between the MEK1 variants MEK1-Cys218 and MEK1-Cys222 is the position of the reaction site, this positional difference is the most likely cause of the rate difference. Substrate protection was not required for these reactions, raising the interesting prospect of regioselectivity, but the slow rate compared to the optimised procedures previously described (Section 4.5.5) meant that it was unnecessary to employ this method.

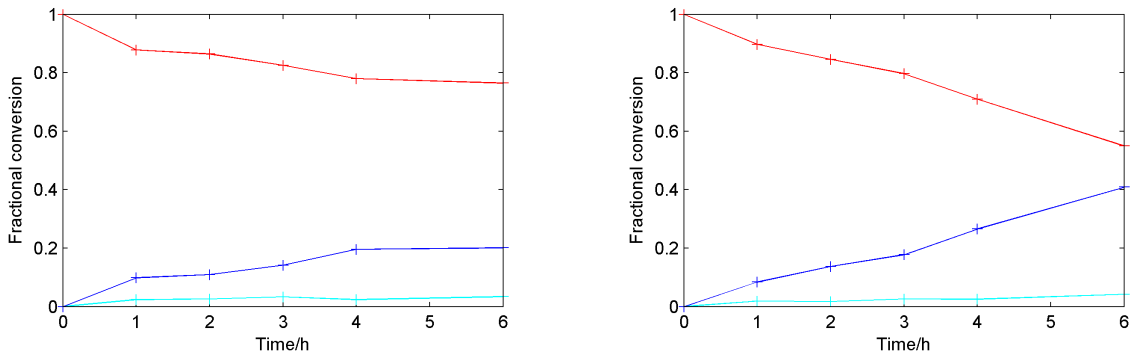


Fig. 4.23 Kinetic analysis of (attempted) conversion of unmodified MEK1 variant (red) to its corresponding Dha product (blue) upon treatment with 1,4-diiodobutane. Different MEK1 variants were subjected to this treatment: **(a)** MEK1-Cys218 and **(b)** MEK1-Cys222. The reaction goes *via* a mono-alkylated adduct intermediate (cyan). Substrate protection was not used in these reactions.

4.8.2 Use of Other Protectors in MEK1-Dha218 Formation

The slow reaction between 1,4-diiodobutane and MEK1-Cys218 meant that this reaction could not be used for MEK1-Dha218 production. Attention therefore came back to the use of substrate protectors for reaction regioselectivity.

The previous use of ATP as a substrate protector in Dha formation on MEK1-Cys218 had resulted in no reaction occurring at the desired site. The lack of reaction in the presence of ATP suggested that the previous protection conditions were too stringent to allow the reaction to occur at position 218. This stringency was relaxed by systematically lowering the ATP concentration. As expected, over-alkylated MEK1-Dha218 occurred at low ATP concentration while little reaction occurred at high ATP concentration. However, none of the conditions of intermediate ATP concentration resulted in the sequestration of MEK1-Dha218 where at best, MEK1-Dha218 was a minority species (Fig. 4.24). It was also found that the ATP concentration originally used (10 mM) was in fact near the minimum ATP concentration needed for full inhibition of the over-alkylation. Modulation of protector concentration therefore could not be used to confer regioselectivity.

As had been undertaken previously on MEK1-Cys222 (Section 4.5.4), smaller substrate inhibitors were investigated for the purpose of reducing protection stringency. The truncated adenosine phosphates, ADP and AMP were again tried, but neither of these allowed any reaction to occur between MEK1-Cys218 and dibromide **1**. Further truncation to adenosine did however allow for a reaction to occur and for

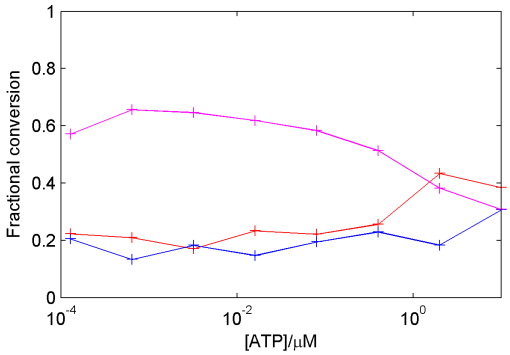


Fig. 4.24 Quantitative analysis of MEK1-Cys218-derived products across an ATP dilution series. The desired product, MEK1-Dha218 (blue) was never the most abundant species over the range of ATP concentrations. MEK1-Cys218 (red) was the prevalent species down to [ATP]∼1 mM, while over-alkylated MEK1-Dha218 (magenta) was prevalent over the rest of the ATP concentration range tested. 6 h reaction endpoint used.

the first time, the Dha product became the major species in the reaction mixture (Fig. 4.25). Although incomplete, this gave tentative signs that the substrate protection strategy could still be used to produce MEK1-Dha218. The work on MEK1-Dha218 formation was subsequently left as work in progress in favour of the biological evaluation of MEK1-pCys222, described in the next Chapter.

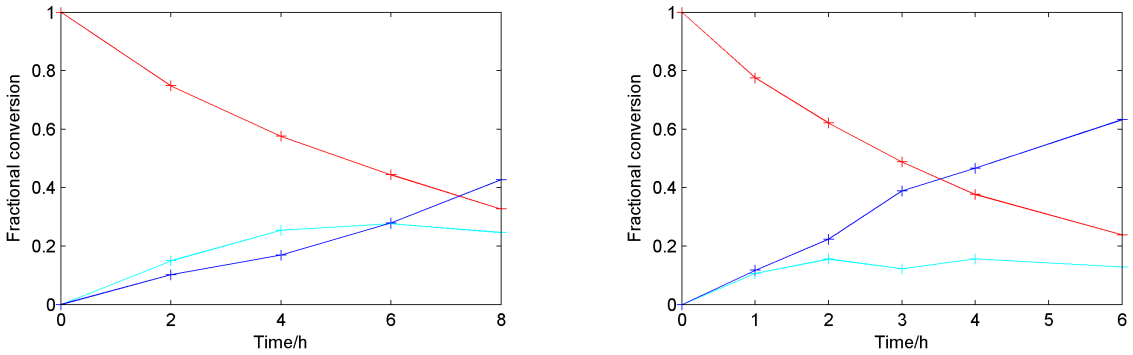


Fig. 4.25 Kinetic analysis of conversion of unmodified MEK1 variant (red) to Dha product (blue). **(a)** MEK1-Cys218 was treated with dibromide **1** in the presence of adenosine. The rate of reaction was slower than when **(b)** MEK1-Cys222 was treated with dibromide **1** in the presence of ATP (reproduced from Fig. 4.7a for convenience of comparison). The reaction goes *via* a mono-alkylated bromide adduct intermediate (cyan).

4.9 Other MEK1 Variants

Like the chemical modification study for MEK1-Dha218 formation, the study of global chemical modification on MEK1 is also a work in progress. The two phosphorylation sites have been explored separately by chemical modification, since it made reaction optimisation easier to undertake. A logical progression

of the work from a biological perspective would be to investigate both sites together in a single variant. The study of double phosphorylation by chemical modification would thus call for the MEK1 variant MEK1-G(−19)F/S218C/S222C/C277S/C376S (Table 4.4, Entry 1, referred to as MEK1-Cys218Cys222). The plasmid containing the gene of MEK1-Cys218Cys222 was made and the encoded protein subsequently expressed and purified, but no further investigation was carried out.

Table 4.4 Summary of further MEK1 variants.

Entry	Mutation Profile							Remarks
1	G(−19)F			S218C	S222C	C277S	C376S	Double modification
2	G(−19)F			S218E				Constitutively active
3	G(−19)F			S218E	S222E			Constitutively active
4	G(−19)F	C121S			S222C	C277S	C376S	Off-target attack prevention
5	G(−19)F		C207S		S222C	C277S	C376S	Off-target attack prevention
6	G(−19)F	C121S		S218C		C277S	C376S	Off-target attack prevention
7	G(−19)F		C207S	S218C		C277S	C376S	Off-target attack prevention

The MEK1 variant library was further extended to create variants that would cover most of the future prospects of the project. DNA constructs of the other constitutively active variants were made to complement MEK1-S222E, filling in all the other permutations of S→E phosphorylation site mutation (Table 4.4, Entry 2 and 3). These variants would be used in future biological studies as controls in the same way as MEK1-S222E (Section 5.4). For the purpose of chemical modification, a number of variants were made where further native cysteines were mutated to Ser (Table 4.4, Entries 4–7). Their intended use was to eliminate over-alkylation, if methods such as substrate protection could not deliver regioselectivity. Since Dha formation at both native phosphorylation sites seems successful, these variants are unlikely to be needed.

4.10 Summary

The study on MEK1 followed the workflow established for p38α in Chapter 2 with variant design, followed by protein expression, purification, modification and characterisation. Yet, the results obtained during the MEK1 study are quite different from that of p38α. Firstly, starting from a variant of *de novo* design meant there were additional challenges to overcome in optimising the construct design and protein expression before chemical modification could begin. Once chemical modification had commenced however, differences in physical and chemical behaviour of MEK1 over p38α meant that further optimisation parameters above rate and conversion of the reaction had to be considered. From a physical perspective,

MEK1 was more liable to precipitate than had been observed for p38α while from a chemical perspective, the increased number of cysteines in MEK1 over p38α meant that despite the careful variant design, off-target modification still occurred. This meant that there was added emphasis on optimising for protein stability and reaction regioselectivity. An increased use of protein stabilisation agents in the buffer and substrate protection were able to overcome both of these problems respectively.

Two MEK1 variants were implicated in this study: MEK1-Cys222 and MEK1-Cys218. Each one presented a chemical modification site corresponding to a native phosphorylation site in MEK1-WT. Although only MEK1-Cys222 was progressed to the desired outcome of chemical phosphorylation, the independent study of the variants raised interesting insights into their behaviour upon chemical modification, and the factors that influenced them. In particular, the requirement of different substrate protecting agents in Dha formation in the two variants showed that in this case, chemical modification behaved as a probe for the structure and extent of the ATP binding site. The use of substrate protection with ATP successfully eliminated over-alkylation in the formation of MEK1-Dha222, but it was more difficult to strike the same balance in producing MEK1-Dha218. Some initial success was made by using adenosine as substrate protector on MEK1-Dha218 with as yet incomplete conversion.

It was hoped that studies on differences between MEK1-Cys222 and MEK1-Cys218 in chemical modification could translate to studies in the differences of the biological action of their phosphocysteine derivatives. This was particularly the case for MEK1 over p38α since both native phosphorylation sites of MEK1 are Ser, making the study of both sites by chemical modification more biologically meaningful. However, the formation of MEK1-Dha218 proved to be more challenging than the formation of MEk1-Dha222, which subsequently made MEK1-pCys222 easier to produce. The completion of the synthetic procedures on MEK1-Cys222 made its series of chemical variants ripe for further studies into its biological action, where the question of whether phosphocysteine is a sufficient mimic of phosphoserine could be addressed. This is the subject of discussion in the next Chapter.

4.11 Experimental

4.11.1 General Measures

Synthetic Chemistry

General considerations in synthetic chemistry were similar to those previously described (Section 2.11.1).

Molecular Biology

All bacterial handling was done in a sterile environment, either within close proximity to a Bunsen burner flame or inside a biological safety cabinet. The pET28a-MEK1-S222C/C277S/C376S plasmid was purchased from Genscript as a synthetic construct. The *E. coli* expression codon-optimised gene was ligated into the multiple cloning region of pET28a using NdeI and BamHI as the restriction sites. LB and TB media were bought as pre-formulated dry granules from Melford and diluted according to the manufacturer's specifications. YT, SB, SOC and NZY$^+$ media were prepared from biological laboratory reagents. 2YT medium was bought as pre-formulated dry granules from Sigma Aldrich and diluted according to the manufacturer's specifications. All other biological laboratory buffer and media reagents were bought from common suppliers and used as purchased. BL21(DE3) and NovaBlue *E. coli* chemically competent cells (Novagen) were handled according to the manufacturer's instructions, thawing on ice before incubation with DNA. NovaBlue was used for cloning, while BL21(DE3) was used for protein expression. XL1-Blue and XL10-Gold *E. coli* chemically super-/ultra-competent cells (Agilent) likewise were handled according to the manufacturer's instructions. XL1-Blue and XL10-Gold were used for cloning. All other procedures for general measures for mutagenesis, bioinformatics, protein expression and purification, SDS-PAGE protein gel analysis, the making of solid LB agar medium and glycerol stocks were similar to those previously described (Section 2.11.1).

Gene Sequence Analysis

DNA sequencing was done by Geneservice, Oxford using the primers T7 forward (5′-TAATACGACTCACTATAGGG-3′) and T7 reverse (5′-GCTAGTTATTGCTCAGCGG-3′) for genes contained in pET vectors. Plasmid sequences displayed are consensus sequences spliced together from the results of a T7 forward/T7 reverse pair read. The forward and reverse complimentary sequences returned from the gene sequencing

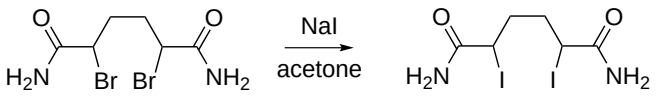
service were analysed by sequence alignment of the two returned sequences, followed by base-by-base string comparison to the desired reference sequence. Sequence data manipulation and comparison was done using scripts written in MATLAB. In the case of serious mismatching between observed and desired sequences, the raw data were visualised and evaluated using FinchTV version 1.4.0. Examples of the MATLAB codes and functions used are in Appendix A.

Protein Modification Reactions

ATP was bought from Sigma Aldrich and used as purchased. Dibromide **1** was synthesised as previously described (Section 2.11.2). All other protein modification reaction general measures were as previously described (Section 2.11.1). LC-MS analysis was done on a Micromass LCT Premier instrument using either a ProSwift® RP-4H monolith, 1×50 mm column (Dionex), or a Chromolith® FastGradient RP-18 endcapped 50×2 mm monolithic HPLC column (Merck). A linear gradient was run from 5–95% of solvent mixture B (1% formic acid in acetonitrile) into solvent mixture A (1% formic acid in water). The gradient was run with a flow rate of 0.4 mL/min over 4 min for the ProSwift column and with a flow rate of 0.3 mL/min over 6 min for the Chromolith column. All other general measures regarding protein modification reactions and quantification of MS data were as previously described (Section 2.11.1). Gel densitometry of samples analysed by SDS-PAGE was done either using ImageQuant (GE Healthcare) as previously described (Section 3.8.3) or a script written in MATLAB (Appendix E). In the case where ImageQuant was used, the data were output into the appropriate programs (Excel or MATLAB) for further mathematical and statistical analysis.

4.11.2 Synthesis of Protein Modification Reagents

Preparation of 2,5-diiodo-adipic acid diamide (26)



1 **26**

A suspension of dibromide **1** (0.115 g, 0.38 mmol, 1.0 eq.) in dry acetone (15 mL) was heated under reflux. Dry DMF was added (1 mL) to give a solution before a solution of sodium iodide (1.2 g, 8.0 mmol, 21 eq.) in acetone (5 mL) was added. The mixture was heated under reflux with stirring for 4 h, after which

time white precipitate had formed and the reaction mixture had turned yellow. The mixture was filtered while hot and the solvent removed under vacuum to give a yellow, solid residue which was triturated with water (20 mL). The remaining precipitate was filtered, washed with water (10 mL) and dried under high vacuum to give *2,5-diiodo-adipic acid diamide* (**26**, 55 mg, 36%) as a yellow powder: m.p. (MeOH) 201–202 °C dec.; ^{1}H NMR (400 MHz, DMSO-d_6) δ ppm 7.63 (2 H, br s, NH_2), 7.11 (2 H, br s, NH_2), 4.25–4.35 (2 H, m, 2×CHI), 1.64–1.96 (4 H, m, CH_2CH_2); ^{13}C NMR (101 MHz, DMSO-d_6) δ ppm (both diastereomers reported) 172.6, 172.5 (C=O), 36.6, 36.4 (CH_2*C*HBr), 25.4, 24.6 (*C*H_2CHBr); *m/z* (HRMS: ESI+) Calculated: 418.8724, Observed: 418.8728 ([M+Na]$^{+}$).

4.11.3 Protein Expression and Purification

Cloning of pET28a-MEK1-S222C/C277S/C376S Plasmid

Transformation: Synthetic pET28a-MEK1-S222C/C277S/C376S plasmid (Genscript, 1 µL, 200 ng/µL) was transformed into NovaBlue *E. coli* chemically competent cells (50 µL) by heat shock. The DNA was incubated with the thawed cells on ice for 5 min before heat shock was performed at 42 °C for 30 s. After incubating on ice for a further 2 min, the cells were fed with SOC medium (250 µL) and incubated in a shaker at 37 °C, 250 rpm for 30 min.

Bacterial Culturing: The cells were plated onto solid LB agar medium containing kanamycin (50 µg/mL) at three different densities (200 µL, 50 µL and 20 µL + 20 µL SOC medium) and the plates were incubated at 37 °C for 17 h. The plate with the lowest cell concentration gave distinct colonies while the other plates gave bacterial lawns. 3 colonies were selected and cultured separately in liquid LB medium (10 mL) containing kanamycin (50 µg/mL) in a shaking incubator at 37 °C, 250 rpm for 16 h.

DNA Purification and Storage: After glycerol stocks were made, the remaining cells were then pelleted by centrifugation (3000 rpm, 20 min, 4 °C). The resulting cell pellet was resuspended in P1 buffer and the plasmid DNA purified out using the QIAprep® Miniprep kit. The plasmid DNA was eluted in EB buffer and stored at −20 °C. DNA sequencing confirmed that the sequence of the gene was intact: plasmid concentration 118 – 127 ng/µL

Sequence of pET28a-MEK1-S222C/C277S/C376S:

```
   1 ATG GCC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG
  46 CCG CGC GGC AGC CAT ATG CCG AAA AAA AAA CCG ACC CCG ATT CAA
  91 CTG AAC CCG GCT CCG GAT GGC TCC GCT GTC AAT GGC ACC TCA AGC
 136 GCA GAA ACG AAC CTG GAA GCG CTG CAG AAA AAA CTG GAA GAA CTG
 181 GAA CTG GAT GAA CAG CAA CGT AAA CGC CTG GAA GCC TTT CTG ACC
 226 CAG AAA CAA AAA GTG GGC GAA CTG AAA GAT GAC GAT TTC GAA AAA
 271 ATC AGT GAA CTG GGC GCC GGT AAC GGC GGT GTG GTT TTT AAA GTC
 316 AGC CAT AAA CCG TCT GGT CTG GTG ATG GCA CGT AAA CTG ATT CAC
 361 CTG GAA ATC AAA CCG GCT ATT CGT AAC CAG ATT ATC CGC GAA CTG
 406 CAA GTG CTG CAT GAA TGC AAT TCT CCG TAT ATT GTT GGC TTT TAT
 451 GGT GCG TTC TAC AGT GAC GGC GAA ATT TCC ATC TGT ATG GAA CAC
 496 ATG GAC GGC GGT AGC CTG GAT CAG GTT CTG AAA AAA GCA GGC CGT
 541 ATC CCG GAA CAA ATT CTG GGT AAA GTC TCT ATT GCT GTG ATC AAA
 586 GGC CTG ACG TAC CTG CGT GAA AAA CAT AAA ATC ATG CAC CGC GAT
 631 GTG AAA CCG TCA AAC ATC CTG GTT AAT TCG CGC GGT GAA ATT AAA
 676 CTG TGC GAC TTT GGC GTT AGC GGT CAG CTG ATT GAT AGT ATG GCG
 721 AAC TCT TTC GTC GGC ACC CGT AGT TAT ATG TCC CCG GAA CGC CTG
 766 CAG GGT ACG CAT TAC TCA GTG CAA TCG GAT ATC TGG TCA ATG GGC
 811 CTG TCG CTG GTT GAA ATG GCC GTT GGT CGT TAT CCG ATT CCG CCG
 856 CCG GAC GCC AAA GAA CTG GAA CTG ATG TTT GGC TCT CAG GTT GAA
 901 GGT GAT GCG GCC GAA ACC CCG CCG CGT CCG CGT ACC CCG GGT CGT
 946 CCG CTG AGC TCT TAT GGT ATG GAC AGC CGT CCG CCG ATG GCA ATC
 991 TTC GAA CTG CTG GAT TAC ATT GTG AAT GAA CCG CCG CCA AAA CTG
1036 CCG AGC GGC GTG TTT AGC CTG GAA TTT CAG GAC TTC GTC AAC AAA
1081 TGT CTG ATC AAA AAT CCG GCA GAA CGT GCT GAT CTG AAA CAA CTG
1126 ATG GTC CAC GCT TTT ATT AAA CGC TCC GAC GCG GAA GAA GTG GAT
1171 TTC GCC GGT TGG CTG TCT TCG ACC ATT GGC CTG AAT CAA CCG TCA
1216 ACG CCG ACC CAC GCT GCT GGT GTG
```

General Procedure for Mutagenesis of pET28a-MEK1-S222C/C277S/C376S Plasmid

Mutagenesis Primers: The following primers were used to make the MEK1 mutant library:

Mutation	Direction	Primer Sequence	Remarks
C222S	forward	5′-GCTGATTGATAGTATGGCGAACTCTTTCGTCGGCACC-3′	Revert to WT
C222S	reverse	5′-GGTGCCGACGAAAGAGTTCGCCATACTATCAATCAGC-3′	Revert to WT
S277C	forward	5′-GATGTTTGGCTGCCAGGTTGAAGGTGATGCGGC-3′	Revert to WT
S277C	reverse	5′-GCCGCATCACCTTCAACCTGGCAGCCAAACATC-3′	Revert to WT
S376C	forward	5′-CGGTTGGCTGTGCTCGACCATTGGCCTGAATC-3′	Revert to WT
S376C	reverse	5′-GATTCAGGCCAATGGTCGAGCACAGCCAACCG-3′	Revert to WT
C222E	forward	5′-GCTGATTGATAGTATGGCGAACGAATTCGTCGGCACCCG-3′	–
C222E	reverse	5′-CGGGTGCCGACGAATTCGTTCGCCATACTATCAATCAGC-3′	–
S218C	forward	5′-CGGTCAGCTGATTGATTGTATGGCGAACTCTTTCG-3′	–
S218C	reverse	5′-CGAAAGAGTTCGCCATACAATCAATCAGCTGACCG-3′	–
S218C	forward	5′-GCGGTCAGCTGATTGATTGTATGGCGAACTGC-3′	for S218C/S222C
S218C	reverse	5′-GCAGTTCGCCATACAATCAATCAGCTGACCGC-3′	for S218C/S222C
S218E	forward	5′-CGGTCAGCTGATTGATGAAATGGCGAACTCTTTCGTCGGC-3′	–
S218E	reverse	5′-GCCGACGAAAGAGTTCGCCATTTCATCAATCAGCTGACCG-3′	–
S218E	forward	5′-CGGTCAGCTGATTGATGAAATGGCGAACGAATTCGTCGGC-3′	for S218E/S222E
S218E	reverse	5′-GCCGACGAATTCGTTCGCCATTTCATCAATCAGCTGACCG-3′	for S218E/S222E
C121S	forward	5′-CTGCAAGTGCTGCATGAATCTAATTCTCCGTATATTGTTGGC-3′	–
C121S	reverse	5′-GCCAACAATATACGGAGAATTAGATTCATGCAGCACTTGCAG-3′	–
C207S	forward	5′-CGCGGTGAAATTAAACTGTCTGACTTTGGCGTTAGCGG-3′	–
C207S	reverse	5′-CCGCTAACGCCAAAGTCAGACAGTTTAATTTCACCGCG-3′	–

Mutagenesis Reaction: A similar procedure to that previously described for p38α (Section 2.11.3) was used. The reactions were assembled with the components listed below in the given order:

Substance	Concentration	Quantity/µL
dNTPs	10 mM each	0.5
forward primer	12.5 µM	0.5
reverse primer	12.5 µM	0.5
MilliQ water	–	20.0
10× reaction buffer	–	2.5
Pfu DNA polymerase	2.5 U/µL	0.5
plasmid	10 or 50 ng/µL	0.5
final volume	–	25.0

The quantities given in the above table are for an individual reaction. In assembling the reactions, all the components were combined into a master mix, excluding the plasmid template. An individual master mix was made for every mutation made. The master mix was then divided into 2 for reactions with 2 different plasmid concentrations (10 and 50 ng/µL). For difficult mutagenesis reactions, DMSO (1.25 µL per reaction) was added to the reaction mixture (the volume of water was adjusted to maintain the final volume). The assembled reactions were heated in a thermal cycler using the following programme:

Segment	No. of Cycles	Temp./°C		Time
1	1	95	–	30 s
2	16	95	denaturation	30 s
		55	annealing	1 min
		68	extension	15 min
3	1	4	–	∞

DpnI (0.5 µL) was then added to each mutagenesis reaction and the mixtures incubated at 37 °C for 2 h. Presence of plasmid-length DNA was confirmed by gel electrophoresis on an agarose gel. Samples (2.5 µL) were loaded onto an agarose gel (0.8% agarose in TAE buffer, supplemented with 5.3 µL of a 1% solution of ethidium bromide per 50 mL gel) and the gel run at 150 V. The reactions with lowest template plasmid concentration and which still contained synthetic plasmid-length DNA were taken forward for each mutation made.

Transformation: For each mutation, XL1-Blue *E. coli* supercompetent cells or XL10-Gold *E. coli* ultracompetent cells (25 µL) were thawed on ice and mutagenesis reaction mixture (2.5 µL) added. The tube was flicked for mixing before incubation of the mixture on ice for 30 min. After heat shock at 42 °C,

40 s, the cells were returned onto ice, incubating for a further 5 min. The cells were then fed with SOC or NZY$^+$ medium (250 µL) and incubated with shaking at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The transformed cells (250 µL) were selected on LB/agar plates supplemented with kanamycin (50 mg/mL final concentration). After spreading the cells on the plate, the plates were incubated at 37 °C for 14–17 h, producing colonies. 3 colonies were selected for each mutation and further cultured by inoculating them into liquid LB medium (3×10 mL), again supplemented with the required antibiotic. The cultures were incubated with shaking at 37 °C, 250 rpm for 16–18 h.

DNA Purification and Storage: After appropriate glycerol stocks were made, the remainder of the cultures were pelleted by centrifugation (3750 rpm, 10 min, 4 °C) and the supernatant discarded. The pellets were resuspended in P1 buffer (250 µL) from the Miniprep kit (Qiagen) and plasmid purification was performed using the rest of the kit as to the manufacturer's guidelines. DNA sequencing confirmed the colonies which contained the desired gene sequence: plasmid concentration: 70–188 ng/µL

Sequence of MEK1-WT with the positions of mutation highlighted:

```
  1 ATG TTC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG
 46 CCG CGC GGC AGC CAT ATG CCG AAA AAA AAA CCG ACC CCG ATT CAA
 91 CTG AAC CCG GCT CCG GAT GGC TCC GCT GTC AAT GGC ACC TCA AGC
136 GCA GAA ACG AAC CTG GAA GCG CTG CAG AAA AAA CTG GAA GAA CTG
181 GAA CTG GAT GAA CAG CAA CGT AAA CGC CTG GAA GCC TTT CTG ACC
226 CAG AAA CAA AAA GTG GGC GAA CTG AAA GAT GAC GAT TTC GAA AAA
271 ATC AGT GAA CTG GGC GCC GGT AAC GGC GGT GTG GTT TTT AAA GTC
316 AGC CAT AAA CCG TCT GGT CTG GTG ATG GCA CGT AAA CTG ATT CAC
361 CTG GAA ATC AAA CCG GCT ATT CGT AAC CAG ATT ATC CGC GAA CTG
406 CAA GTG CTG CAT GAA TGC AAT TCT CCG TAT ATT GTT GGC TTT TAT
451 GGT GCG TTC TAC AGT GAC GGC GAA ATT TCC ATC TGT ATG GAA CAC
496 ATG GAC GGC GGT AGC CTG GAT CAG GTT CTG AAA AAA GCA GGC CGT
541 ATC CCG GAA CAA ATT CTG GGT AAA GTC TCT ATT GCT GTG ATC AAA
586 GGC CTG ACG TAC CTG CGT GAA AAA CAT AAA ATC ATG CAC CGC GAT
631 GTG AAA CCG TCA AAC ATC CTG GTT AAT TCG CGC GGT GAA ATT AAA
676 CTG TGC GAC TTT GGC GTT AGC GGT CAG CTG ATT GAT AGT ATG GCG
721 AAC TCT TTC GTC GGC ACC CGT AGT TAT ATG TCC CCG GAA CGC CTG
766 CAG GGT ACG CAT TAC TCA GTG CAA TCG GAT ATC TGG TCA ATG GGC
811 CTG TCG CTG GTT GAA ATG GCC GTT GGT CGT TAT CCG ATT CCG CCG
856 CCG GAC GCC AAA GAA CTG GAA CTG ATG TTT GGC TGC CAG GTT GAA
```

```
 901 GGT GAT GCG GCC GAA ACC CCG CCG CGT CCG CGT ACC CCG GGT CGT
 946 CCG CTG AGC TCT TAT GGT ATG GAC AGC CGT CCG CCG ATG GCA ATC
 991 TTC GAA CTG CTG GAT TAC ATT GTG AAT GAA CCG CCG CCA AAA CTG
1036 CCG AGC GGC GTG TTT AGC CTG GAA TTT CAG GAC TTC GTC AAC AAA
1081 TGT CTG ATC AAA AAT CCG GCA GAA CGT GCT GAT CTG AAA CAA CTG
1126 ATG GTC CAC GCT TTT ATT AAA CGC TCC GAC GCG GAA GAA GTG GAT
1171 TTC GCC GGT TGG CTG TGC TCG ACC ATT GGC CTG AAT CAA CCG TCA
1216 ACG CCG ACC CAC GCT GCT GGT GTG
```

In obtaining this library, the necessary mutations were made in a stepwise fashion, purifying the resulting mutant plasmid after each round of muagenesis.

Expression Optimisation for the Protein Expression of MEK1 Variants

Bacterial Culturing: Glycerol stocks of previously transformed *E. coli* containing the gene for the MEK1-G(−19)F/S222C/C277S/C376S were streaked onto LB/agar plates supplemented with kanamycin (50 mg/mL). The plates were incubated at 37 °C for 14 h. 3 colonies were selected and cultured further by inoculating into fresh liquid LB medium (3×10 mL), again supplemented with kanamycin (50 mg/mL) and the cultures incubated with shaking at 37 °C, 250 rpm for 15 h to give the starter cultures.

Protein Expression: One of the starter cultures was used to inoculate the test cultures. 5 different media were tested:

- LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl)
- YT medium (1.6% tryptone, 0.5% yeast extract, 0.5% NaCl)
- 2YT medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl)
- TB medium (1.2% tryptone, 2.4% yeast extract, 0.94% K_2HPO_4, 0.22% KH_2PO_4, 0.4% glycerol)
- SB medium (3.2% tryptone, 2% yeast extract, 0.5% NaCl)

Aliquots of media (3×10 mL for each medium) were inoculated with starter culture (50 µL) and these new test cultures were grown to $OD_{600} \sim 0.6$ (∼1.0 for TB medium) in an incubator shaker at 37 °C, 250 rpm. The temperature was then reduced to 18 °C before induction with IPTG at one of 3 concentrations (0.2, 0.5 or 1.0 mM final concentration, all 3 concentrations used for each type of medium). The cultures were further incubated at 18 °C, 250 rpm for 17 h.

Protein Purification: The cells from the test cultures were harvested by centrifugation (3750 rpm, 10 min, 4 °C), the supernatants discarded and the pellets flash frozen in liquid nitrogen. The pellets were re-

suspended in BugBuster Master Mix (300 µL each) and incubated on a laboratory wheel at room temperature for 30 min. The lysates were clarified by centrifugation (13 200 rpm, 10 min, 4 °C) and the supernatents taken. The clarified lysates were filtered before loading onto self-packed Ni^{2+} affinity columns (100 µL of a 50% slurry, Novagen) and the columns shaken at 4 °C, 700 rpm for 1.5 h. The columns were washed (5×200 µL) with Ni^{2+} affinity wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 50 mM imidazole, 5% glycerol, 0.5 mM TCEP) and eluted (2×100 µL) with Ni^{2+} elution buffer (same as wash buffer but with 500 mM imidazole), collecting both elutions as separate fractions. Samples of the pellets, flow-through fractions and the 1st elution fractions were analysed by SDS-PAGE to determine the relative amounts of desired protein under each condition.

General Procedure for Protein Expression of MEK1 Variants

Except for MEK1-S222E, all MEK1 variants were expressed using the following, representative procedure.

Transformation: Plasmid encoding the mutant MEK1-G(−19)F/S222C/C277S/C376S (referred to as MEK1-Cys222) was transformed into BL21(DE3) *E. coli* chemically competent cells by heat shock. BL21(DE3) *E. coli* competent cells (25 µL) were thawed on ice and plasmid DNA (1 µL) added. The mixture was incubated on ice for 5 min before heat shock was performed at 42 °C for 30 s and immediately returning on ice. After a further incubation period of 2 min, the cells were fed with SOC medium (125 µL) before incubation with shaking at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The freshly transformed cells (20 µL) were streaked onto LB/agar plates supplemented with kanamycin (50 µg/mL final concentration). 3 colonies were selected and cultured further in liquid LB medium (3×10 mL), again supplemented with kanamycin to give starter cultures.

Protein Expression: After suitable glycerol stocks had been made, one of the starter cultures was further cultured. 2YT medium (5×800 mL), again supplemented with kanamycin was inoculated with the starter culture (5×1.6 mL). These new cultures were grown to $OD_{600} \sim 0.1$ in an incubator with shaking at 37 °C, 180 rpm before the temperature was turned to 20 °C. The cultures were grown further to $OD_{600} = 0.6$–0.7. Protein expression was then induced by the addition of IPTG (1 mM final concentration) and incubation continued at the same conditions for 17 h. The cells were finally harvested by centrifugation

(8000 rpm, JA10 rotor, 5×10 min, 4 °C), the pellets placed into bags, flash frozen in liquid nitrogen and stored at −80 °C.

Protein Expression of MEK1-S222E

Transformation: Plasmid (1.00 µL, 188 ng/µL) was transformed into BL21(DE3) chemically competent cells (25 µL) by heat shock. The plasmids were incubated with the cells on ice for 5 min before heat shock was performed for 30 s at 42 °C. After shock, the cells were incubated again on ice for 2 min before being fed with SOC medium (125 µL). The cells were then incubated in a shaker at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The cells were plated onto solid LB agar medium supplemented with kanamycin (50 µg/mL) at 3 different cell concentrations (5 µL and 20 µL + 40 µL SOC medium, 50 µL) and incubated at 37 °C for 17 h. All plates gave distinct colonies. 3 colonies of each construct from the lower density plates were selected and cultured further in liquid LB medium (15 mL) supplemented with kanamycin (50 µg/mL) in a shaking incubator at 37 °C, 250 rpm for 17 h.

Stock Preparation and Protein Expression: Glycerol stocks of all the selected colonies were made by diluting glycerol solution (250 µL of a 60% stock solution in water) with bacterial broth (723 µL) and supplementing with glucose (27.7 µL of a 2 M stock solution). The glycerol stocks were stored at −80 °C. One colony of each construct was then cultured further a third time. Each construct was inoculated into LB medium (3×800 mL) supplemented with kanamycin (50 µg/mL) with the starter culture (3×4.5 ml). The cultures were grown to $OD_{600} \sim 0.15$ in a shaking incubator at 37 °C, 180 rpm before the temperature was turned down to 20 °C and the cultures grown further to $OD_{600} \sim 0.8$. Induction was then done by adding IPTG (800 µL of a 1 M solution in water, final concentration: 1 mM). Expression was allowed to occur over 17 h before the cells were harvested by centrifugation (9500 rpm, JA10 rotor, 3×10 min, 4 °C). The pellets were flash frozen in liquid nitrogen and stored at −80 °C.

General Procedure for Protein Purification of MEK1 Variants

All MEK1 variants, except for MEK1-Cys218Cys222 were purified using the following, representative procedure.

General Considerations: All the columns used were carefully cleaned as appropriate between separations, particularly as mutants of the same protein were being purified. For Ni^{2+} affinity chromatography,

the nickel was stripped (15 CV) with stripping buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 50 mM EDTA) and the column washed with water before recharging with $NiSO_4$ (0.1 M solution in water). For the gel filtration column, the column was washed (2 CV) with sodium hydroxide (0.5 M) and water before equilibration.

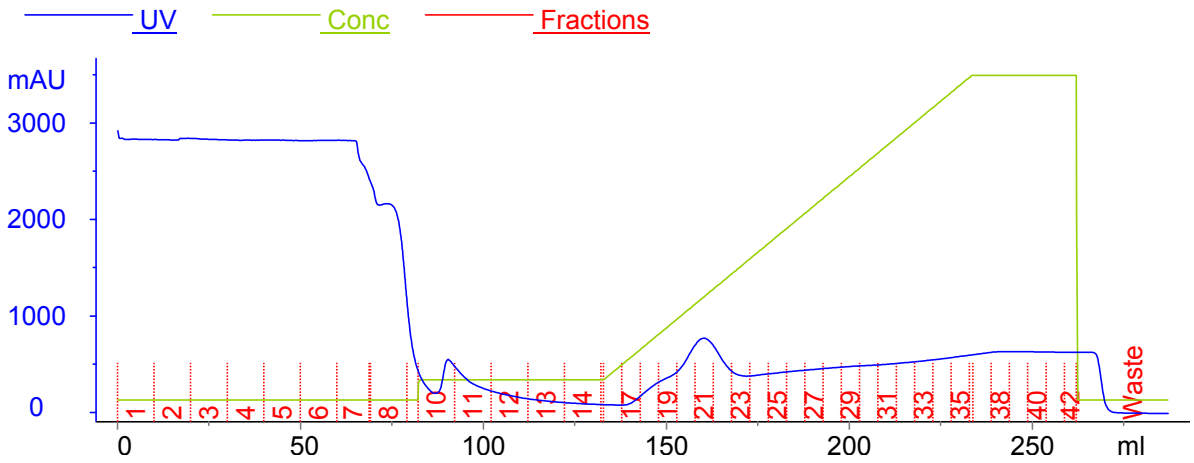
Cell Lysis: The cell pellets (30 g) were combined, thawed on ice and re-suspended (100 mL) in ice cold Ni^{2+} affinity lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 5% glycerol, 0.5 mM TCEP) with stirring at 4 °C. Lysozyme (120 mg) was added and the mixture stirred for a further 3–4 h before sonication using a sonicator equipped with a microtip (20×3 s blasts, 60% amplitude, 2770 J total energy). DNase (1 mg) was added and the mixture stirred for a further 1 h before the lysate was clarified by centrifugation (20 000 rpm, JA25.50 rotor, 40 min, 4 °C). The lysate was further clarified by filtration using 0.2 µm syringe filters.

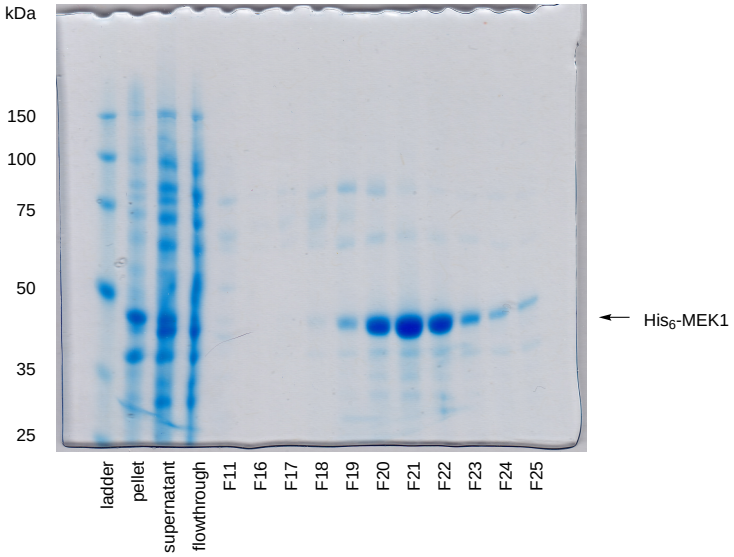
Ni^{2+} Affinity Chromatography: Using an ÄKTA Purifier™ FPLC system, the clarified lysate was loaded (110 mL) onto a pre-packed 5 mL HisTrap™ HP Ni^{2+} affinity column, chasing through (17.5 mL, 3.5 CV) with more lysis buffer. The column was washed (50 mL, 10 CV) with Ni^{2+} affinity wash buffer (same as lysis buffer but with 50 mM imidazole) and the bound protein eluted using a linear gradient (100 mL, 20 CV, 5 mL fractions) from wash buffer to Ni^{2+} affinity elution buffer (same as lysis buffer but with 500 mM imidazole). The fractions containing the desired protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and combined.

Size Exclusion Chromatography and Storage: The collected fractions were concentrated by Vivaspin (MWCO 10 000 Da) to a final volume ∼2 mL before filtering. The concentrated proteins were loaded onto a HiLoad™ 16/60 Superdex 200 gel filtration column and filtered (180 mL, 1.5 CV, 1.2 mL fractions) into MEK1 storage buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP). The fractions containing the desired protein were determined from the UV trace of the FPLC report file and their purity was determined from SDS-PAGE analysis. The pure and impure fractions of the desired protein were collected separately, re-concentrated by Vivaspin (MWCO 10 000), divided into aliquots, flash frozen in liquid nitrogen and stored at −80 °C.

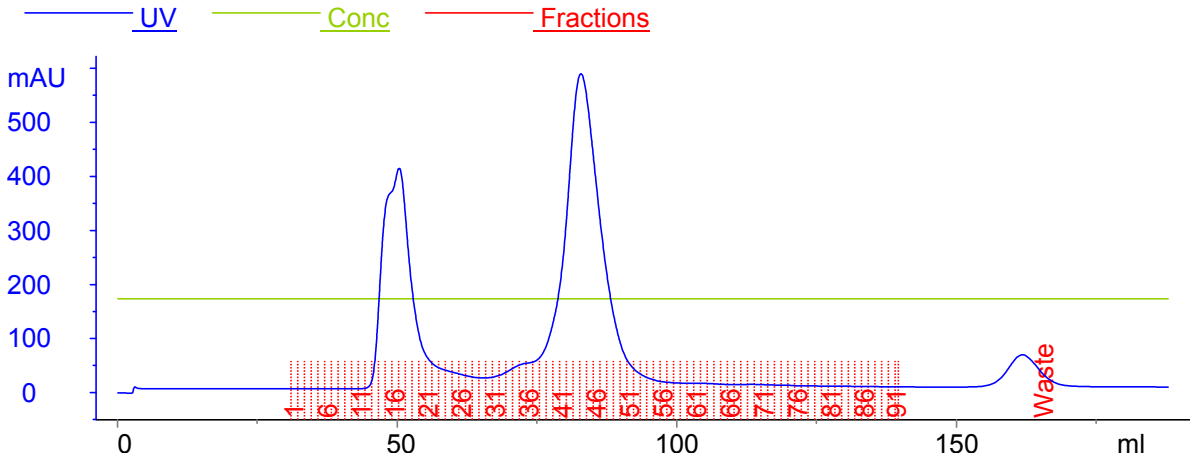
MEK1-Cys222:

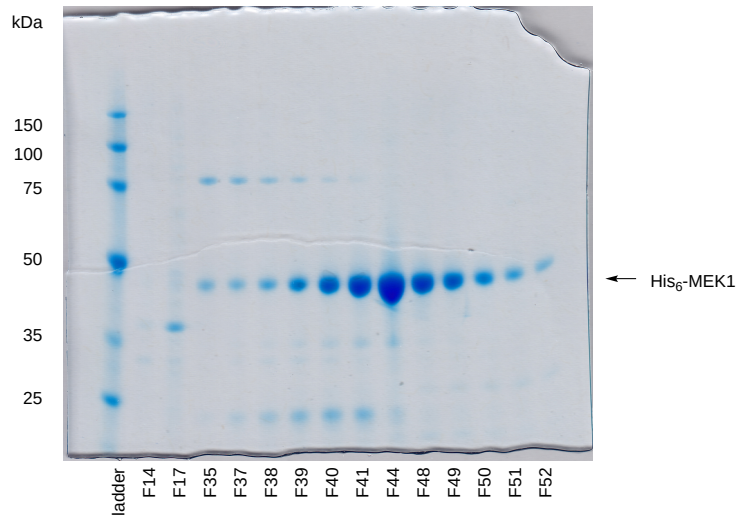
After Ni^{2+} affinity chromatography:



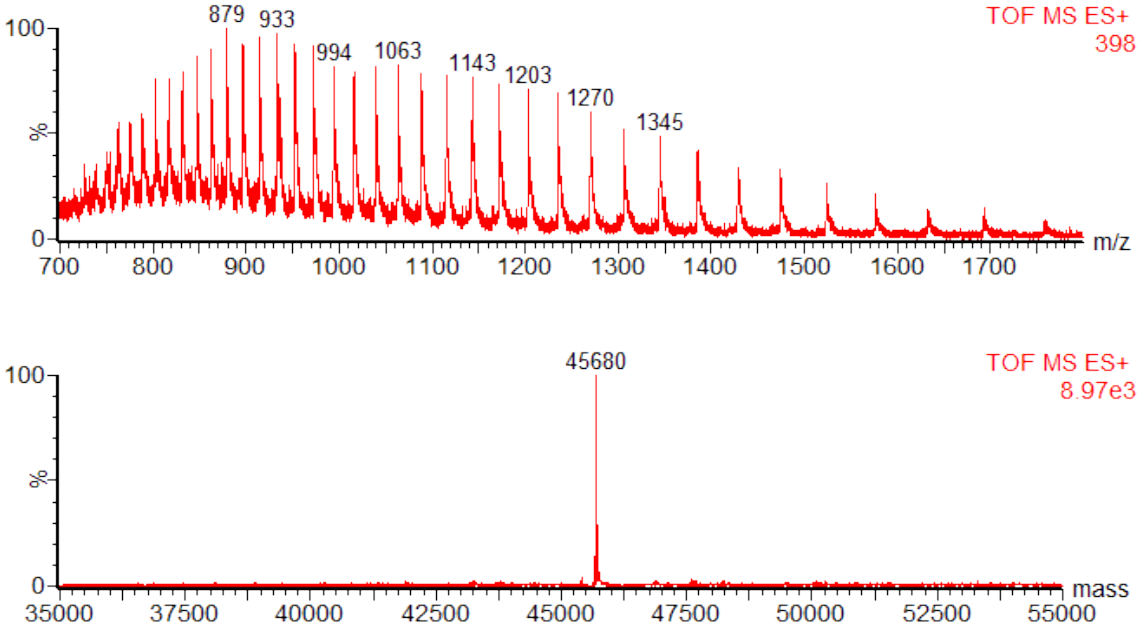


After gel filtration:





MEK1-Cys222: (calculated mass: 45 676, observed mass: 45 680),



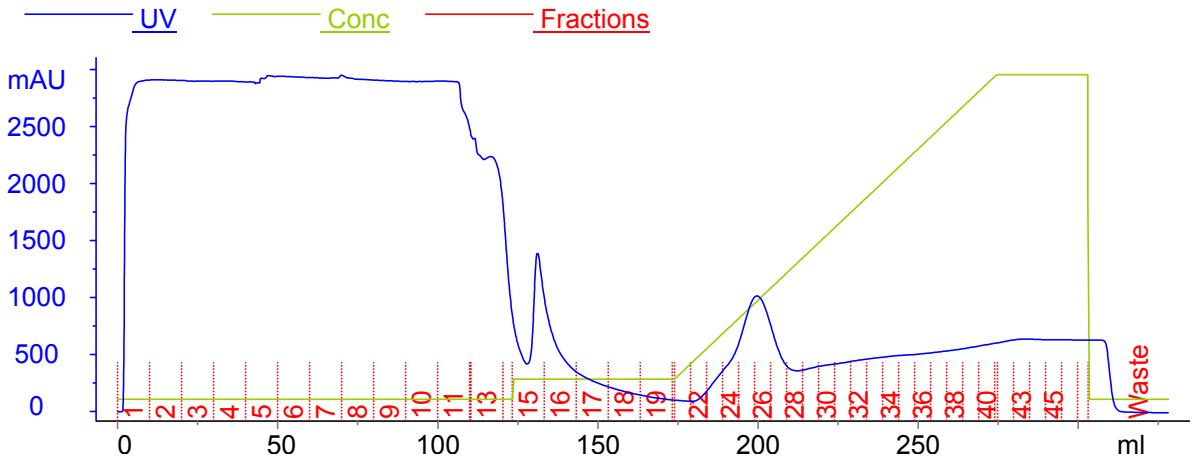
Protein sequence obtained by translation of the corresponding gene sequence:

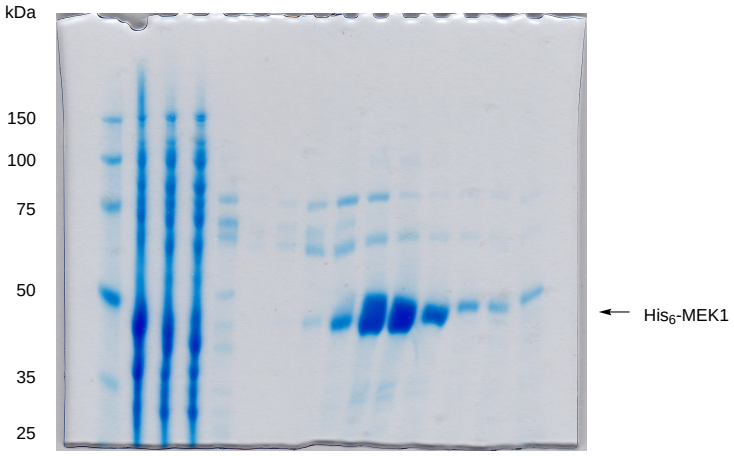
```
tag                                  MFSSHHHHHH SSGLVPRGSH
  1 MPKKKPTPIQ LNPAPDGSAV NGTSSAETNL EALQKKLEEL ELDEQQRKRL
 51 EAFLTQKQKV GELKDDDFEK ISELGAGNGG VVFKVSHKPS GLVMARKLIH
101 LEIKPAIRNQ IIRELQVLHE CNSPYIVGFY GAFYSDGEIS ICMEHMDGGS
151 LDQVLKKAGR IPEQILGKVS IAVIKGLTYL REKHKIMHRD VKPSNILVNS
201 RGEIKLCDFG VSGQLIDSMA NCFVGTRSYM SPERLQGTHY SVQSDIWSMG
```

```
251 LSLVEMAVGR YPIPPPDAKE LELMFGSQVE GDAAETPPRP RTPGRPLSSY
301 GMDSRPPMAI FELLDYIVNE PPPKLPSGVF SLEFQDFVNK CLIKNPAERA
351 DLKQLMVHAF IKRSDAEEVD FAGWLSSTIG LNQPSTPTHA AGV
```

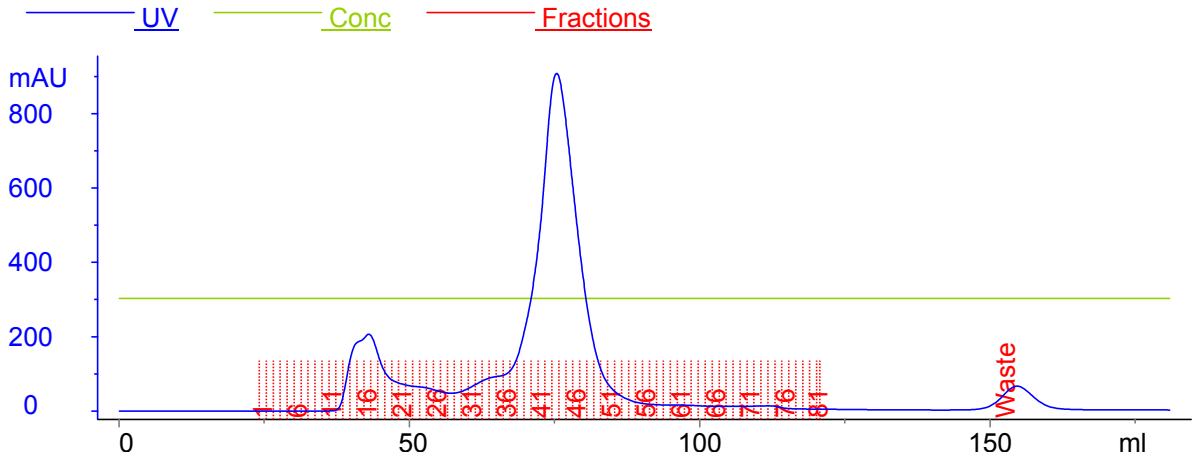
MEK1-Cys218:

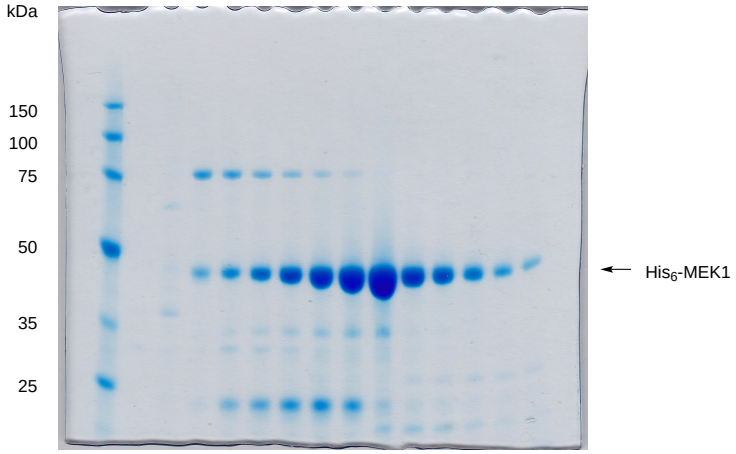
After Ni^{2+} affinity chromatography:



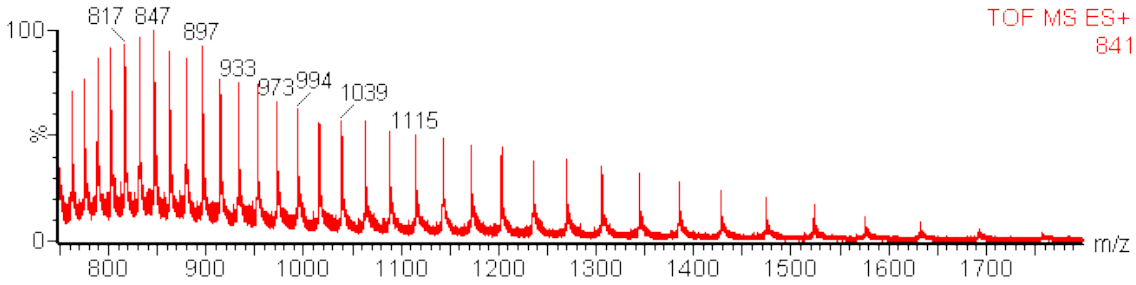


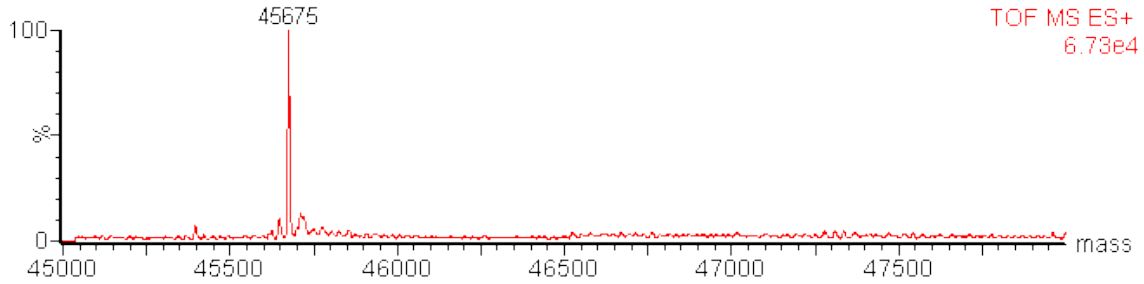
After gel filtration:





MEK1-Cys218: (calculated mass: 45 676, observed mass: 45 675),



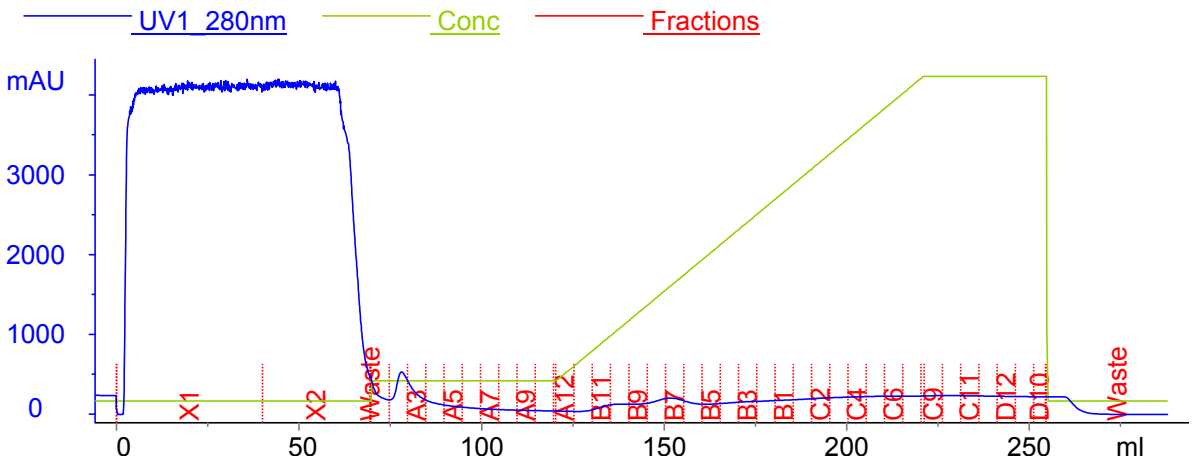


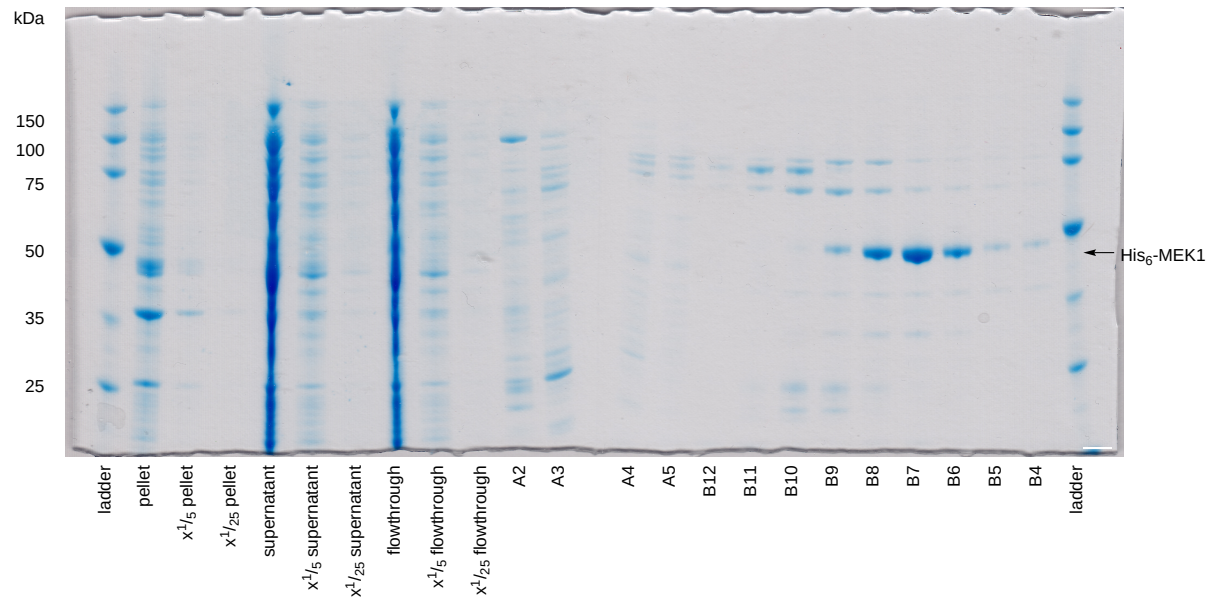
Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                     MFSSHHHHHH SSGLVPRGSH
  1 MPKKKPTPIQ LNPAPDGSAV NGTSSAETNL EALQKKLEEL ELDEQQRKRL
 51 EAFLTQKQKV GELKDDDFEK ISELGAGNGG VVFKVSHKPS GLVMARKLIH
101 LEIKPAIRNQ IIRELQVLHE CNSPYIVGFY GAFYSDGEIS ICMEHMDGGS
151 LDQVLKKAGR IPEQILGKVS IAVIKGLTYL REKHKIMHRD VKPSNILVNS
201 RGEIKLCDFG VSGQLIDCMA NSFVGTRSYM SPERLQGTHY SVQSDIWSMG
251 LSLVEMAVGR YPIPPPDAKE LELMFGSQVE GDAAETPPRP RTPGRPLSSY
301 GMDSRPPMAI FELLDYIVNE PPPKLPSGVF SLEFQDFVNK CLIKNPAERA
351 DLKQLMVHAF IKRSDAEEVD FAGWLSSTIG LNQPSTPTHA AGV
```

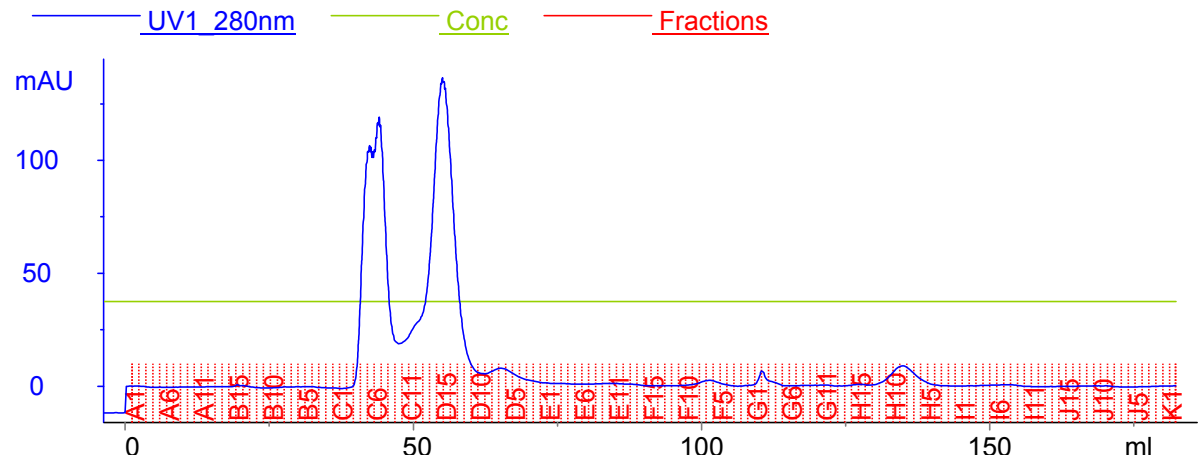
MEK1-S222E:

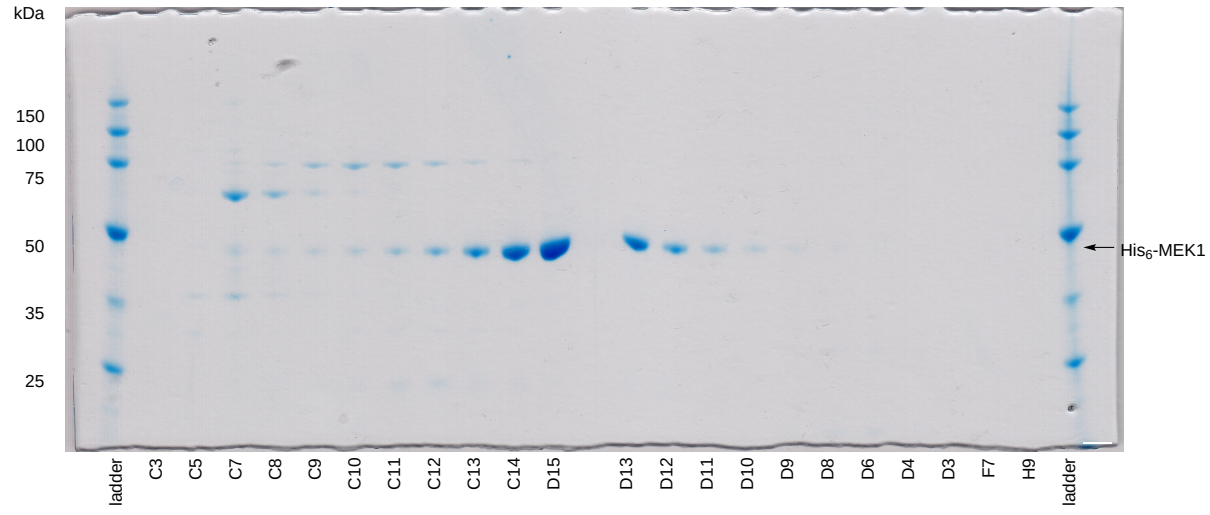
After Ni^{2+} affinity chromatography:



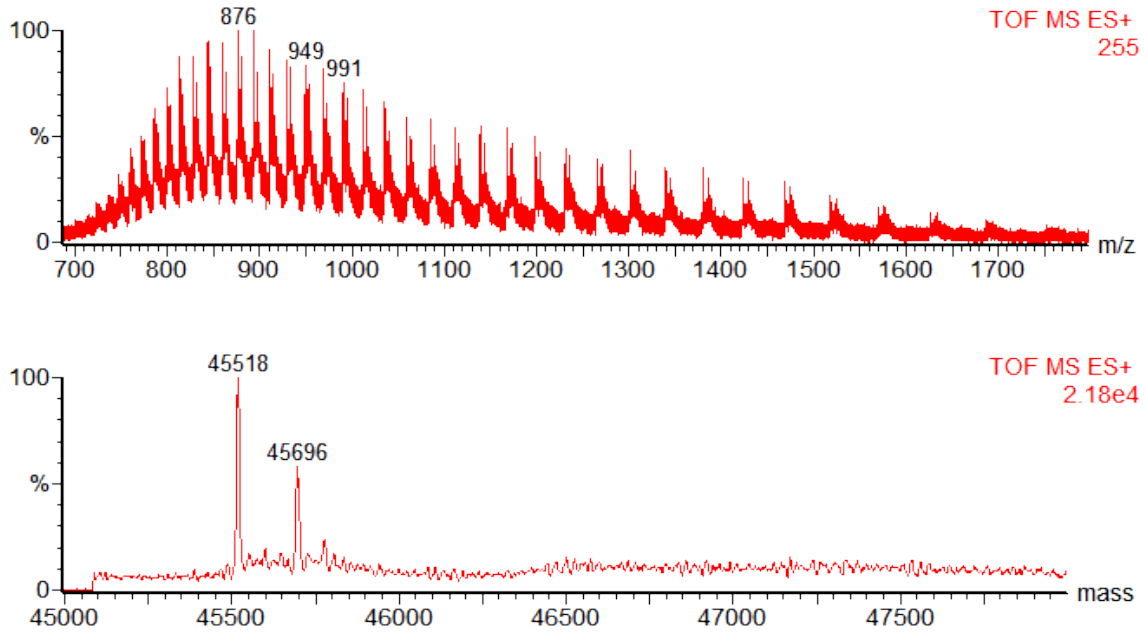


After gel filtration:





MEK1-S222E: (calculated mass: 45 513, observed masses: 45 518 and 45 696, the second one for *N*-terminal gluconylation).



Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                        GSSHHHHHH SSGLVPRGSH
  1 MPKKKPTPIQ LNPAPDGSAV NGTSSAETNL EALQKKLEEL ELDEQQRKRL
 51 EAFLTQKQKV GELKDDDFEK ISELGAGNGG VVFKVSHKPS GLVMARKLIH
101 LEIKPAIRNQ IIRELQVLHE CNSPYIVGFY GAFYSDGEIS ICMEHMDGGS
151 LDQVLKKAGR IPEQILGKVS IAVIKGLTYL REKHKIMHRD VKPSNILVNS
201 RGEIKLCDFG VSGQLIDSMA NEFVGTRSYM SPERLQGTHY SVQSDIWSMG
251 LSLVEMAVGR YPIPPPDAKE LELMFGCQVE GDAAETPPRP RTPGRPLSSY
301 GMDSRPPMAI FELLDYIVNE PPPKLPSGVF SLEFQDFVNK CLIKNPAERA
351 DLKQLMVHAF IKRSDAEEVD FAGWLCSTIG LNQPSTPTHA AGV
```

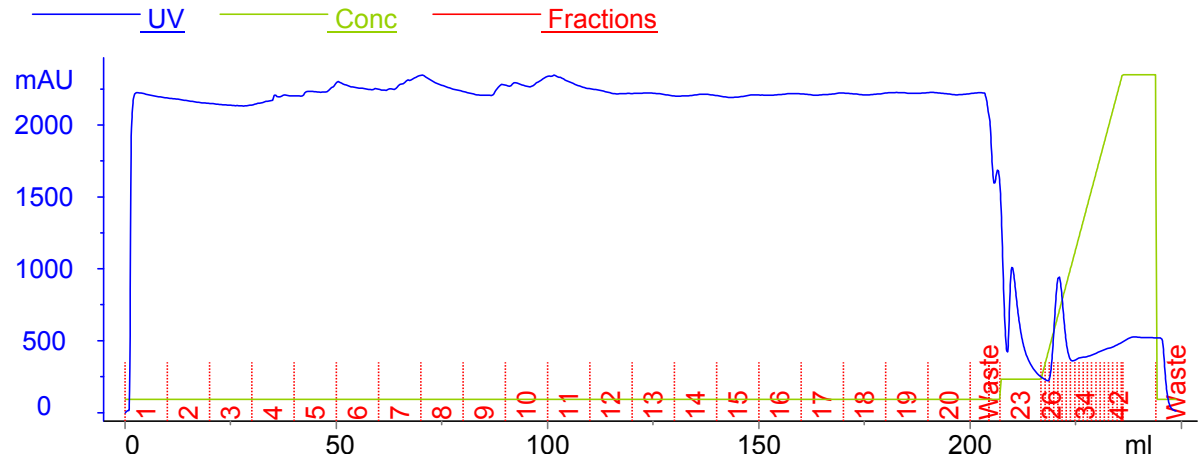
Protein Purification of MEK1-Cys218Cys222

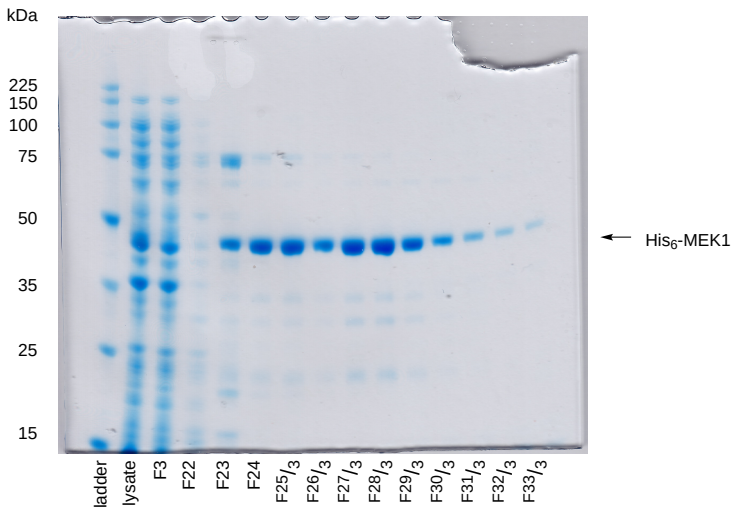
MEK1-Cys218Cys222 was purified using a modified protocol from the one already established for previous MEK1 variants.

Cell Lysis: The cell pellets were combined, thawed on ice and re-suspended (160 mL) in ice cold MEK1 Ni^{2+} affinity lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 5% glycerol, 0.5 mM TCEP) with stirring at 4 °C. Lysozyme (170 mg) was added and the mixture stirred for a further 2 h before sonication using a sonicator equipped with a microtip (25×2 s blasts, 60% amplitude, 2128 J total energy). DNase (1 mg) was added and the mixture stirred for a further 1 h before the lysate was clarified

by centrifugation (20 000 rpm, JA25.50 rotor, 45 min, 4 °C). The lysate was further clarified by sequential filtration through 0.8 µm and 0.2 µm syringe filters.

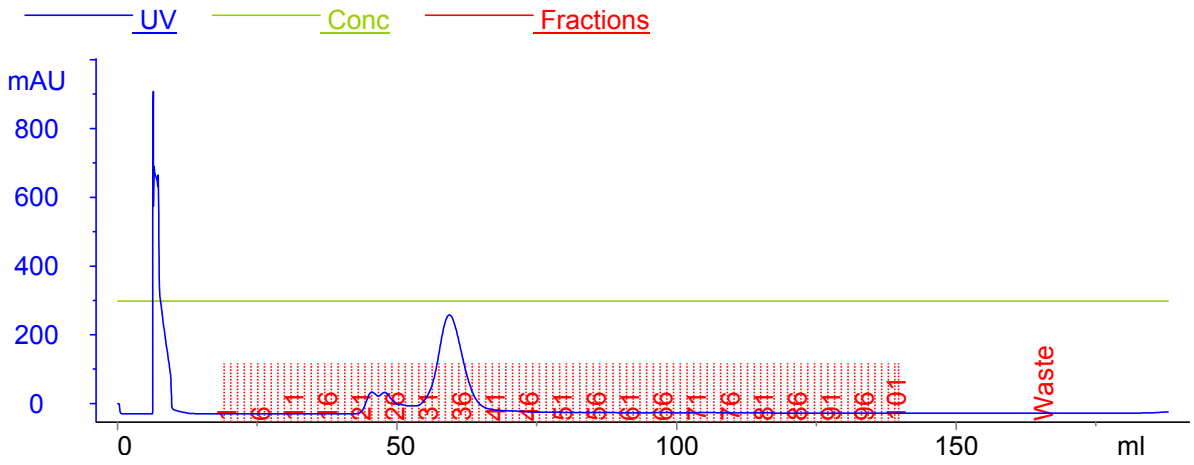
Ni^{2+} Affinity Chromatography: Using an ÄKTA Purifier™ FPLC system, the clarified lysate was loaded (155 mL) onto a pre-packed 1 mL HisTrap™ HP Ni^{2+} affinity column, chasing through (3.5 mL, 3.5 CV) with more lysis buffer. The column was washed (10 mL, 10 CV) with MEK1 Ni^{2+} affinity wash buffer (same as lysis buffer but with 50 mM imidazole) and the bound protein eluted using a linear gradient (20 mL, 20 CV, 1 mL fractions) from wash buffer to MEK1 Ni^{2+} affinity elution buffer (same as lysis buffer but with 500 mM imidazole). The fractions containing the protein of interest were eluted into collection tubes already containing (2 mL each) lysis buffer. The fractions containing the desired protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and combined.

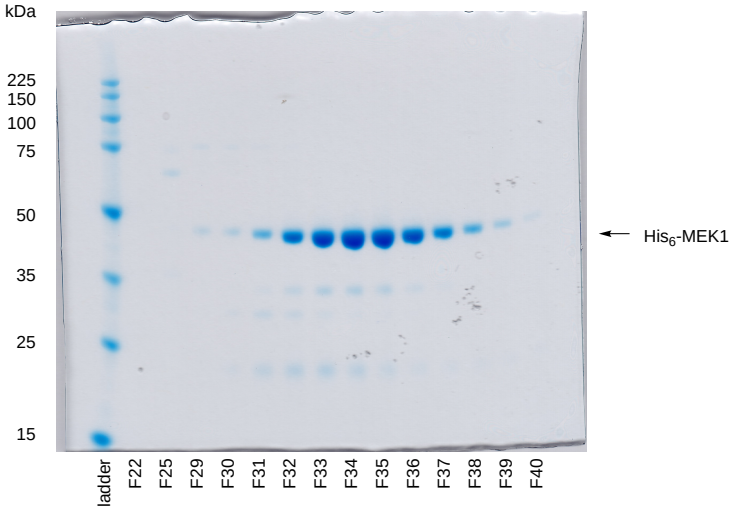




Size Exclusion Chromatography and Storage: The collected fractions were concentrated by Vivaspin (MWCO 10 000) to a final volume ∼1 mL before filtering through a syringe filter. Chasing through

with MEK1 storage buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP), the concentrated proteins were loaded onto a HiLoad™ 16/60 Superdex 200 gel filtration column and filtered (180 mL, 1.5 CV, 1.2 mL fractions) into MEK1 storage buffer. The fractions containing the desired protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and collected. The collected fractions were re-concentrated by Vivaspin (MWCO 10 000, 3.11 mg/mL final concentration), divided into aliquots (27×61 µL), flash frozen in liquid nitrogen and stored at −80 °C: Yield: 5.1 mg in MEK1 storage buffer.



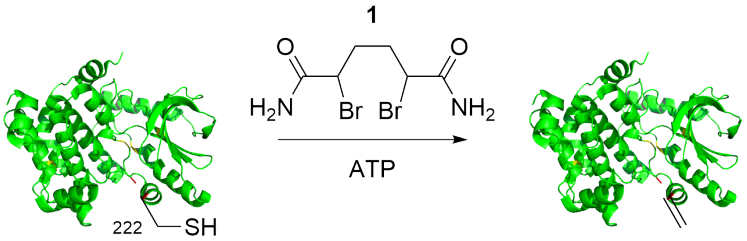


Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                     MFSSHHHHHH SSGLVPRGSH
  1 MPKKKPTPIQ LNPAPDGSAV NGTSSAETNL EALQKKLEEL ELDEQQRKRL
 51 EAFLTQKQKV GELKDDDFEK ISELGAGNGG VVFKVSHKPS GLVMARKLIH
101 LEIKPAIRNQ IIRELQVLHE CNSPYIVGFY GAFYSDGEIS ICMEHMDGGS
151 LDQVLKKAGR IPEQILGKVS IAVIKGLTYL REKHKIMHRD VKPSNILVNS
201 RGEIKLCDFG VSGQLIDCMA NCFVGTRSYM SPERLQGTHY SVQSDIWSMG
251 LSLVEMAVGR YPIPPPDAKE LELMFGSQVE GDAAETPPRP RTPGRPLSSY
301 GMDSRPPMAI FELLDYIVNE PPPKLPSGVF SLEFQDFVNK CLIKNPAERA
351 DLKQLMVHAF IKRSDAEEVD FAGWLSSTIG LNQPSTPTHA AGV
```

4.11.4 Chemical Modification: Optimised Procedures

Synthesis of MEK1-Dha222 (24)

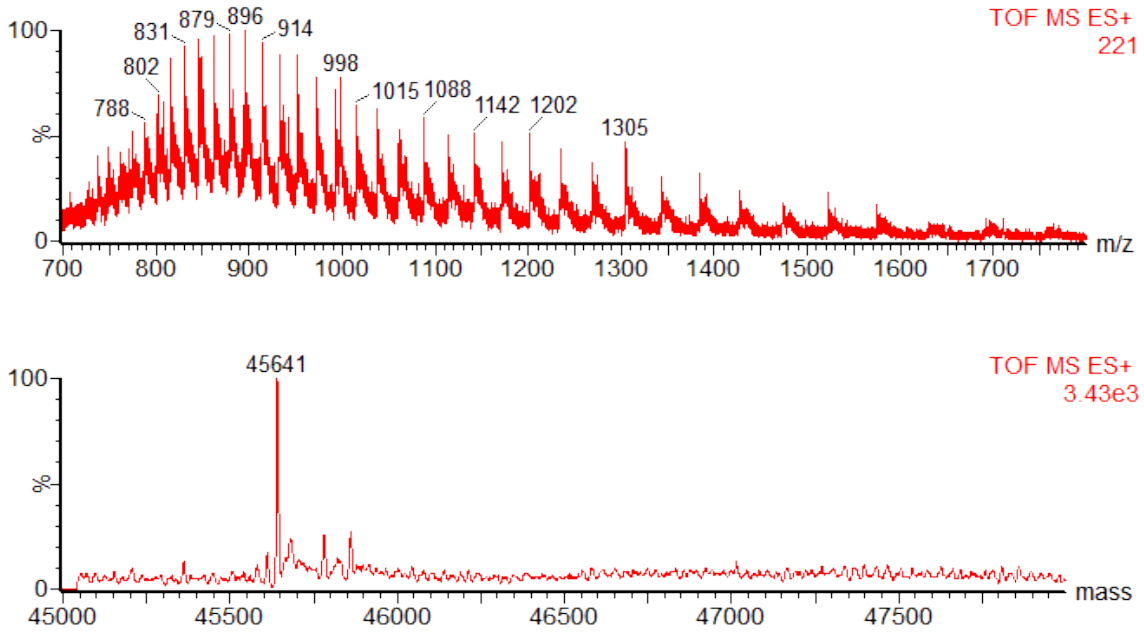


23 **24**

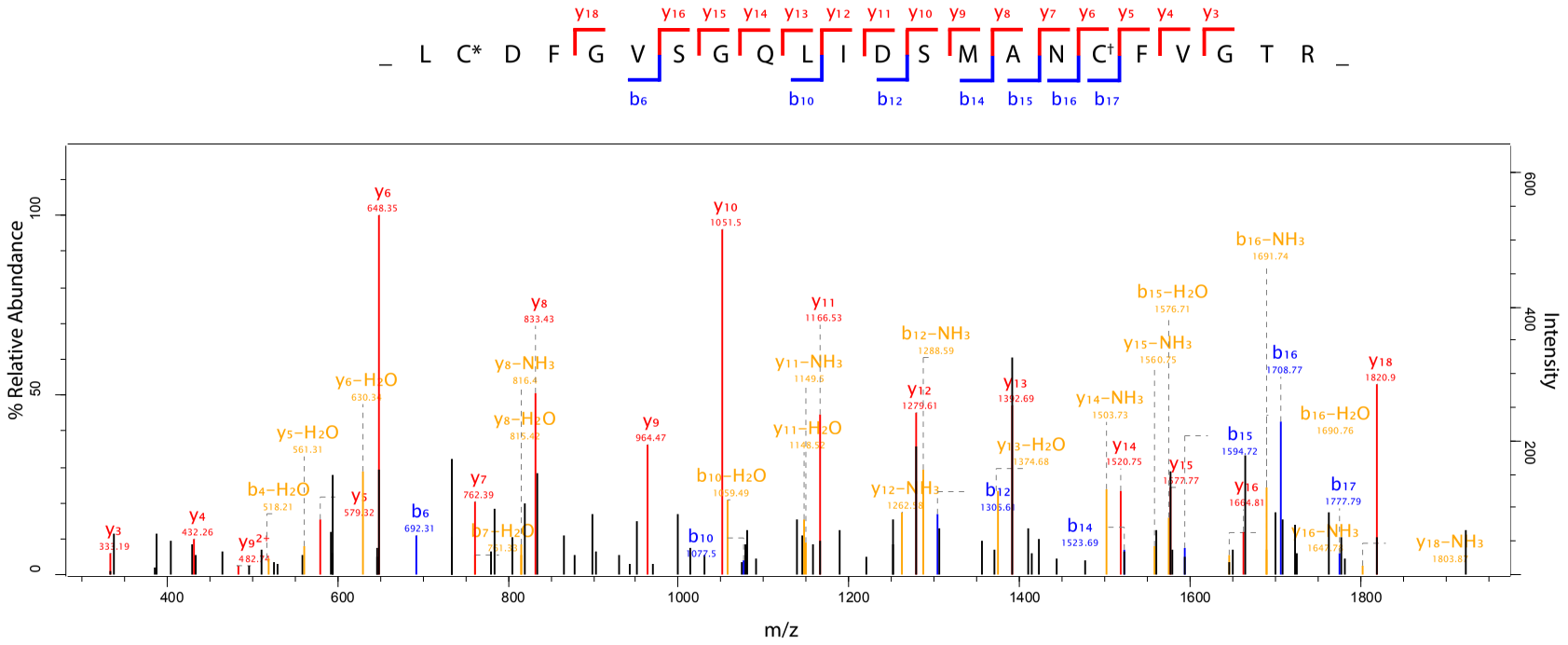
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

MEK1-Cys222 (**23**) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 10% glycerol) using a G-25 MiniTrap™ desalting column (GE Healthcare). After filtration of the protein, the concentration of the protein was adjusted with reaction buffer. The diluted, desalted MEK1-Cys222 (570 µL of a 1.56 mg/mL solution, 19.5 nmol) was divided into aliquots (3×190 µL). To each aliquot was added ATP (29.6 µL of a 100 mM solution in water, pH 7.0, final concentration: 10 mM) and the mixtures incubated at room temperature for 30 min. Dibromide **1** (76 µL of a 3.9 mg/mL suspension in reaction buffer, 0.98 µmol, 150 eq.) was then added to each reaction aliquot and the mixtures incubated with shaking at 37 °C, 600 rpm for 8 h. LC-MS of a re-combined sample after this time showed >85% conversion to the Dha product. The reaction aliquots were combined and desalted again using a G-25 MidiTrap™ desalting column (GE Healthcare), followed by repeated concentration/dilution by

Vivaspin (MWCO 10 000) to give MEK1-Dha222, which was either kept on ice for immediate use, or flash frozen and stored at −80 °C. LC-MS/MS analysis of a digested sample prepared using the method as described in Section 3.8.7 showed that Dha had been installed at the desired site: Yield: 380 µL of a 1.34 mg/mL solution in MEK1 reaction buffer, 51%, (calculated mass: 45 642, observed mass: 45 641).

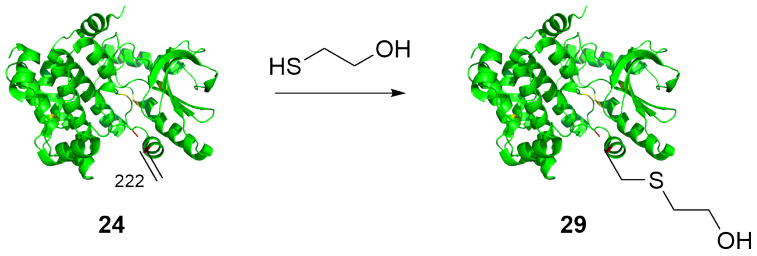


Raw file	Scan	Method	Score	Mass	Gene names
ORE01_RR_131220_MEK1_Dha_FASP_02	5763	ITMS; CID	97.15	2355.09	MAP2K1



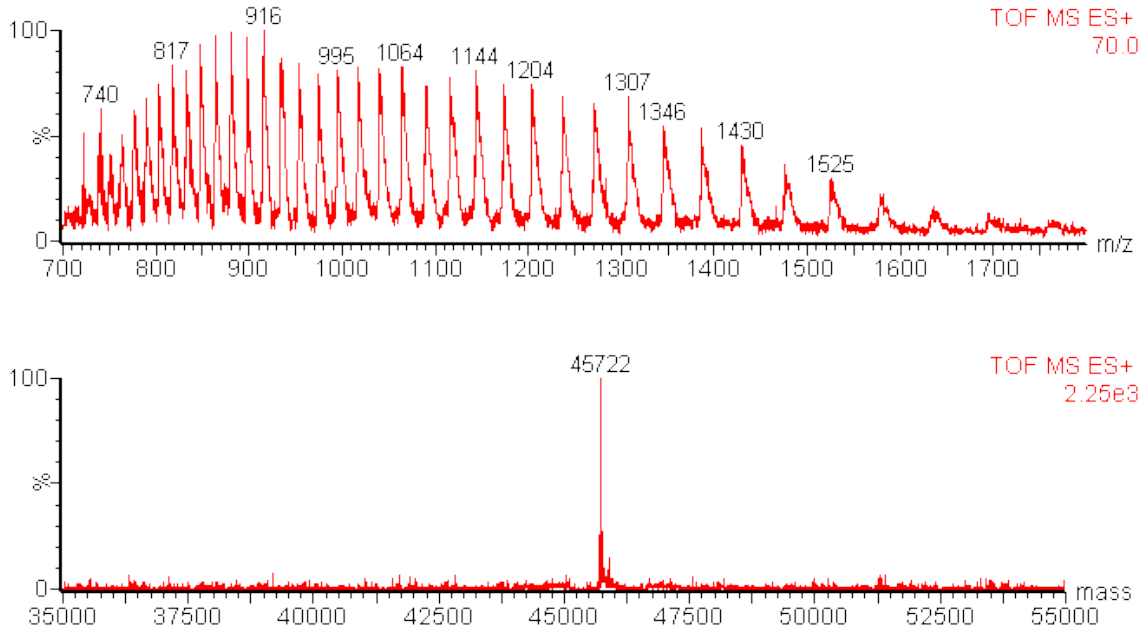
C* denotes Carbamidomethyl modification; C† denotes Dehydroalanine (Dha) modification

Synthesis of MEK1-Ethanolylcysteine222 (29)

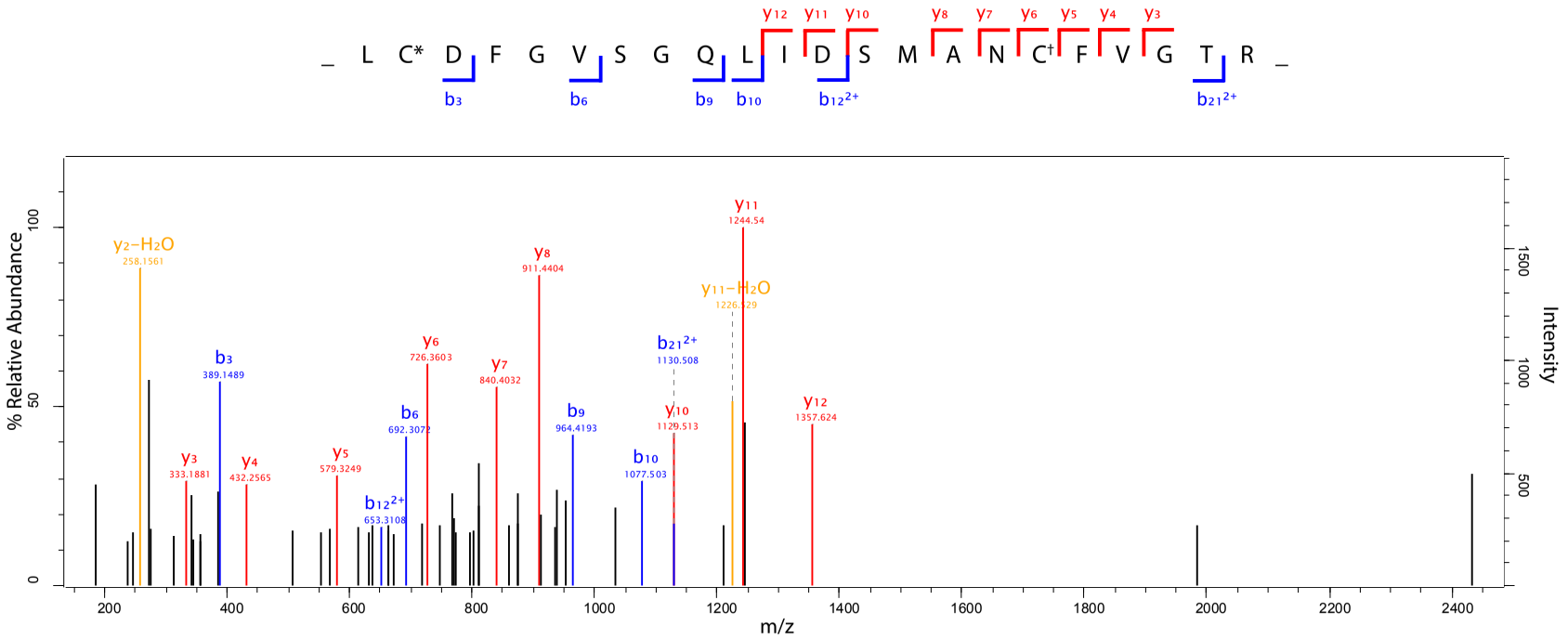


24 **29**

In collaboration with Sébastien Galan, to demonstrate that the installed Dha was reactive towards nucleophilic thiols, MEK1-Dha222 (**24**) was treated with β-mercaptoethanol. To an aliquot of MEK1-Dha222 (50 µL of a 1.0 mg/mL solution, 1.1 nmol, 1.0 eq.) in MEK1 reaction buffer was added β-meracaptoethanol (1 µL of a 10% v/v solution in MEK1 reaction buffer, 1.4 µmol, 1260 eq.). The mixture was incubated at room temperature for 1 h. LC-MS after this time showed complete conversion to MEK1-ethanolylcysteine222. LC-MS/MS analysis of a digested sample prepared using the method as described in Section 3.8.7 showed that the adduct was installed in the desired position: (calculated mass: 45 720, observed mass: 45 722)

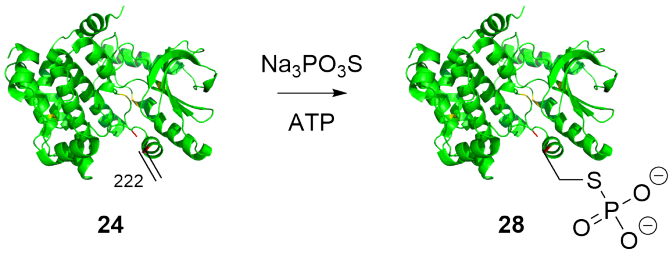


Raw file	Scan	Method	Score	Mass	Gene names
ORE01_RR_131218_MEK_bME_01	3181	FTMS; HCD	57.48	2433.11	MAP2K1



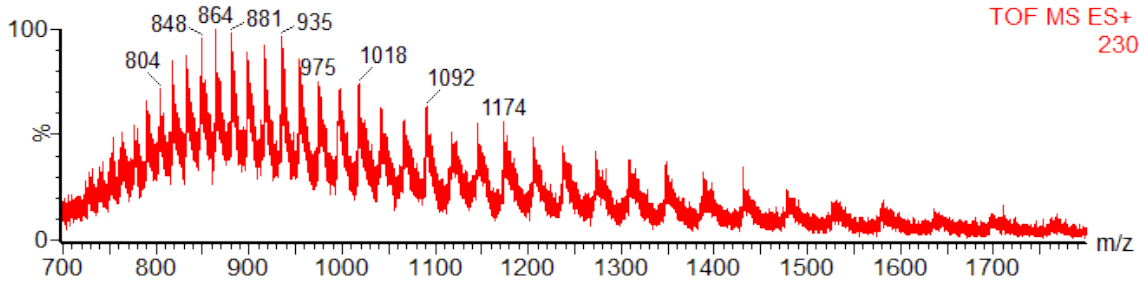
C* denotes Carbamidomethyl modification; C† denotes β-Mercaptoethanol adduct with Dehydroalanine (Dha) modification

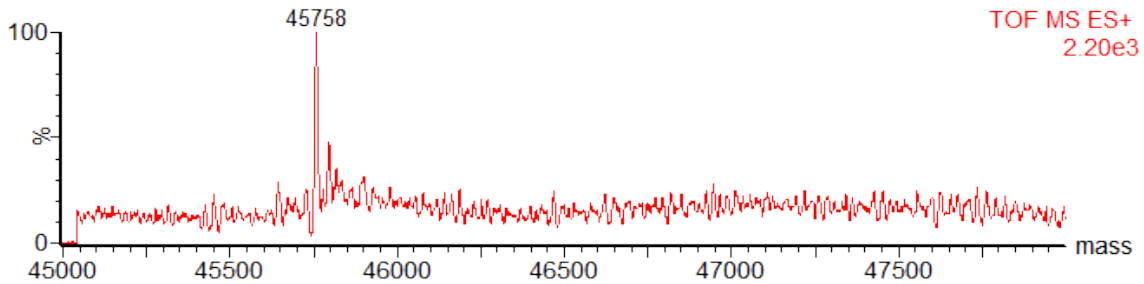
Synthesis of MEK1-pCys222 (28)



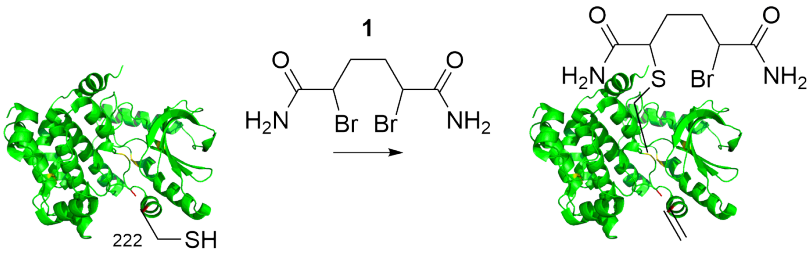
24 28

To MEK1-Dha222 (**24**) (412 µL of a 1.04 mg/mL solution, 9.4 nmol, 1.0 eq.) in MEK1 reaction buffer was added ATP (11.8 µL of a 500 mM solution in water, pH 7.0, final concentration: 10 mM). Separately, sodium thiophosphate (96.45 mg) was dissolved in water (79.9 µL) and HCl (41.9 µL of a 5 M solution) added to adjust the pH to 8.0. The thiophosphate solution (4×16.0 µL of a 530 mg/mL solution, 4×47 µmol, 4×5000 eq.) was added batchwise (5 min intervals) to the protein and the resultant mixture incubated at room temperature for 18 h. LC-MS after this time showed >85% conversion to the thiophosphate adduct. The reaction mixture was desalted by dialysis against MEK1 reaction buffer to give MEK1-pCys222, which was either kept on ice for immediate use, or flash frozen stored at −80 °C: Yield: 670 µL of a 0.46 mg/mL solution in MEK1 reaction buffer, 72%, (calculated mass: 45 756, observed mass: 45 758).



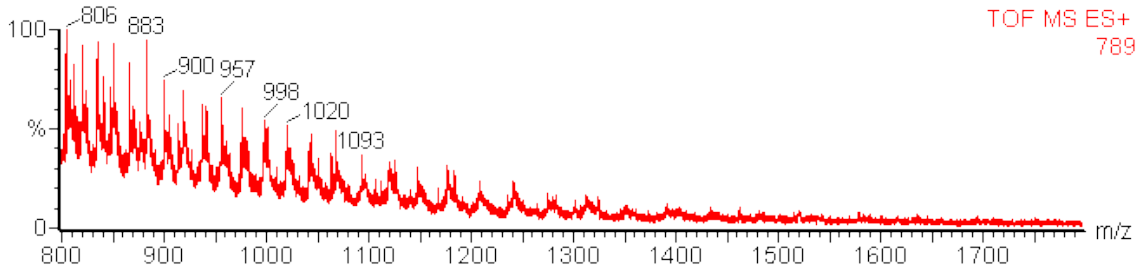


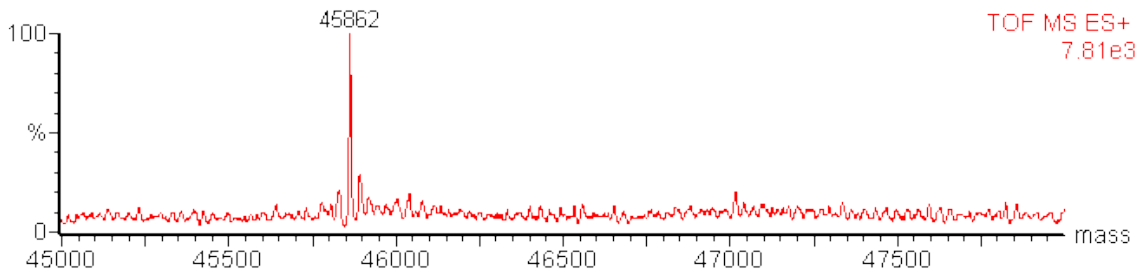
Synthesis of Over-alkylated MEK1-Dha222



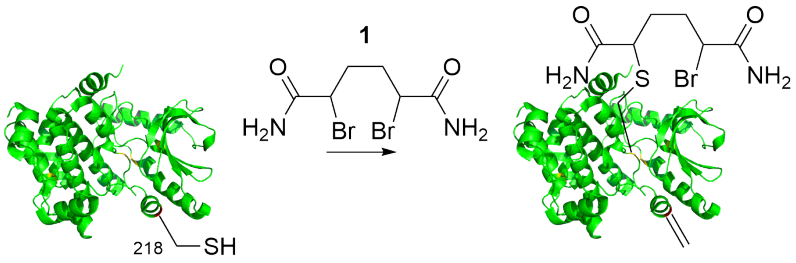
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

An aliquot of MEK1-Cys222 (**23**) (45 µL of a 4.1 mg/mL solution, 4.1 nmol, 1.0 eq.) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol) using a G-25 SpinTrap™ desalting column (GE Healthcare). To the resulting solution (140 µL) was added a suspension of **1** (46.9 µL of a 3.9 mg/mL suspension, 0.61 µmol, 150 eq.) in reaction buffer. The mixture was incubated with shaking at 37 °C, 500 rpm and the reaction monitored by LC-MS. LC-MS after 1 h showed conversion to the single Dha product and single bromide adduct, while LC-MS after 2 h showed the emergence of a further peak with mass corresponding to the formation of a further bromide adduct, postulated to be **25**. LC-MS after 6 h showed >90% conversion to the postulated over-alkylated product **25**; (calculated mass: 45 863, observed mass: 45 862).



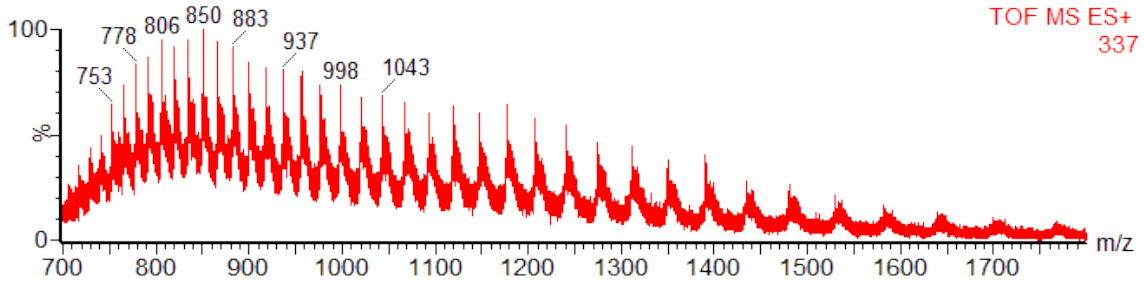


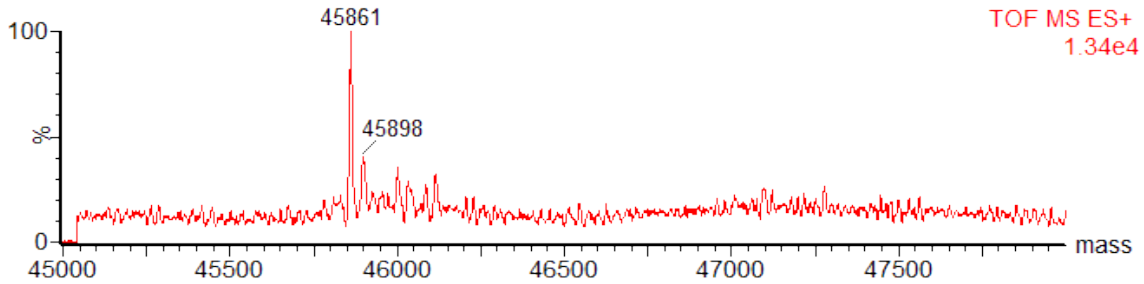
Synthesis of Over-alkylated MEK1-Dha218



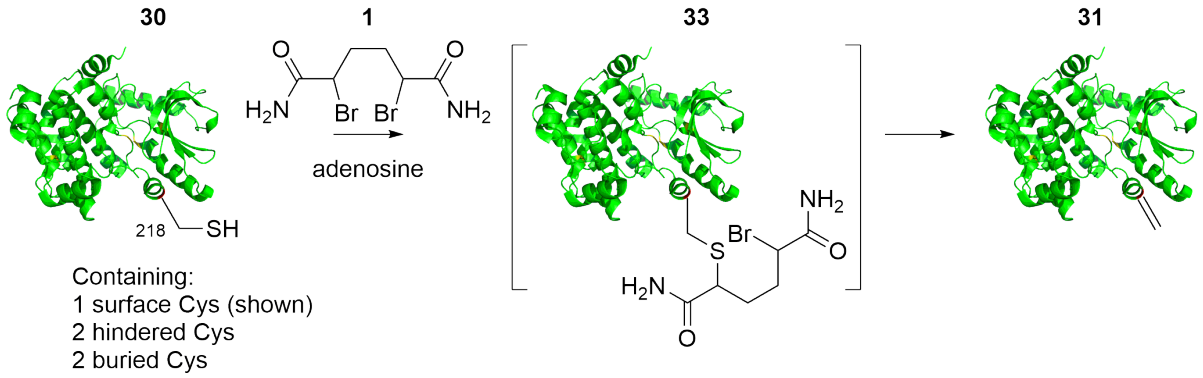
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

An aliquot of MEK1-Cys218 (**30**) (50 µL of a 4.0 mg/mL solution, 4.4 µmol, 1.0 eq.) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) to MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol). A suspension of **1** (50.9 µL of a 3.9 mg/mL suspension, 0.66 µmol, 150 eq.) was added and the mixture shaken at 37 °C, 550 rpm for 6 h. LC-MS analysis after this time showed ~60% conversion to the single Dha, single bromide adduct **32**: (calculated mass: 45 863, observed mass: 45 861).

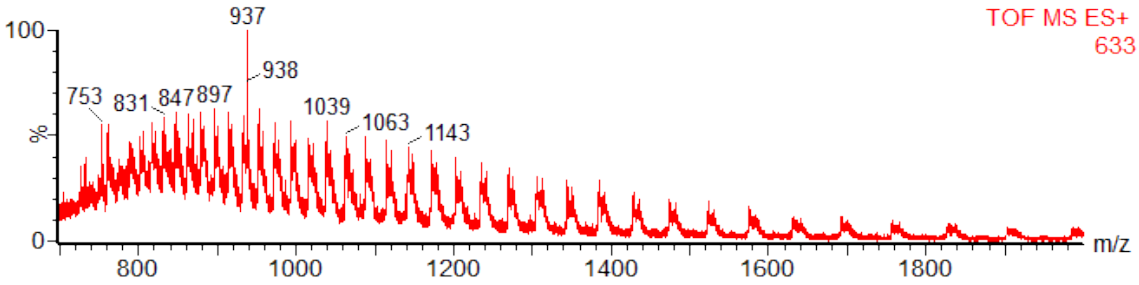


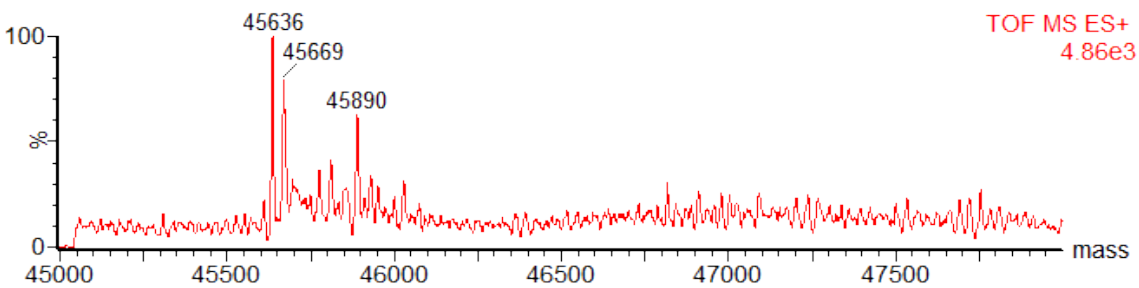


Reaction Procedure Conferring Highest Conversion to MEK1-Dha218



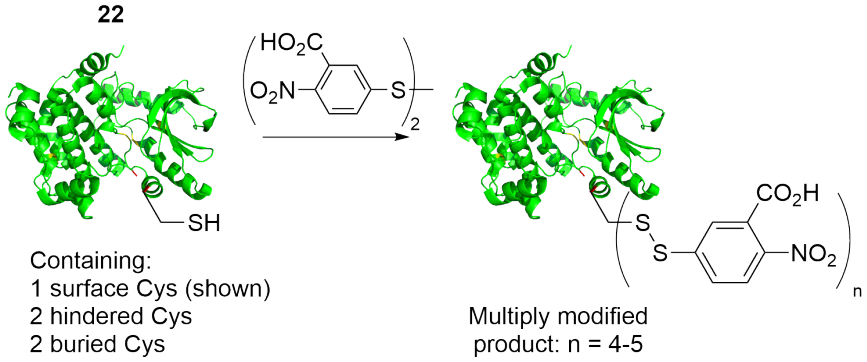
MEK1-Cys218 (**30**) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (same as MEK1 storage buffer but with 50 mM HEPES pH 8.0 and no TCEP) using a G-25 MiniTrap™ desalting column (GE Healthcare). The resultant protein solution (60 µL of a 1.62 mg/mL solution, 2.1 nmol, 1.0 eq.) was filtered and further diluted with MEK1 reaction buffer (13.6 µL). Adenosine (1.0 µL of a 800 mM) in DMSO was added to the protein and the mixture incubated at room temperature for 30 min. Dibromide **1** (24.7 µL of a 3.9 mg/mL suspension in reaction buffer, 0.32 µmol, 150 eq.) was added and the mixture incubated with shaking at 37 °C, 600 rpm. LC-MS after 8 h showed that >60% of the starting MEK1-Cys218 had been consumed, forming either MEK1-Dha218 (**31**) or the mono-bromide intermediate **33**: **31**: (calculated mass: 45 642, observed mass: 45 636), **33** (calculated mass: 45 897, observed mass: 45 890).



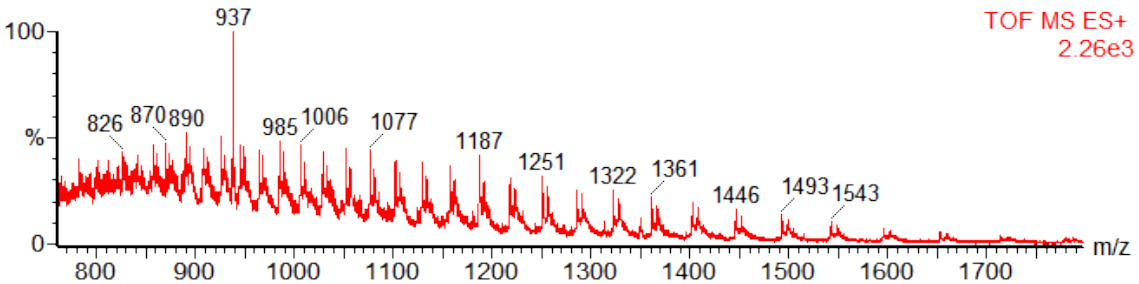


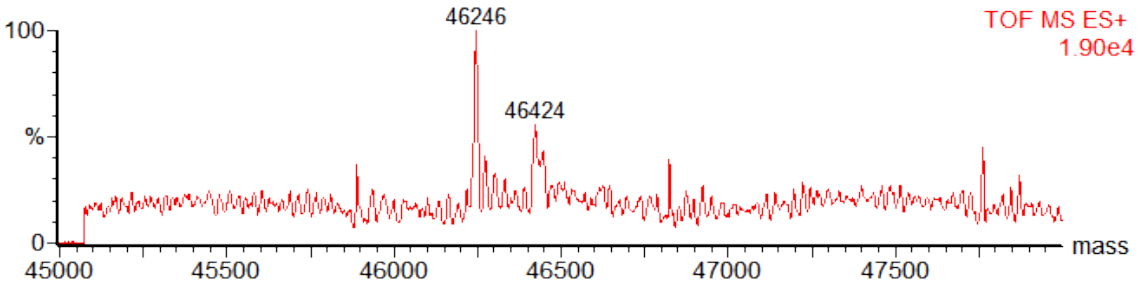
4.11.5 Chemical Modification: Control Experiments

Treatment of MEK1-S222C/C277S/C376S (22) with Ellman's Reagent

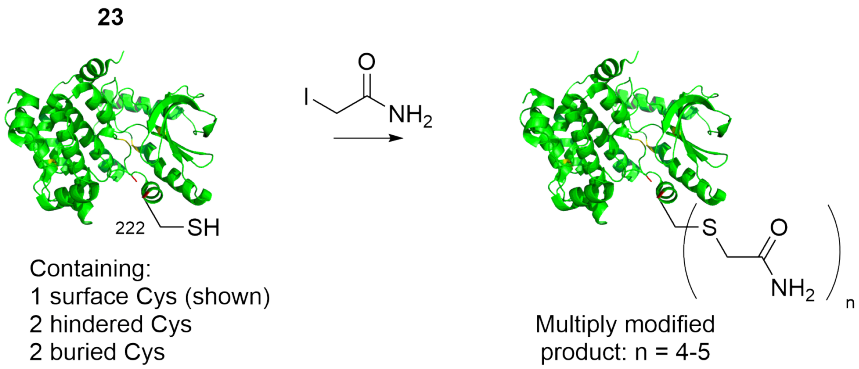


To a freshly thawed solution of MEK1-S222C/C277S/C376S (without G(−19)F mutation, **22**) (60 µL of a 2.4 mg/mL solution, 3.2 nmol, 1.0 eq.) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was added a solution of Ellman's reagent (61 µL of a 2.1 mg/mL solution, 0.32 µmol, 100 eq.) in HEPES buffer (50 mM HEPES pH7.5). The mixture was mixed end-over-end at 4 °C for 30 min before a sample (15 µL) was diluted with HEPES buffer (35 µL) and analysed by LC-MS. LC-MS showed masses corresponding to modification of 4–5 of the 5 cysteines of the protein with Ellman's reagent; 4 modifications (calculated masses: 46 243 and 46 421, observed masses: 46 246 and 46 424); 5 modifications (calculated mass: 46 440, observed mass: 46 444).

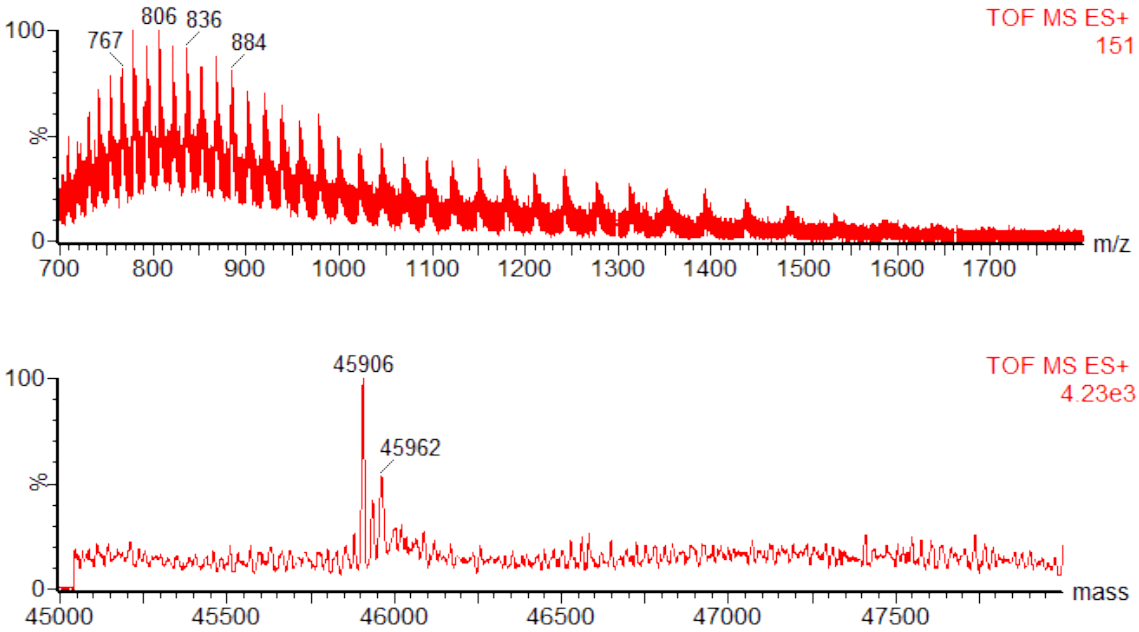




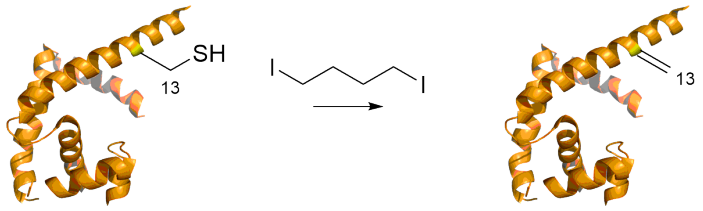
Treatment of MEK1-Cys222 (23) with Iodoacetamide



An aliquot of MEK1-Cys222 (**23**) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) to MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol). To the resulting protein solution (140 µL of a 2.3 mg/mL solution, 7.0 nmol, 1.0 eq. was added iodoacetamide (70.5 µL of a 100 mM, 7.0 µmol, 1000 eq.) in the same buffer. The reaction tube was wrapped in foil before being mixed end-over-end at room temperature for 3 h, monitoring reaction progress by LC-MS. LC-MS after this time showed products with masses corresponding to the 4– and 5–alkylated products in a 2:1 ratio: (calculated masses: 45 904 and 45 961, observed masses: 45 906 and 45 962).

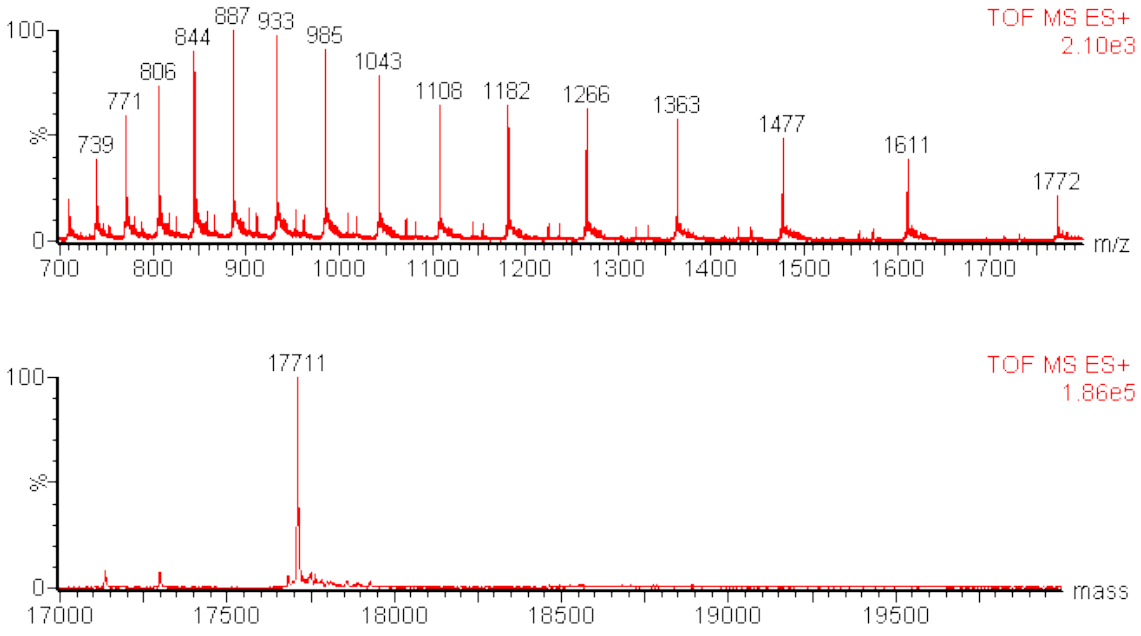


Synthesis of SarZ-Dha13



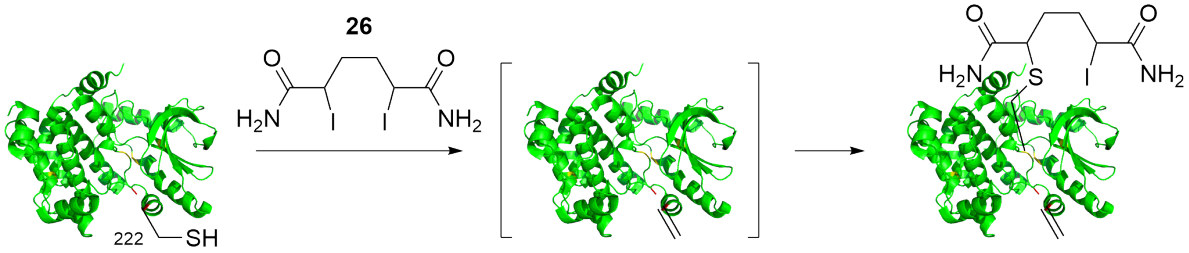
34 **35**

An aliquot of SarZ (**34**) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) into phosphate buffer (50 mM sodium phosphate pH 8.0). To the resulting protein solution (120 µL of a 1.0 mg/mL solution, 6.8 nmol, 1.0 eq.) was added 1,4-diiodobutane (1.34 µL, 10.1 µmol, 1500 eq.) and the mixture shaken with heating at 37 °C, 600 rpm for 2 h. A lower stoichiometric amount of 1,4-diiodobutane (0.2 µL, 1.5 µmol, 225 eq.) was also tried with a longer reaction time (3 h). Both reactions were followed by LC-MS and reaction kinetics analysed from the data collected. At the endpoint, both reactions showed >90% conversion to the Dha product **35**: (calculated mass: 17 710, observed mass: 17 711).



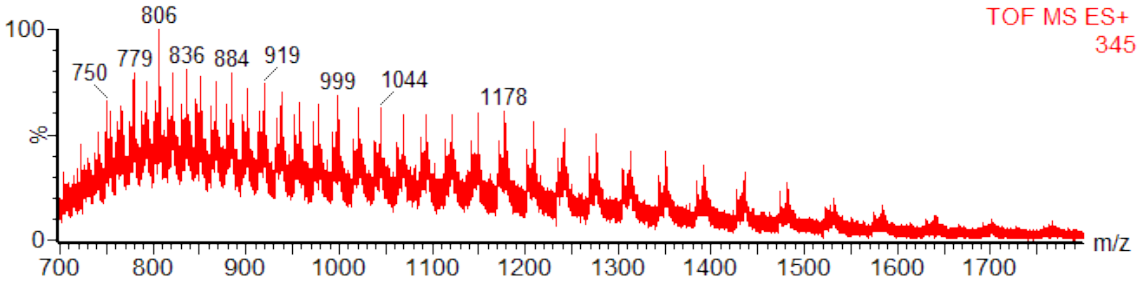
4.11.6 Chemical Modification: Optimisation Experiments on MEK1-Cys222

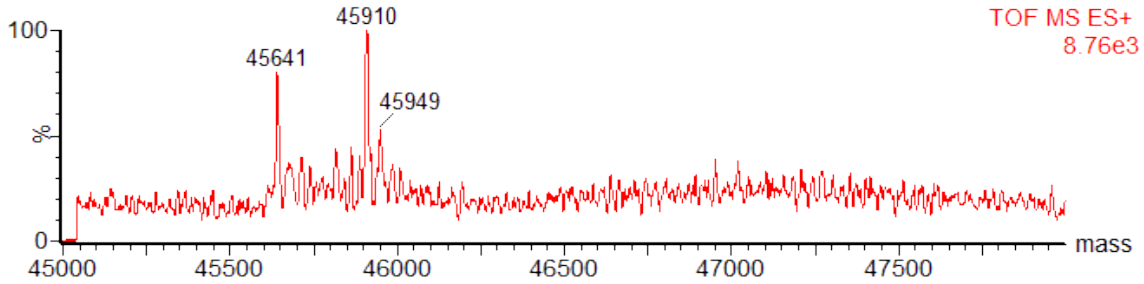
Treatment of MEK1-Cys222 (23) with Diiodide 26



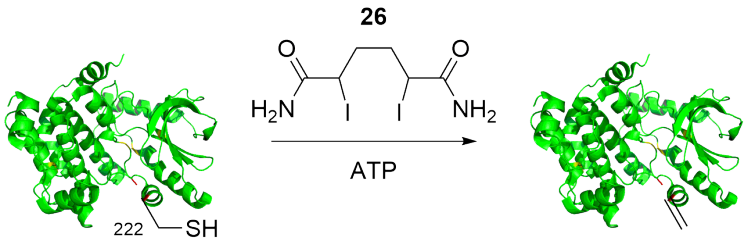
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

An aliquot of MEK1-Cys222 (**23**) (65 µL of a 3.1 mg/mL solution, 4.4 nmol, 1.0 eq.) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) into MEK1 reaction buffer (the same as storage buffer but with 50 mM HEPES pH 8.0 and without TCEP). A suspension of diiodide **26** (51.4 µL of a 5.1 mg/mL suspension, 0.66 µmol, 150 eq.) in MEK1 reaction buffer was added and the mixture shaken at 37 °C, 550 rpm for 6 h while the reaction was monitored by LC-MS. LC-MS at the 6 h endpoint showed some remaining MEK1-Dha222 (**24**, ~30% conversion) but with ~40% conversion, the predominant product was the single Dha, single iodide adduct **36**: (calculated mass: 45 910, observed mass: 45 910).





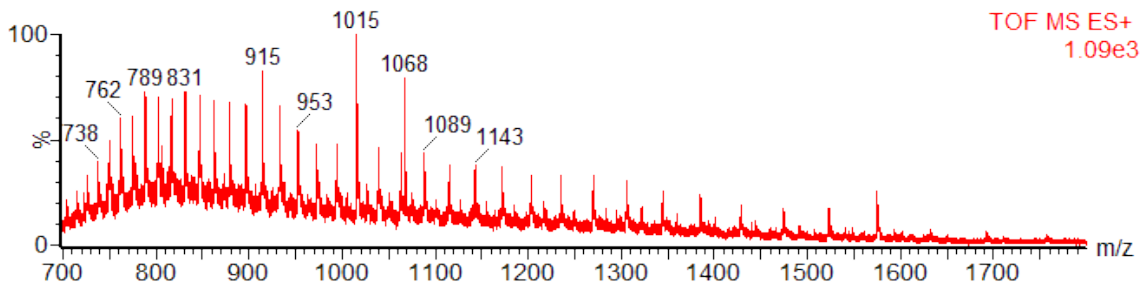
Treatment of MEK1-Cys222 (23) with Diiodide 26 and Substrate Protection

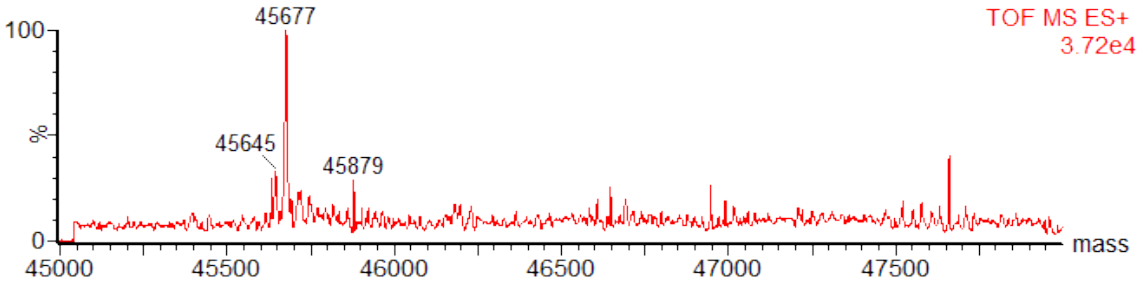


23 **24**

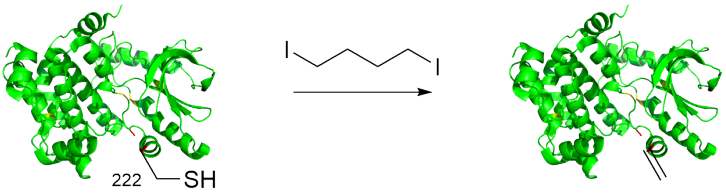
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

An aliquot of MEK1-Cys222 (**23**) (65 µL of a 3.1 mg/mL solution, 4.4 nmol, 1.0 eq.) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) to MEK1 reaction buffer (the same as storage buffer but with 50 mM HEPES pH 8.0 and without TCEP). A solution of ATP (21.3 µL of a 100 mM solution at ~pH 7, 10 µM final concentration) in the same buffer was added and the mixture incubated at room temperature for 5 min. Finally, a suspension of diiodide **26** (51.4 µL of a 5.1 mg/mL suspension, 0.66 µmol, 150 eq.) in the same buffer was added and the mixture shaken at 37 °C, 550 rpm for 6 h while the reaction was monitored by LC-MS. LC-MS at the 6 h endpoint showed ~20% conversion to the single Dha product **24**: (calculated mass: 45 642, observed mass: 45 645).





Treatment of MEK1-Cys222 (23) with 1,4-Diiodobutane

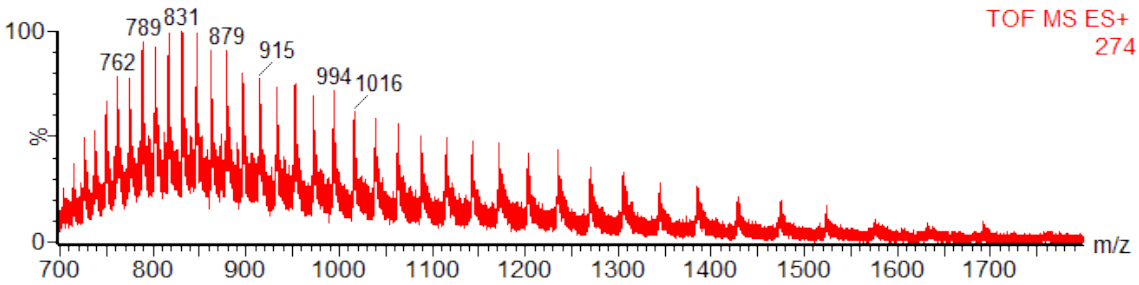


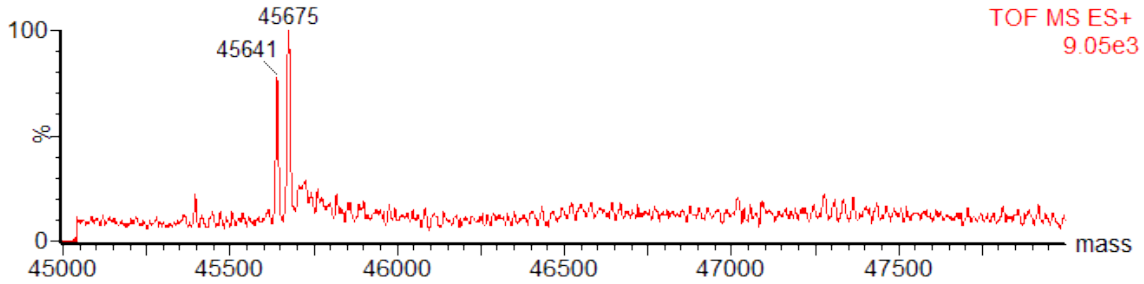
23

Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

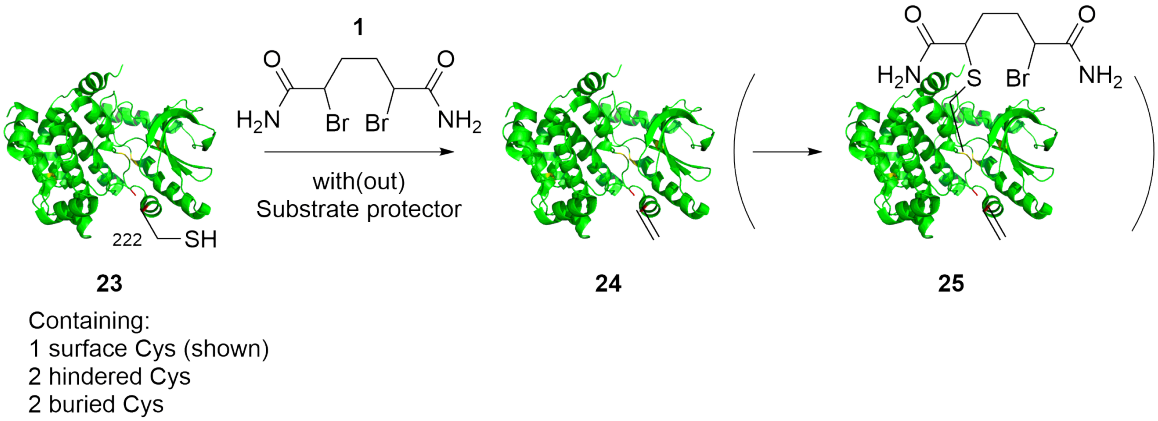
24

An aliquot of MEK1-Cys222 **23** was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) to phosphate buffer (50 mM sodium phosphate pH 8.0). To the resulting protein solution (140 µL of a 1.8 mg/mL solution, 5.5 nmol, 1.0 eq.) was added 1,4-diiodobutane (1.1 µL, 8.3 µmol, 1500 eq.) and the mixture shaken at 37 °C, 550 rpm for 6 h. The mixture remained heterogeneous during the whole period of the reaction. LC-MS after this time showed ~40% to the single Dha product **24**: (calculated mass: 45 642, observed mass: 45 641).





General Conditions for Screening Substrate Protecting Agents during MEK1-Dha222 Formation

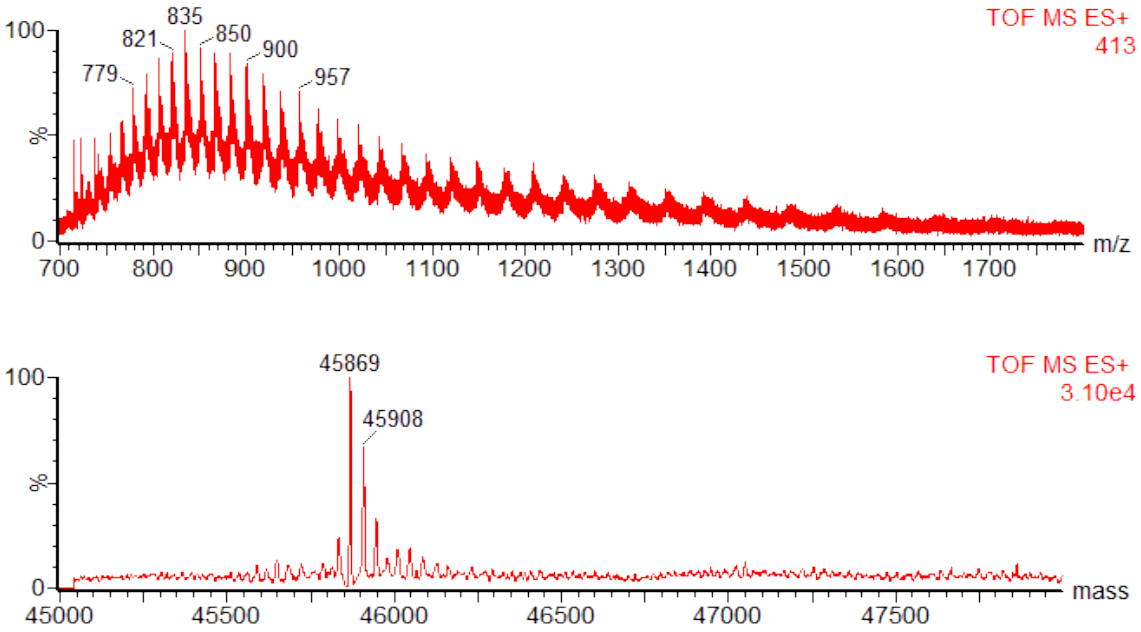


MEK1-Cys222 (**23**) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (same as MEK1 storage buffer but with 50 mM HEPES pH 8.0 and no TCEP) using a G-25 SpinTrap™ desalting column (GE Healthcare). The concentration of the protein was adjusted with reaction buffer. To the diluted, desalted MEK1-Cys222 (65 µL of a 1.55 mg/mL solution, 2.2 nmol, 1.0 eq.) was added substrate protector (one of water, AMP, ADP or ATP) (10.1 µL of 100 mM protector stock, pH 7.0) and the mixture incubated at room temperature for 30 min. Dibromide **1** (25.6 µL of a 3.9 mg/mL suspension in reaction buffer, 0.33 µmol, 150 eq.) was then added and the mixture incubated with shaking at 37 °C, 600 rpm for 6 h. LC-MS after this time showed >80% conversion to the Dha product **24** in all cases (calculated mass: 45 642, observed mass: 45 648), except for when no substrate protector was used. In this case, further alkylation to the single Dha, single bromide adduct **32** was observed (calculated mass: 45 863, observed mass: 45 869).

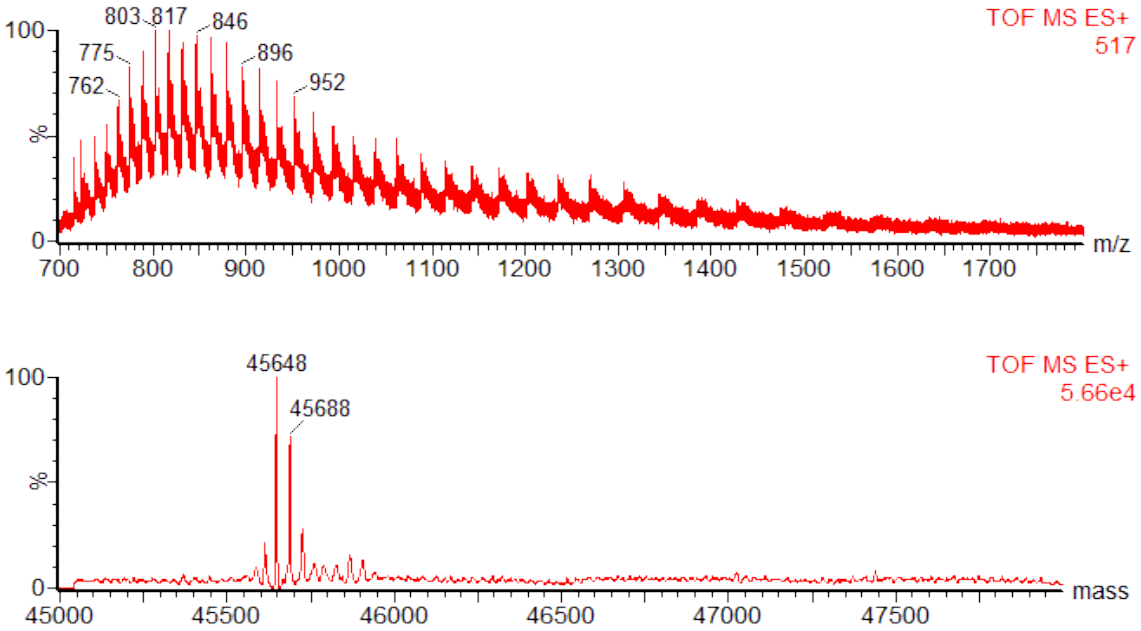
Formation of precipitate was also observed after this time. The precipitate was pelleted by centrifugation (17 000 × *g*, 15 min, 4 °C) and the supernatant carefully decanted by pipette, ensuring that the same volume

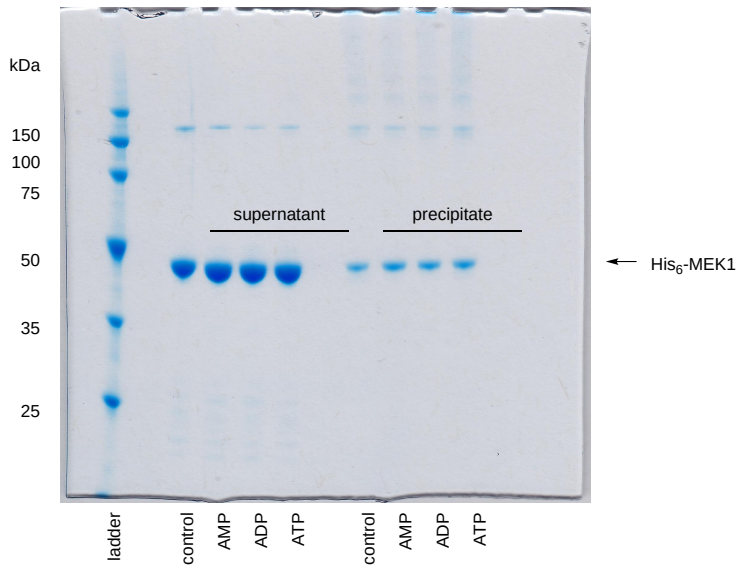
of supernatant liquid was removed from each reaction. Samples of the supernatant and pellet fractions were then analysed by SDS-PAGE.

with no protector (control):

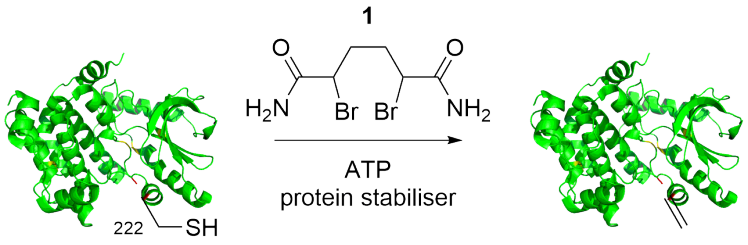


with AMP, ADP or ATP (representative MS):





General Conditions for Screening Protein Stabilising Additives during MEK1-Dha222 Formation



23 **24**

Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

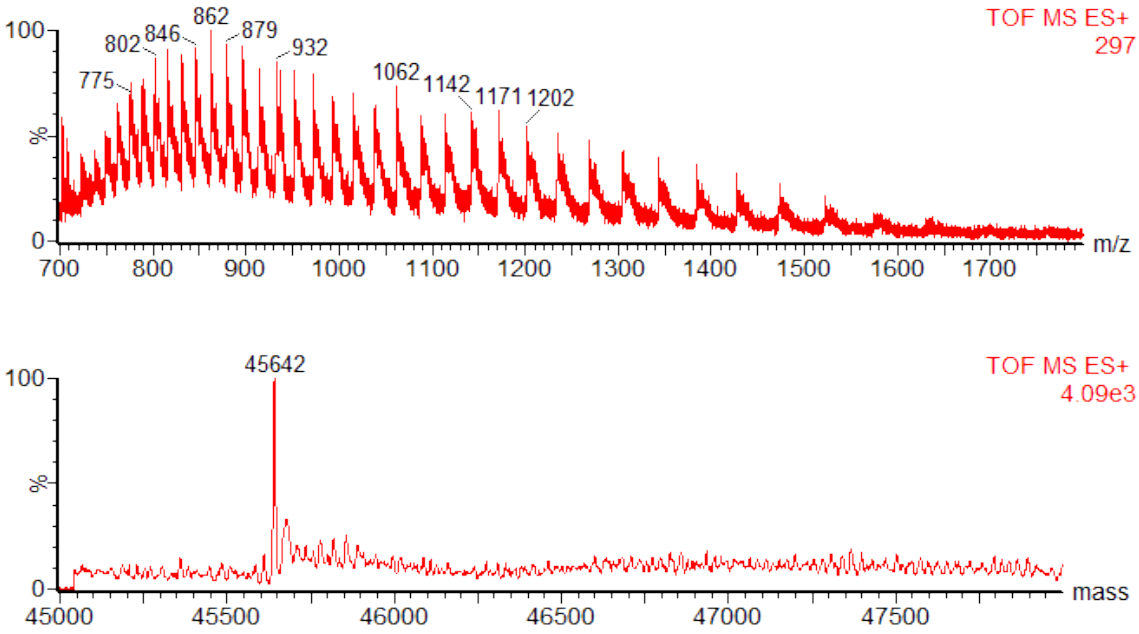
MEK1-Cys222 (**23**) in MEK1 storage buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (same as MEK1 storage buffer but with 50 mM HEPES pH 8.0 and no TCEP) using a G-25 SpinTrap™ desalting column (GE Healthcare). The concentration of the protein was adjusted with reaction buffer. To the diluted, desalted MEK1-Cys222 (50 µL of a 1.83 mg/mL solution, 2.0 nmol, 1.0 eq.) was added protein stabiliser (9.16 µL of additive stock, see table below) and ATP (9.16 µL of a 100 mM solution in water, pH 7.0) and the mixture incubated at room temperature for 30 min. Dibromide **1** (23.3 µL of a 3.9 mg/mL suspension in reaction buffer, 0.30 µmol, 150 eq.) was then added and the mixture incubated with shaking at 37 °C, 600 rpm for 6 h. LC-MS after this time showed >80% conversion to the Dha product in all cases. Formation of precipitate was also observed after this time. The precipitate was pelleted by centrifugation

(17 000 ×*g*, 15 min, 4 °C) and the supernatant carefully decanted by pipette, ensuring the same volume of supernatant liquid was removed from each reaction. Samples of the supernatant and pellet fractions were then analysed by SDS-PAGE: (calculated mass: 45 642, observed masses: 45 640–45 642).

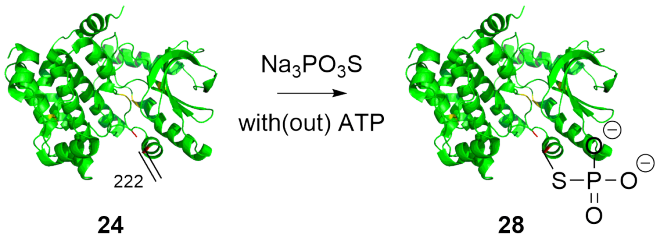
List of protein stabilising additives used:

Entry	Additive	Stock Concentration
1	–	–
2	Brij-35	0.3%
3	Triton X-100	0.2%
4	TWEEN-20	0.2%
5	PEG 8000	0.5%
6	Glycerol	60%
7	Arginine	1.0 M

Representative MS (extra glycerol as additive):

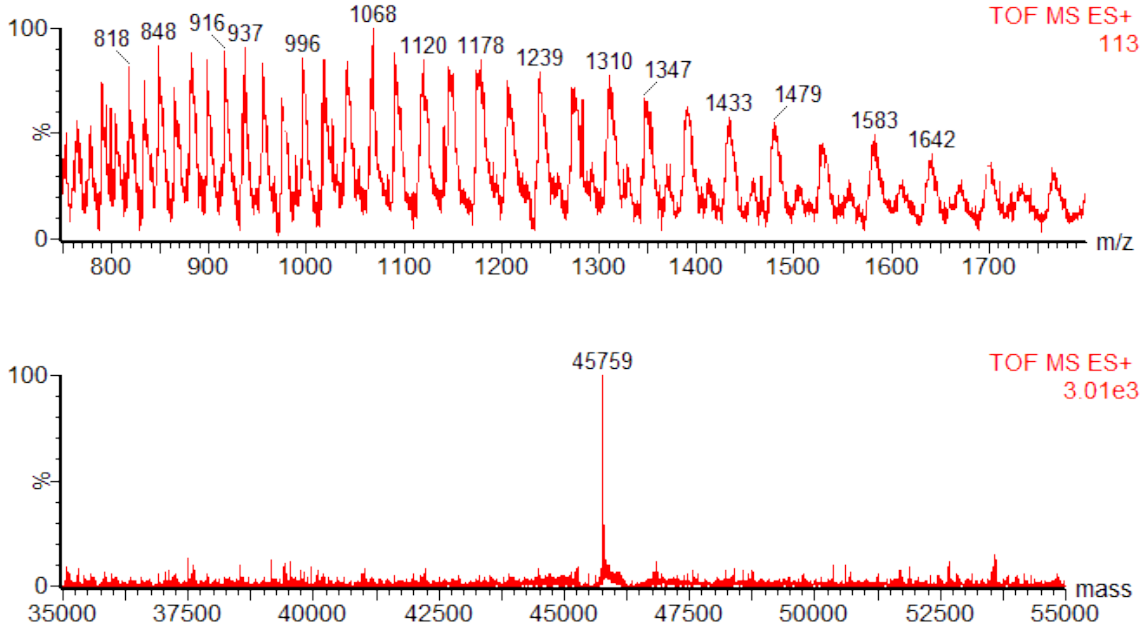


General Conditions for Screening Reaction Conditions for MEK1-pCys222 Formation



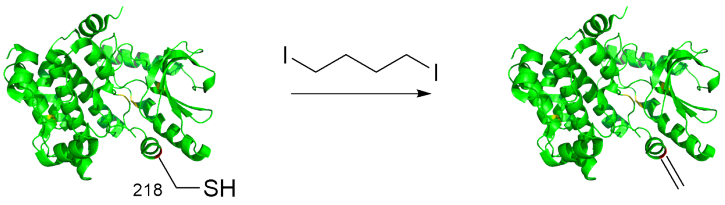
24 **28**

To MEK1-Dha222, **24** (35.0 µL of a 1.04 mg/mL solution, 0.80 nmol, 1.0 eq.) in MEK1 reaction buffer was added either ATP (1.0 µL of a 500 mM solution in water, pH 7.0) or water (same volume as ATP). Separately, sodium thiophosphate (15.67 mg) was dissolved in water (12.97 µL) and HCl (6.81 µL of a 5 M solution) added to adjust the pH to 8.0. The thiophosphate solution (4×1.35 µL of a 530 mg/mL solution, 4×4.0 µmol, 4×5000 eq.) was added batchwise (5 min intervals) to the protein and the resultant mixture incubated at either room temperature or 37 °C. Reaction progress was followed by LC-MS analysis. The optimal condition was found to be room temperature for 18 h. LC-MS at the endpoint under these conditions showed >85% conversion to the thiophosphate adduct. The reaction mixture was desalted by dialysis against MEK1 reaction buffer to give MEK1-pCys222, **28**: (calculated mass: 45 756, observed mass: 45 759).



4.11.7 Chemical Modification: Optimisation Experiments on MEK1-Cys218

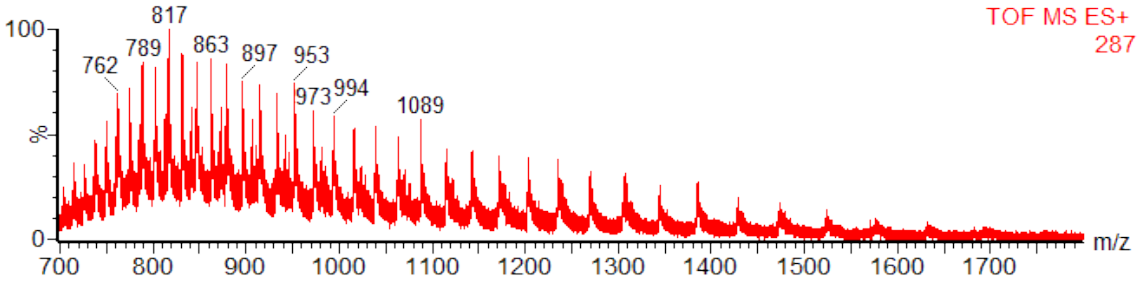
General Procedure for Treatment of MEK1-Cys218 (30) with 1,4-Diiodobutane

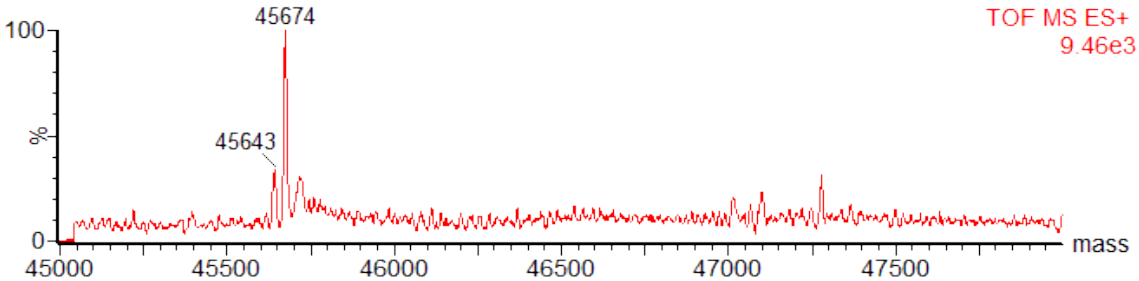


30 **31**

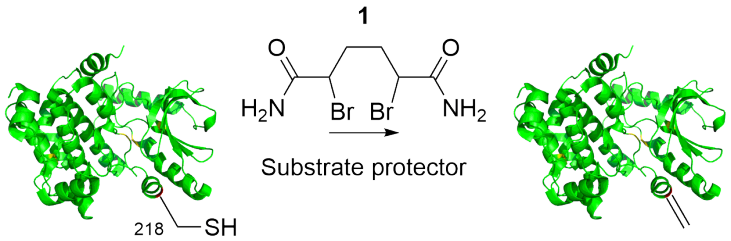
Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

An aliquot of MEK1-Cys218 (**30**) (60 µL of a 4.0 mg/mL solution, 5.3 µmol, 1.0 eq.) was thawed on ice and the buffer exchanged using a G-25 SpinTrap™ desalting column (GE Healthcare) to the desired buffer (either MEK1 reaction buffer or phosphate buffer). 1,4-Diiodobutane (1.04 µL, 7.9 µmol, 1500 eq.) was added and the mixture shaken at 37 °C, 550 rpm for 4 h, following the reaction progress by LC-MS. LC-MS after 6 h showed < 20% conversion to the single Dha product **31**. The mixture remained heterogeneous during the whole period of the reaction: (calculated mass: 45 642, observed mass: 45 643–45 646).





General Conditions for Screening Substrate Protecting Agents for MEK1-Dha218 Formation

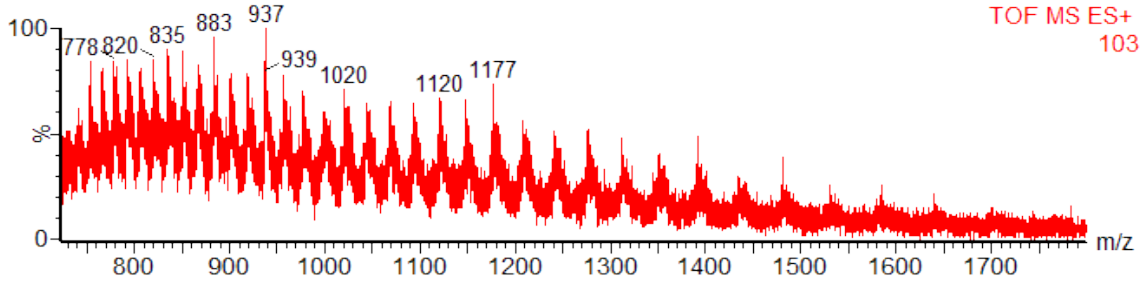


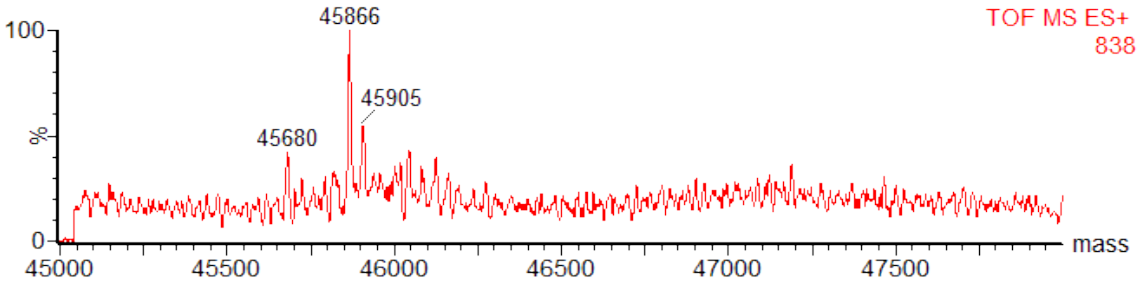
30 **31**

Containing:
1 surface Cys (shown)
2 hindered Cys
2 buried Cys

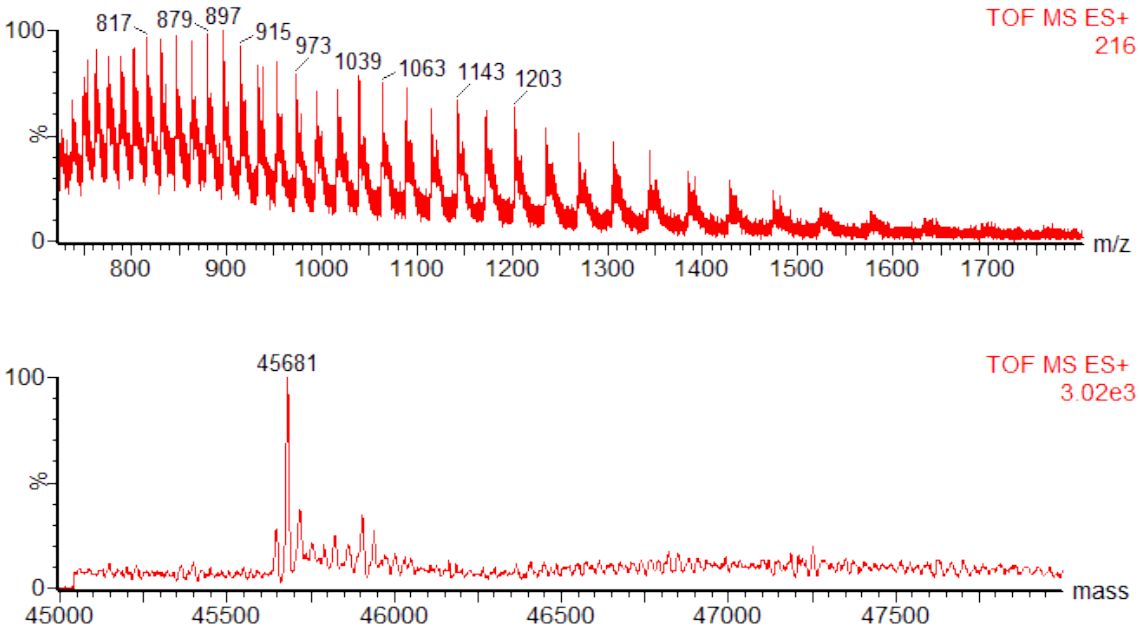
A similar procedure as the one in Section 4.11.6 was used here. MEK1-Cys218 in MEK1 storage buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP) was thawed on ice and the buffer exchanged to MEK1 reaction buffer (same as MEK1 storage buffer but with no TCEP) using a G-25 MiniTrap™ desalting column (GE Healthcare). The concentration of the protein was adjusted with reaction buffer. To the diluted, desalted MEK1-Cys222 (100 µL of a 1.67 mg/mL solution, 3.7 nmol, 1.0 eq.) was added substrate protector (one of water, AMP, ADP or ATP) (15.6 µL of 100 mM protector stock, pH 7.0) and the mixture incubated at room temperature for 30 min. Dibromide (40.7 µL of a 3.9 mg/mL suspension in reaction buffer, 0.53 µmol, 140 eq.) was then added and the mixture incubated with shaking at 37 °C, 600 rpm for 6 h. LC-MS after this time showed no conversion to the Dha product in all cases when substrate protector was used. In the case where no substrate protector was used, over-alkylation to the single Dha, single bromide adduct was observed (calculated mass: 45 863, observed mass: 45 866).

with no protector (control):





with AMP, ADP or ATP (representative MS):



4.11.8 Methods for Characterising (Modified) Proteins

4.11.9 MS/MS analysis and DSF

MS/MS and DSF procedures were similar to those previously described (Section 3.8.7 and 2.11.9 respectively).

General Procedure for Measurement of Circular Dichroism Spectra

A similar procedure to that described previously (Section 2.11.9) was used on MEK1. Prior to measuring the spectra, samples not already in MEK1 reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol) underwent buffer exchange using a combination of G-25 SpinTrap™ desalting column (GE Healthcare) and repeated concentration/dilution by Vivaspin (MWCO 10 000). MEK1-pCys222 which

would not tolerate concentration/dilution instead had its buffer exchanged by extensive dialysis (Slide-A-lyser® MWCO 10 000 dialysis cassette). Samples were then diluted (0.46–0.51 mg/mL) to an appropriate final volume (200–220 µL), loaded into a cuvette with thin pathlength (1.0 mm) before the spectra were collected using a Chirascan (Applied Photophysics). Collection parameters: Wavelength: 180–260 nm, Wavelength step size: 0.5 nm. Appropriate blank spectra were taken to account for background signal, one for the blank reaction buffer, and one for reaction buffer containing ATP (10 mM). The latter blank was required as MEK1-pCys222 could not be easily dissociated from ATP. ATP was found to have no influence on ellipticity. Data collected were exported as raw data and reprocessed in MATLAB to give the final plot (mean average of repeat scans taken, spectra smoothed using Savitzky-Golay filtering. Smoothing parameters: MATLAB function: `sgolayfilt` from Signal Processing toolbox, Polynomial order: 1, Window size: 5). Examples of the MATLAB scripts used for the data processing and plotting are given in Appendix C, where Appendix C.10 is most relevant.

Chapter 5

Functional Evaluation of MEK1 Kinase after Chemical Modification

Given the success of chemical modification in producing MEK1-Dha222 and MEK1-pCys222 from MEK1-Cys222, the activity of all these species had to be tested. A suitable assay therefore was to be developed, which in turn required a suitable substrate to be chosen.

5.1 Selection and Production of MEK1 Substrate

MEK1 only has two known substrates, ERK1 and ERK2.[135] In sourcing a suitable construct for expressing a substrate species, a brief survey of literature studies of MEK1 activity was undertaken. This survey revealed that the ERK1 or 2 used originated from a range of species including mouse[38,136] and rat.[137,138] However, given the pharmaceutical orientation of the project, it was decided that investigation on human proteins was an important aspect of the work. In searching specifically for a construct containing the full-length human ERK1 or 2, it was found that a construct for human ERK1 had previously been reported,[11] and was readily accessible. ERK1 was thus chosen for the MEK1 study at hand.

Like in the p38α study, the choice of a native protein substrate was considered important to investigating the biological recognition capabilities of the synthetically phosphorylated MEK1. The choice of full-length ERK1 would therefore sufficiently challenge the MEK1 chemical variants to determine if they had such biological recognition. However, whereas ATF2 as a substrate to p38α could be used directly in the assay, ERK1 could not. This is because ATF2 is a transcription factor and will therefore not interfere with

the assay conditions, while ERK1 is another protein kinase and therefore has activity when phosphorylated by MEK1. Therefore, for a MEK1 activity assay to be successful, MEK1 has to be the only active species present.

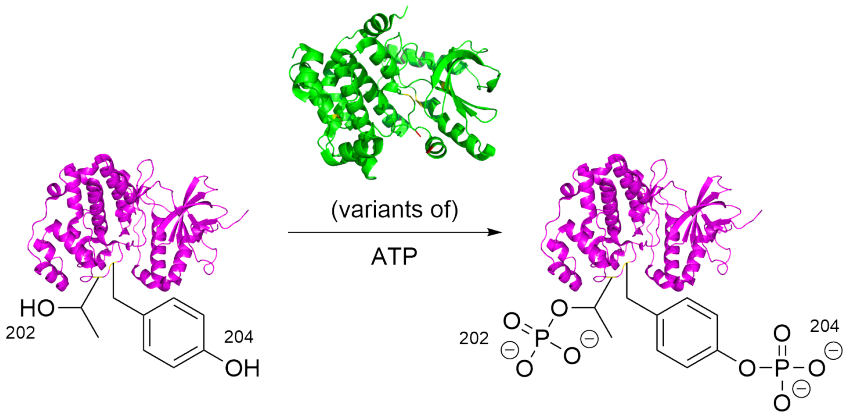
The potential for ERK1 to interfere in the activity assay lies in its ability to autophosphorylate,[11] and its potential to have ATPase activity. For the latter ability, ERK2, the closest homologue to ERK1, has been shown to have ATPase activity,[139] thus it is possible that ERK1 may have the same. ERK1 autophosphorylation would interfere with the assay since the MEK1 contribution to ERK1 phosphorylation would be difficult to distinguish from ERK1 autophosphorylation. Further, ATPase activity would interfere as ATP would be hydrolysed to ADP, thus causing the ATP concentration to become variable. In the case that the ATP concentration becomes low, the rate of phosphorylation observed would then be limited by the low ATP concentration and therefore would not be an accurate representation of MEK1 activity. To avoid such complications, the ERK1 had to be made enzymatically inactive (referred to as "kinase-dead" form), which could be achieved by K71R mutation.[11]

After ensuring that the K71R mutation on ERK1 had been successfully made, ERK1-K71R was expressed according to the reported procedure[11] and purified with some procedure modifications. The purity had been reported to be 60–80% and therefore unacceptability low, so the literature protocol was modified to include a further size exclusion purification step. This step also replaced the buffer exchange step, which had previously been performed by dialysis. The result was improved purity (>95%).

5.2 MEK1 Preliminary Activity Assay

With both the kinase and its substrate in-hand, the activity of all the chemical variants of MEK1-Cys222 could be tested. Many of the design aspects for the assay on MEK1-Cys222-derived chemical variants were based on the p38α preliminary study (Section 3.1). These included the use of MS as the detection method, and the use of near-stoichiometric quantities of MEK1 chemical variant (5 µM MEK1 enzyme, 10 µM ERK1 substrate). The former was important in providing a richer picture of MEK1 action, while the purpose of the latter was to increase the detection sensitivity of the assay and ensure the greatest chance of detecting enzymatic activity. Increasing detection sensitivity was also considered to be important since there was evidence suggesting that MEK1 may have required bis-phosphorylation for full activation.[22]

Given that the MEK1-Cys222-derived chemical variants were mono-phosphorylated at most, their activity could therefore be very low, thus requiring the assay to be as sensitive as possible.



Scheme 5.1 ERK1 is phosphorylated by MEK1 at Thr202 and Tyr204 using ATP as a co-factor.

As with the p38α activity assay, the format of the MEK1 assay involved the detection of the different phosphorylation states of the ERK1 substrate (referred to as ERK1, ERK1-P and ERK1-PP, for unphosphorylated, mono-phophosylated and bis-phosphorylated ERK1 respectively) by MS (Scheme 5.1). In performing the assay, phosphorylation reactions were run with the three chemical variants derived from MEK1-Cys222. Samples were taken at periodic timepoints from the assay reaction and quenched with guanidine solution, before being analysed by MS. It was noted upon closer inspection of the spectra that the use of guanidine over urea for the quenching agent had eliminated the problem of cyanate adduct formation previously observed in the p38α assay (Section 3.5.2).

Analysis of the data gratifying revealed that all 3 species were found to be active, such that MEK1-Dha222 and MEK1-pCys222 converted ERK1 completely to ERK1-PP after 1 h (Fig. 5.1b), while MEK1-Cys222 converted ERK1 completely to ERK1-PP after 4 h (Fig. 5.1b). Given the correlation between enzymatic activity and tertiary structure, and that all the species still had enzymatic activity, this demonstrated that the tertiary structure had remained intact over all the synthetic steps of chemical modification, despite previous concerns (Section 4.6). Overall, this result gave an indication that MEK1-pCys222, as well as MEK1-Dha222 were more active than MEK1-Cys222. The elevated activity of MEK1-Dha222 was particularly surprising, since the apparent activation of a kinase by the installation of Dha was unprecedented.[1,48] Further investigation was therefore required to discern the exact order of activity of the three species, to attempt to get a better insight into the unusual activity of MEK1-Dha222, and to obtain a better mechanistic viewpoint of ERK1 phosphorylation by the MEK1 chemical variants.

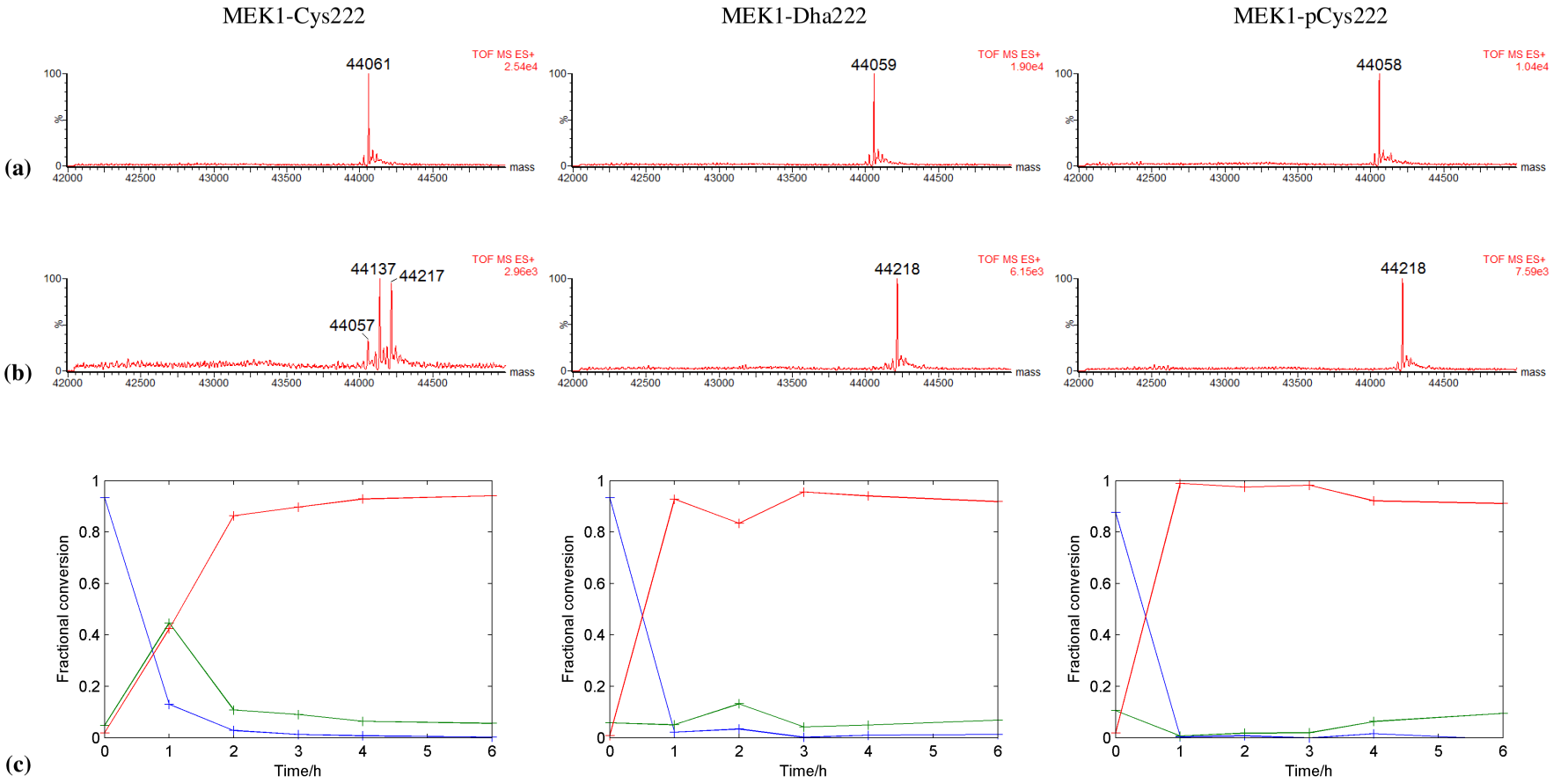


Fig. 5.1 In columns: The subfigures refer (from left to right) to MEK1-Cys222, MEK1-Dha222 and MEK1-pCys222 respectively. **(a)** MS taken at $t = 0\,\mathrm{h}$. ERK1: (calculated mass: 44 058, observed mass: 44 057–44 061). For comparison to **(b)** MS taken at $t = 1\,\mathrm{h}$. ERK1-P: (calculated mass: 44 138, observed mass: 44 137), ERK1-PP: (calculated mass: 44 218, observed mass: 44 217–44 218). The shift to the higher mass of the phosphorylated species can clearly be seen. **(c)** Timecourse plots, showing the conversion to species of the various phosphorylation states of ERK1. Key: ERK1 (blue), ERK1-P (green) and ERK1-PP (red).

5.3 MEK1 Activity Assay Optimisation

In order to distinguish the order of activity of the various MEK1-X222 chemical variants, the resolution of the data had to be improved. Improving resolution would require more points to be taken per unit time. Slowing down the enzymatic reaction would facilitate this process, where decreasing enzyme:substrate ratio is one method of doing this. Further, such a decrease in enzyme:substrate ratio would be required if the data were to be kinetically quantified by steady state kinetics, since as previously mentioned (Section 3.5), the requirement of steady state kinetics is that substrate concentration ([S], ERK1 in this case) is much greater than enzyme concentration ([E], MEK1-X222 chemical variant in this case).[105,108] Decrease of enzyme:substrate ratio is achieved by either lowering enzyme concentration, or increasing substrate concentration. Since ERK1 concentration had been chosen to make its detection by LC-MS more facile, ERK1 concentration was kept constant, and MEK1 concentration varied.

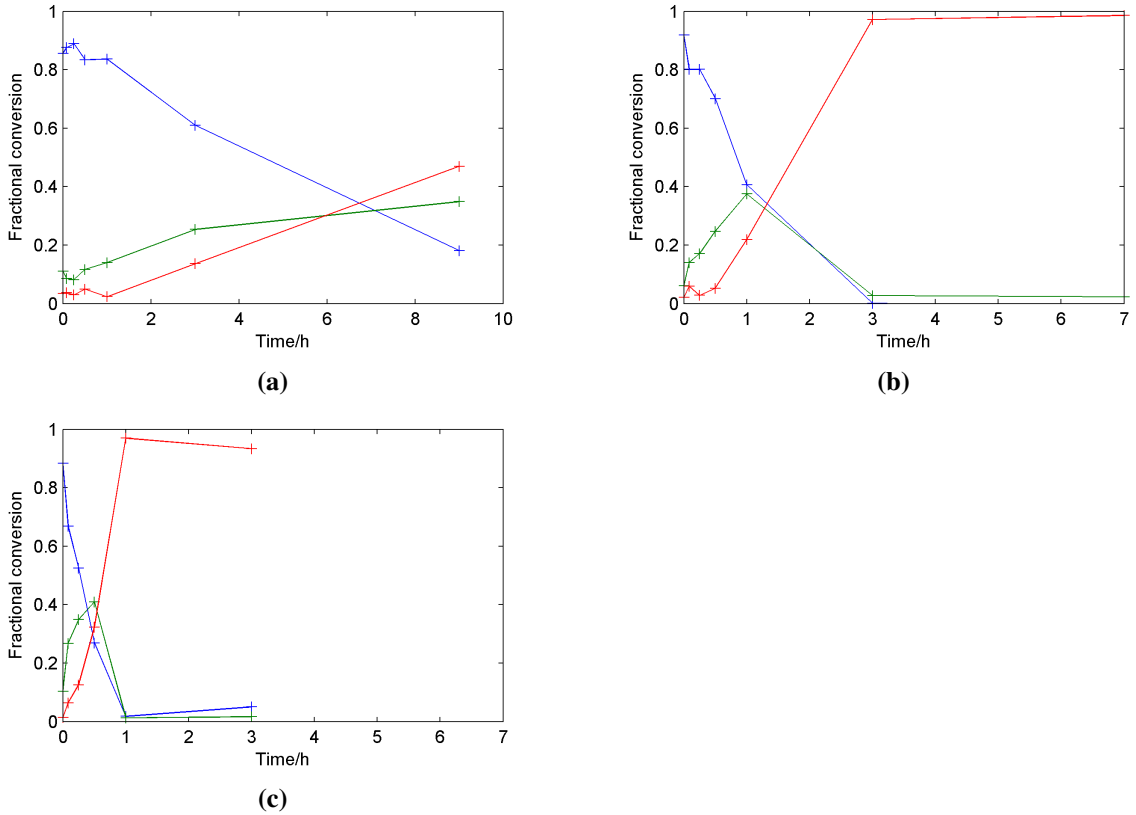


Fig. 5.2 Timecourses of ERK1 phosphorylation collected with varying MEK1-pCys222 concentration: **(a)** [MEK1-pCys222] = 0.1 µM, **(b)** [MEK1-pCys222] = 0.5 µM, **(c)** [MEK1-pCys222] = 1.0 µM. Key: ERK1 (blue), ERK1-P (green) and ERK1-PP (red).

In using MEK1-pCys222 to gauge a suitable enzyme concentration, a range of MEK1-pCys222 concentrations were tried, and the effects on ERK1 phosphorylation again monitored (Fig. 5.2). The timecourse

plots revealed that the lowest MEK1-pCys222 concentration tested (0.1 μM, Fig. 5.2a) was within the [S] ≫ [E] requirement of steady state kinetics, while still giving a discernible/distinguishable signal in a reasonable time. However, in order for the timecourse to also reveal mechanistic information on ERK1 phosphorylation by the MEK1 chemical variants, the timecourse had to be run to the exhaustive phosphorylation of ERK1. This criterion was not met with the aforementioned MEK1-pCys222 concentration, but it was by the medium concentration tested (0.5 μM, Fig. 5.2b). This higher concentration would also more practically allow the ERK1 conversion profiles of the other MEK1 chemical variants to be measured in full, since the other MEK1 chemical variants were likely to be slower than MEK1-pCys222. Further, although not as ideal as the low concentration, the intermediate MEK1 concentration had the potential to be used for saturation kinetic studies. Since the practical limit on [ERK1] was ∼50 μM, [MEK1 variant] = 0.5 μM would still allow the condition of [S] ≫ [E] to be satisfied. Finally, the highest MEK1-pCys222 concentration (Section 5.2c) would not meet the condition of [S] ≫ [E] and was hence ruled out, leaving [MEK1 variant] = 0.5 μM as the concentration of choice for the optimised assay.

5.4 Mechanistic Insights

Using the new assay conditions, more refined datasets for the MEK1-X222 chemical variants were collected (Table 5.1, Entries 1–3), and a number of positive controls were also added to the assay by adding two further MEK1 variants, MEK1-A19D and MEK1-S222E (Table 5.1, Entry 4 and 5). As MEK1-WT was unavailable, MEK1-A19D (prepared by Dr. Balakumar Vijayakrishnan) was used as a surrogate, since it was the readily available MEK1 variant with fewest mutations. This variant contained an *N*-terminal mutation, a mutation which is within the ERK binding domain.[140] The mutation could therefore alter the recruitment of the substrate, which would alter the overall kinetics of the variant, but was unlikely to affect the enzyme's catalytic efficiency. On the other hand, MEK1-S222E was used as a positive comparison for phosphate mimic MEK1-pCys222, since this variant has been shown to be constitutively active[10,12] and represented the most active phosphate mimic described in the literature. MEK1-Dha222 had no natural comparison and therefore did not have a positive control associated with it.

A number of negative control experiments were also added to the testing panel (Table 5.1, Entry 6 and 7). As ERK1 is precedented to autophosphorylate,[11] the autophosphorylation ability of ERK1-K71R was tested by treating it with ATP, but without any MEK1 variant being present. A test for the biological recognition of MEK1-pCys222 was also added, with p38α-WT being introduced as the substrate in place

Table 5.1 Summary of all the combinations of MEK1 variant and substrate tested.

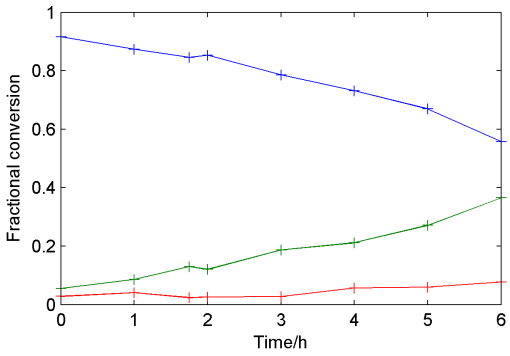
Entry	MEK1 (Chemical) Variant	Substrate	Remarks
1	MEK1-Cys222	ERK1-K71R	–
2	MEK1-Dha222	ERK1-K71R	–
3	MEK1-pCys222	ERK1-K71R	–
4	MEK1-S222E	ERK1-K71R	–
5	MEK1-A19D	ERK1-K71R	–
6	–	ERK1-K71R	Autophosphorylation test
7	MEK1-Cys222	p38α-WT	Biological recognition test

of ERK1-K71R. p38α and ERK1 both come from the same tier of the MAP kinase family and are therefore homologues.[2] The premise of this experiment was that if the MEK1 mimic were to incorrectly recognise a kinase, the most likely candidate for misrecognition would be a another kinase which is homologous to ERK1, such as p38α.

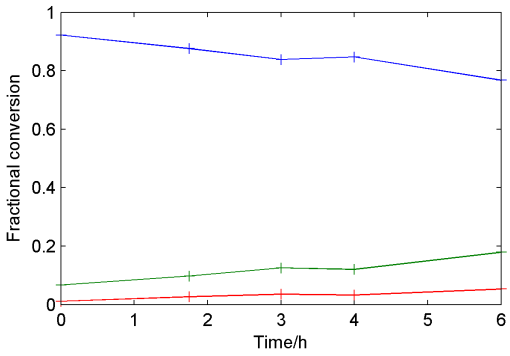
With the updated assay conditions allowing for higher resolution of enzymatic activity, and using the complete conversion of ERK1 to ERK1-PP as the reference endpoint, it could be confirmed that the order of activity of MEK1-X222 chemical variants was MEK1-Cys222 $<$ MEK1-Dha222 $<$ MEK1-pCys222 (Fig. 5.3a, 5.3c and 5.3d). This order of activity was expected, and demonstrated that while Dha installation could confer some activity to MEK1, it was not as efficient as phosphocysteine in doing so. On widening the comparison to the positive controls, it was found that the controls had similar activities to their intended counterparts: MEK1-A19D (Fig. 5.3b) had activity most similar to MEK1-Cys222, while MEK1-S222E (Fig. 5.3e) had activity similar to that of MEK1-pCys222.

The two negative controls were both negative. In the experiment where no MEK1 variant was present, no phosphorylation of ERK1 was observed either, thus showing that ERK1 was not able to autophosphorylate. Basal level autophosphorylation of ERK1-WT has been reported,[141] so the lack of autophosphorylation in the assay demonstrated that the K71R mutation was successful in preventing autophosphorylation from occurring. The result from this negative control could thus also be used as a baseline comparison for the basal activity displayed by MEK1-Cys222 and MEK1-A19D. The use of p38α-WT as the substrate also did not result in any substrate phosphorylation. This showed that despite being a semi-synthetic construct, MEK1-pCys222 was still specific to the phosphorylation of ERK1, and that the biological recognition of MEK1 had not been disrupted by the introduction of phosphocysteine.

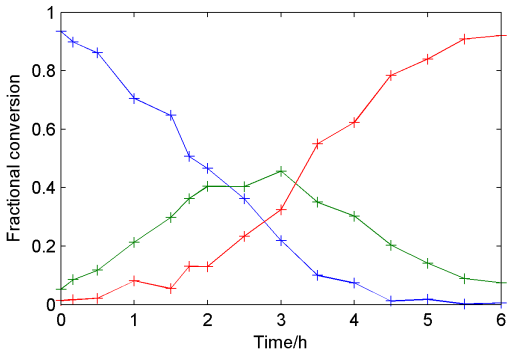
To allow for a more detailed ordering of activity for the 5 variants of MEK1, the initial rates of ERK1 consumption (ν_{0_ERK1}) were calculated, along with the initial rates for ERK1-P and ERK1-PP formation



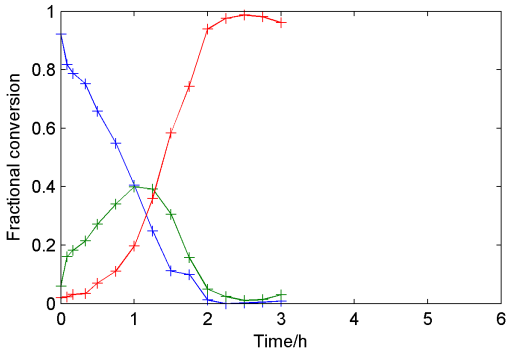
(a) with MEK1-Cys222



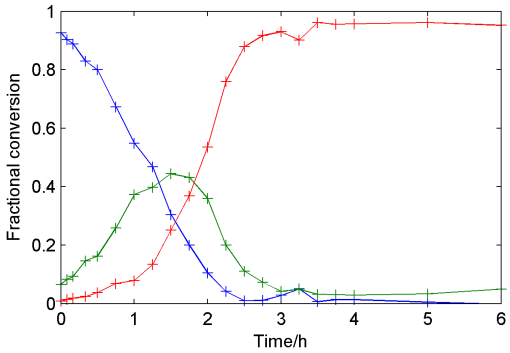
(b) with MEK1-A19D



(c) with MEK1-Dha222



(d) with MEK1-pCys222



(e) with MEK1-S222E

Fig. 5.3 Timecourses for ERK1 species evolution upon treatment with various MEK1 variants that displayed activity, either due to basal activity or due to activation. Key: ERK1 (blue), ERK1-P (green) and ERK1-PP (red).

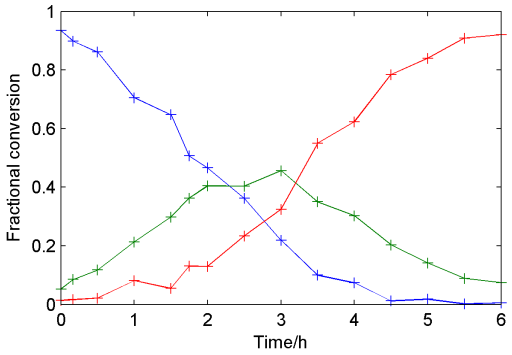
($\nu_{0\_\mathrm{ERK1\text{-}P}}$ and $\nu_{0\_\mathrm{ERK1\text{-}PP}}$ respectively) (Table 5.2). These values were also calculated for the ERK1 autophosphorylation negative control for a baseline comparison. A full kinetic analysis was not undertaken, mainly due to the constraints of time, since large amounts of MS machine time would be required. It was found that the trends between MEK1 variants for the consumption of ERK1 were also reflected in the initial rates of ERK1-P and ERK1-PP formation.

Table 5.2 Initial rates of the consumption or formation of the species corresponding to the various phosphoforms of ERK1.

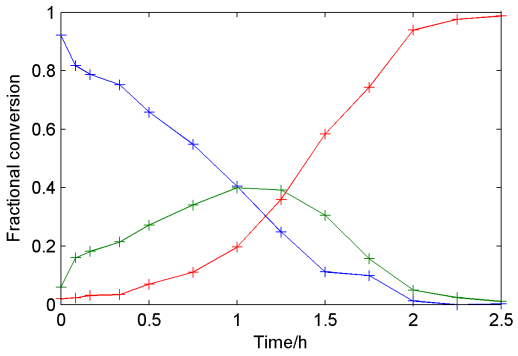
MEK1 Variant	ν_{0_ERK1}/μM h^{-1}	$\nu_{0\_\mathrm{ERK1\text{-}P}}$/μM h^{-1}	$\nu_{0\_\mathrm{ERK1\text{-}PP}}$/μM h^{-1}
MEK1-Cys222	0.58 ± 0.05	0.49 ± 0.04	0.08 ± 0.02
MEK1-Dha222	2.29 ± 0.21	1.73 ± 0.08	0.56 ± 0.16
MEK1-pCys222	4.52 ± 0.44	3.30 ± 0.47	1.22 ± 0.17
MEK1-A19D	0.24 ± 0.03	0.18 ± 0.02	0.07 ± 0.01
MEK1-S222E	3.71 ± 0.27	2.97 ± 0.25	0.74 ± 0.06
−ve control	0.06 ± 0.02	0.02 ± 0.01	0.04 ± 0.01

As ERK1 consumption is the marker for overall activity, the $\nu_{0-\mathrm{ERK1}}$ values were used to grade the activity of the MEK1 variants. It was thus found that MEK1-A19D was less active than MEK1-Cys222, while MEK1-S222E was closer in activity to MEK1-pCys222 than to MEK1-Dha222. Although MEK1-A19D was less active than MEK1-Cys222, comparison to the value of the negative control showed that it did still display activity. This demonstrated that basal activity was not just a trait of the variant MEK1-Cys222, where basal activity might have arisen from the various mutations made (C277S and C376S, as well as S222C). Instead, basal activity is a feature native to inactive MEK1, a notation which is also found in the literature.[12,13] As for MEK1-S222E, its placing demonstrated that the introduction of negative charge at position 222 using glutamate was less efficient at producing an active variant, compared to using phosphocysteine. A similar observation had previously been reported where the effects on kinase activation using phosphocysteine, glutamate and aspartate in place of phosphothreonine were investigated, showing that phosphocysteine caused more activation of the kinase than either of the two acidic residues.[48] It was concluded in the literature study that phosphocysteine was a more faithful replacement of phosphothreonine by virtue of enzymatic activity. By extension, given that phosphocysteine induces more enzymatic activity than glutamate, it is therefore likely that phosphocysteine is also a more faithful mimic of phosphoserine than the acidic residues glutamate and aspartate, although it cannot be stated conclusively since no comparison to phosphoserine has so far been made.

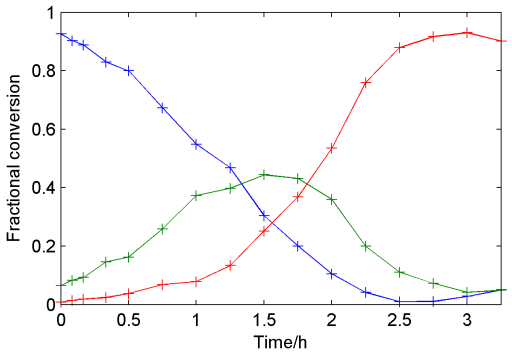
As mentioned previously (Section 3.3), the purpose of using MS as the assay detection method was to see what mechanistic insights could be gained, since it could distinguish between the different phosphorylation states of the substrate. For such further insight, closer attention was given to the profiles of the timecourses, particularly for those that had the ERK1 converted fully to ERK1-PP. Of the 5 MEK1 variants tested, the timecourses of MEK1-Dha222 (Fig. 5.3c), MEK1-pCys222 (Fig. 5.3d) and MEK1-S222E (Fig. 5.3e) fitted this criterion. An initial, visual inspection of the three timecourses suggested that for each timecourse, the relationship between the three curves of the three ERK1 phosphoforms were the same. This was characterised by the ERK1 curves crossing each other in corresponding positions on the three timecourses, a character which was clearer upon re-plotting of the graphs with different timescales (Fig. 5.4). The relationship between the timecourses was therefore a re-scaling in the time dimension.



(a) with MEK1-Dha222



(b) with MEK1-pCys222



(c) with MEK1-S222E

Fig. 5.4 Timecourses for ERK1 species evolution upon treatment with various MEK1 variants that displayed activity: **(a)** reproduced from Fig. 5.3c for ease of comparison, **(b)** re-plotted from Fig. 5.3d, **(c)** re-plotted from Fig. 5.3e. The re-plotting of **(b)** and **(c)** highlights the observation that the overall profiles of the timecourses are the same in all 3 cases. Key: ERK1 (blue), ERK1-P (green) and ERK1-PP (red).

The ratios between initial rates were calculated as a more rigorous approach to describe the profile relationship between the curves of a given timecourse. Taking ratios in this fashion normalises the relationships between the curves of a given timecourse, thus allowing the ratios to be compared with those

of other timecourses. Upon comparison of the ratios between timecourses, it was found that the ratio between initial rate of ERK1 consumption (ν_{0_ERK1}) and initial rate of ERK1-P production ($\nu_{0\_\mathrm{ERK1\text{-}P}}$) was the same for all 5 MEK1 variants (Table 5.3). As ERK1-P can only be produced from ERK1, the ratio between ERK1 consumption and ERK1-P production should always be the same, however, there is a practical limitation on the measurement of rate, since the measurement cannot be made instantaneously. Further, the conversion of ERK1 to ERK1-P is not the only reaction occurring: conversion of ERK1-P to ERK1-PP is also occurs at the same time. The second reaction does not affect the intrinsic reaction rate of the first, but it does affect the measurement of the initial rates, particularly for $\nu_{0\_\mathrm{ERK1\text{-}P}}$. This discrepancy results in ν_{0_ERK1} and $\nu_{0\_\mathrm{ERK1\text{-}P}}$ being unequal, and resulting in ratios that are not unity. Hence, the equality in ν_{0_ERK1}:$\nu_{0\_\mathrm{ERK1\text{-}P}}$ ratio is significant and it can be stated that the relationship between the ERK1 and ERK1-P curves between timecourses is also the same.

Table 5.3 Initial rates of the consumption of ERK1 and formation of ERK1-P. The ratio between these two rates has also been calculated.

MEK1 Variant	ν_{0_ERK1}/µM h^{-1}	$\nu_{0\_\mathrm{ERK1\text{-}P}}$/µM h^{-1}	ν_{0_ERK1}: $\nu_{0\_\mathrm{ERK1\text{-}P}}$ ratio
MEK1-Cys222	0.58 ± 0.05	0.49 ± 0.04	1.16 ± 0.16
MEK1-Dha222	2.29 ± 0.21	1.73 ± 0.08	1.33 ± 0.14
MEK1-pCys222	4.52 ± 0.44	3.30 ± 0.47	1.37 ± 0.24
MEK1-A19D	0.24 ± 0.03	0.18 ± 0.02	1.37 ± 0.28
MEK1-S222E	3.71 ± 0.27	2.97 ± 0.25	1.25 ± 0.16

Given that the ν_{0_ERK1}:$\nu_{0\_\mathrm{ERK1\text{-}P}}$ ratio between timecourses are similar, it would be expected that the ratios ν_{0_ERK1}:$\nu_{0\_\mathrm{ERK1\text{-}PP}}$ and $\nu_{0\_\mathrm{ERK1\text{-}P}}$:$\nu_{0\_\mathrm{ERK1\text{-}PP}}$ would also be similar between timecourses, since the values describe the products in a closed system. Calculation of these values gave similarities, with all values between 3.7–7.1 for ν_{0_ERK1}:$\nu_{0\_\mathrm{ERK1\text{-}PP}}$, and between 2.7–6.1 for $\nu_{0\_\mathrm{ERK1\text{-}P}}$:$\nu_{0\_\mathrm{ERK1\text{-}PP}}$, but the range was outside that required to demonstrate equality. It therefore cannot be concluded that the relationship to the ERK1-PP curve to the ERK1 and ERK1-P curves is always the same, in all timecourses. However, given the sigmoidal nature of the ERK1-PP curve, accurate determination of $\nu_{0\_\mathrm{ERK1\text{-}PP}}$ using linear regression was unreliable and was therefore likely to contribute to the range of ratio values observed. Still, the equality of the ν_{0_ERK1}:$\nu_{0\_\mathrm{ERK1\text{-}P}}$ ratio gave credence to the notation that the profiles of the timecourses were the same, even between different MEK1 variants.

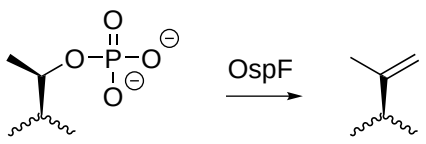
Despite not being able to conclusively show that the timecourse profiles are the same between MEK1 variants, the ERK1-PP curves of MEK1-Dha222, MEK1-pCys222 and MEK1-S222E were all similarly sigmoidal. Further to the similarity, a sigmoidal ERK1-PP curve has been attributed to the distributive mech-

anism exhibited by MEK1, where MEK1 dissociates from ERK1 after the first phosphorylation event.[137] The sigmoidal curve's presence in the timecourse plots suggests that MEK1-Dha222, MEK1-pCys222 and MEK1-S222E also follow this mechanism.

With such similarity in the profiles, it would be reasonable to assume that MEK1-Dha222, MEK1-pCys222 and MEK1-S222E owe their activity to the same molecular mechanism of activation. This however is not the case, with MEK1-pCys222 and MEK1-S222E being activated due to the addition of negative charge, while the likely cause of MEK1-Dha222 activation is due to sterics. The activity of MEK1-Dha222 implies that the change in geometry from $sp^3 \rightarrow sp^2$ on conversion of Cys→Dha causes a significant enough change in the conformation of MEK1, such that it renders the kinase partially active. Along with the similarity in timecourse profiles, it is therefore possible that activation in MEK1-Dha222, MEK1-pCys222 and MEK1-S222E all occur due to the same overall conformation change in MEK1, only that the conformational change is to different extents in the different variants, resulting in the different levels of activity observed. It is thus proposed that MEK1-Dha222 occupies an intermediate conformation between the inactive MEK1-Cys222 and the activated MEK1-pCys222.

5.5 Implications on Kinase Activation by Installation of Dehydroalanine

The increased level of activity of MEK1-Dha222 over MEK1-Cys222 raises the possibility that the activity of a kinase can also be altered *in vivo* by biochemical transformations other than phosphorylation/ dephosphorylation. One such biochemical transformation is performed by OspF, a phosphothreonine lyase from the various pathogenic species of *Shigella*, where on top of dephosphorylation, the enzyme also dehydrates the resulting threonine (Scheme 5.2).[142] This enzyme acts specific against members of the MAPK family of kinases, including ERK1 and p38α. However, it is the dephosphorylation that is likely to cause the dominate effect on the target kinase when causing it to be inactive, with the subsequent dehydration only serving to make the dephosphorylation irreversible. There are however no known reports of a modification, other than phosphorylation which activates a kinase.



Scheme 5.2 The biochemical reaction catalysed by OspF produces

It is likely that it is due to the dominant nature of phosphorylation on the activity of kinases that phosphorylation is the most widely studied transformation in kinases. Other transformations which may occur are likely to have much smaller effects, which would be difficult to detect in an ensemble measurement made in the background of a phosphorylated kinase's activity. This is an area of research that chemical modification could contribute readily to, since it may be easier to prepare some of these entities of interest to homogeneity in a semi-synthetic fashion, rather than through isolation from the native system.

Finally, the discovery of any such alternative modes of kinase activity manipulation by biochemical transformation would add further subtly to the activity of an individual kinase. Subtly in activity has already been observed in kinases where multiple phosphorylation sites are known. Taking MEK1 as an example, further to phosphorylation at Ser218 and Ser222 causing the kinase's activation, phosphorylation at Ser212 causes inactivation of the kinase.[13] Phosphorylation at a combination of phosphorylation sites is thus likely make a kinase's activity more dynamically varied, allowing it more flexibility than the otherwise digital view of active/inactive states. In turn, either the conformation or the intermediate levels of activity of different phosphoforms are likely to contribute to the regulation of biomolecular interactions between a given kinase and its multiple interaction partners, with different phosphorylation states interacting more readily with some partners over others. In this context, alternative modes of kinase activity manipulation, if present, would add a further level of complexity, but may help to explain and better describe the network nature of cellular signalling.

5.6 Summary

Having established a synthetic method to obtain MEK1-pCys222 from MEK1-Cys222, the enzymatic function of the resulting MEK1-X222 species demanded testing. A suitable substrate was found in ERK1-K71R for the assay. Preliminary studies showed that MEK1-Cys222 demonstrated basal activity, while the introduction of negative charge in MEK1-pCys222 had made it more active. These results confirmed literature reports of MEK1 basal activity[12,13] and the activation of MEK1 by introducing negative charge at the native phosphorylation site.[10,12,38] Further, these observations were an extension from the studies on p38α in Chapters 2 and 3, which along with further literature precedent[48] had shown that introduction of negative charge in the form of pCys on kinases had the potential for kinase activation. What was a surprise however is that the synthetic intermediate, MEK1-Dha222 had also been activated upon chemical modification, which was unprecedented.

A more detailed analysis revealed that while MEK1-Dha222 was of intermediate activity between MEK1-Cys222 and MEK1-pCys222, it wasn't as active as MEK1-S222E. MEK1-S222E is a variant arising from the common design strategy of replacing the negative charge of the phosphate with an acidic residue, which is subsequently incorporated through genetic means, a strategy which made MEK1-S222E have activity closer to MEK1-pCys222 than to MEK1-Dha222. Closer observation of the timecourse profiles of the three activated species (MEK1-Dha222, MEK1-pCys222 and MEK1-S222E) showed profile similarities among the three timecourses. This lead to a suspicion that the activation of these three species arises from the same overall conformational change in MEK1, despite the difference in activation strategy at the atomic level. Thus, the activation of MEK1-Dha222 solely through what seems to be the change in bond geometry from $sp^3 \rightarrow sp^2$ warrants further investigation. The activity of MEK1-Dha222 raises the largely unstudied possibility that the activity of kinases could also be controlled by other mechanisms *in vivo*, other than phosphorylation/dephosphorylation.

As mentioned in the previous Chapter, this work on MEK1 was a work in progress at the time of writing. Although it can be concluded that the introduction of negative charge through phosphocysteine activated MEK1 more than using glutamate for the same purpose, the question of whether phosphocysteine is a sufficient mimic of phosphoserine cannot be conclusively answered at this stage. This a particularly true as direct no comparison to phosphoserine has been made. However, in the specific case of MEK1-pCys222, the corresponding MEK1-pSer222 variant would be very difficult, if not impossible to produce in a homogeneous form. It should be noted that MEK1-pCys222 still has the other native phosphorylation site, Ser218, intact as serine. In producing MEK1-pSer222 for the required comparison, an enzymatic or biological solution would be needed. However, given that the upstream kinase to MEK1 has the propensity to phosphorylate both Ser218 and Ser222 nearly simultaneously,[12] the requirement of regioselective phosphorylation at Ser222 is highly unlikely to be met. The strategy f using of alanine mutants to block unwanted phosphorylation at position 218 has been used before,[22] however, this mutation also eliminates any potential of cooperativity[143] or interaction that the hydrogen bonding of the unphosphorylated serine, Ser218 in this case, might have. Evidence of the importance of such an interaction is in the form of discrepancies in biological conclusions between studies where such alanine mutants were used to draw conclusions,[22] compared to corresponding studies where they were not used.[12] In these two cases, the former study which made use of alanine mutants concluded that phosphorylation at both Ser218 and Ser222 was required for MEK1 activation, while the latter study concluded that phosphorylation at either Ser was

sufficient to give full enzymatic activity. The MEK1 study described in this Chapter certainly supports the latter literature MEK1 study. It also demonstrates the importance of retaining the wild-type sequence in the kinase's activation loop, aside for the mutations that are required for direct intervention at the phosphorylation site being investigated, i.e. investigation into phosphorylation at position 222 only warrants mutation at position 222 and nowhere else. Thus, the production of MEK1-pCys222 and the study of its activity demonstrates a key advantage that the method of chemical modification has over more conventional methods of kinase activation: chemical modification is able to access phosphoforms of kinase variants that may otherwise be impossible to obtain homogenously.

Given the status of a work in progress, a number of further studies would be required to conclude whether or not pCys is a sufficient mimic of phosphoserine, even though the evidence here strongly suggests that this is the case. These proposals are discussed in the next Chapter.

5.7 Experimental

5.7.1 General Measures

General measures regarding molecular biology were as previously described (Section 2.11.1), except that BL21(DE3) chemically competent cells were purchased from New England Biolabs (NEB) and handled according to the manufacturer's instructions. General measures regarding gene sequence analysis, MS data analysis and MS data quantification were as previously described (Section 2.11.1). General measures regarding enzymatic reactions and activity assay were as previously described (Section 3.8.1).

5.7.2 Protein Expression and Purification of ERK1

Extraction of NpT7-5-ERK1-WT plasmid

Bacterial Culturing: *E.coli* transformed with NpT7-5-ERK1-WT plasmid was received from Addgene as a bacterial stab. The cells were streaked onto solid LB agar medium containing ampicillin (100 µg/mL) using liquid medium as a carrier (20 µL). The plates were incubated at 37 °C for 18 h. 3 colonies were selected and cultured separately in liquid LB medium (10 mL) containing ampicillin (50 µg/mL) in a shaking incubator at 37 °C, 250 rpm for 18 h.

DNA Purification and Storage: After glycerol stocks were made, the remaining cells were pelleted by centrifugation (3000 rpm, 10 min, 4 °C). The resulting cell pellet was resuspended in P1 buffer and the plasmid DNA purified out using the QIAprep® Miniprep kit. The plasmid DNA was eluted in EB buffer and stored at −20 °C. DNA sequencing confirmed that the sequence of the gene was intact. Sequencing was done with the T7F primer and a custom primer, 5′−TTCCTCTACCAGATCCTGCG−3′ (ERK1-460-479) for sequencing in the forward direction: plasmid concentration 56 – 120 ng/µL

Sequence of ERK1-WT:

```
   1 ATG GCA CAT CAC CAT CAC CAT CAT ATG GCG GCG GCG GCG GCT CAG
  51 GGG GGC GGG GGC GGG GAG CCC CGT AGA ACC GAG GGG GTC GGC CCG
 101 GGG GTC CCG GGG GAG GTG GAG ATG GTG AAG GGG CAG CCG TTC GAC
 151 GTG GGC CCG CGC TAC ACG CAG TTG CAG TAC ATC GGC GAG GGC GCG
 201 TAC GGC ATG GTC AGC TCG GCC TAT GAC CAC GTG CGC AAG ACT CGC
 251 GTG GCC ATC AAG AAG ATC AGC CCC TTC GAA CAT CAG ACC TAC TGC
 301 CAG CGC ACG CTC CGG GAG ATC CAG ATC CTG CTG CGC TTC CGC CAT
 351 GAG AAT GTC ATC GGC ATC CGA GAC ATT CTG CGG GCG TCC ACC CTG
 401 GAA GCC ATG AGA GAT GTC TAC ATT GTG CAG GAC CTG ATG GAG ACT
 451 GAC CTG TAC AAG TTG CTG AAA AGC CAG CAG CTG AGC AAT GAC CAT
 501 ATC TGC TAC TTC CTC TAC CAG ATC CTG CGG GGC CTC AAG TAC ATC
 551 CAC TCC GCC AAC GTG CTC CAC CGA GAT CTA AAG CCC TCC AAC CTG
 601 CTC ATC AAC ACC ACC TGC GAC CTT AAG ATT TGT GAT TTC GGC CTG
 651 GCC CGG ATT GCC GAT CCT GAG CAT GAC CAC ACC GGC TTC CTG ACG
 701 GAG TAT GTG GCT ACG CGC TGG TAC CGG GCC CCA GAG ATC ATG CTG
 751 AAC TCC AAG GGC TAT ACC AAG TCC ATC GAC ATC TGG TCT GTG GGC
 801 TGC ATT CTG GCT GAG ATG CTC TCT AAC CGG CCC ATC TTC CCT GGC
 851 AAG CAC TAC CTG GAT CAG CTC AAC CAC ATT CTG GGC ATC CTG GGC
 901 TCC CCA TCC CAG GAG GAC CTG AAT TGT ATC ATC AAC ATG AAG GCC
 951 CGA AAC TAC CTA CAG TCT CTG CCC TCC AAG ACC AAG GTG GCT TGG
1001 GCC AAG CTT TTC CCC AAG TCA GAC TCC AAA GCC CTT GAC CTG CTG
1051 GAC CGG ATG TTA ACC TTT AAC CCC AAT AAA CGG ATC ACA GTG GAG
1101 GAA GCG CTG GCT CAC CCC TAC CTG GAG CAG TAC TAT GAC CCG ACG
1151 GAT GAG CCA GTG GCC GAG GAG CCC TTC ACC TTC GCC ATG GAG CTG
1201 GAT GAC CTA CCT AAG GAG CGG CTG AAG GAG CTC ATC TTC CAG GAG
1251 ACA GCA CGC TTC CAG CCC GGA GTG CTG GAG GCC CCC
```

Mutagenesis of NpT7-5-ERK1 Plasmid

Mutagenesis Primers: The following primers were used to make the K71R mutant of ERK1 from ERK1-WT:

- K71R forward: 5′−CAAGACTCGCGTGGCCATCCGTAAGATCAGCCCCTTCGAAC−3′
- K71R reverse: 5′−GTTCTGAGCGCACCGGTAGGCATTCTAGTCGGGGAAGCTTG−3′

Although mutagenesis with the ERK1 primers was successful, It was noted that the K71R reverse was incorrectly designed and should have been the reverse complement, not the complement.

A similar procedure to that previously described for MEK1 (Section 4.11.3) was used to perform the rest of the mutagenesis, using XL10-Gold ultracompetent cells for the plasmid cloning in *E. coli* and ampicillin (100 mg/mL final concentration) instead of kanamycin for bacterial selection.

Sequence of ERK1-K71R:

```
   1 ATG GCA CAT CAC CAT CAC CAT CAT ATG GCG GCG GCG GCG GCT CAG
  46 GGG GGC GGG GGC GGG GAG CCC CGT AGA ACC GAG GGG GTC GGC CCG
  91 GGG GTC CCG GGG GAG GTG GAG ATG GTG AAG GGG CAG CCG TTC GAC
 136 GTG GGC CCG CGC TAC ACG CAG TTG CAG TAC ATC GGC GAG GGC GCG
 181 TAC GGC ATG GTC AGC TCG GCC TAT GAC CAC GTG CGC AAG ACT CGC
 226 GTG GCC ATC CGT AAG ATC AGC CCC TTC GAA CAT CAG ACC TAC TGC
 271 CAG CGC ACG CTC CGG GAG ATC CAG ATC CTG CTG CGC TTC CGC CAT
 316 GAG AAT GTC ATC GGC ATC CGA GAC ATT CTG CGG GCG TCC ACC CTG
 361 GAA GCC ATG AGA GAT GTC TAC ATT GTG CAG GAC CTG ATG GAG ACT
 406 GAC CTG TAC AAG TTG CTG AAA AGC CAG CAG CTG AGC AAT GAC CAT
 451 ATC TGC TAC TTC CTC TAC CAG ATC CTG CGG GGC CTC AAG TAC ATC
 496 CAC TCC GCC AAC GTG CTC CAC CGA GAT CTA AAG CCC TCC AAC CTG
 541 CTC ATC AAC ACC ACC TGC GAC CTT AAG ATT TGT GAT TTC GGC CTG
 586 GCC CGG ATT GCC GAT CCT GAG CAT GAC CAC ACC GGC TTC CTG ACG
 631 GAG TAT GTG GCT ACG CGC TGG TAC CGG GCC CCA GAG ATC ATG CTG
 676 AAC TCC AAG GGC TAT ACC AAG TCC ATC GAC ATC TGG TCT GTG GGC
 721 TGC ATT CTG GCT GAG ATG CTC TCT AAC CGG CCC ATC TTC CCT GGC
 766 AAG CAC TAC CTG GAT CAG CTC AAC CAC ATT CTG GGC ATC CTG GGC
 811 TCC CCA TCC CAG GAG GAC CTG AAT TGT ATC ATC AAC ATG AAG GCC
 856 CGA AAC TAC CTA CAG TCT CTG CCC TCC AAG ACC AAG GTG GCT TGG
 901 GCC AAG CTT TTC CCC AAG TCA GAC TCC AAA GCC CTT GAC CTG CTG
 946 GAC CGG ATG TTA ACC TTT AAC CCC AAT AAA CGG ATC ACA GTG GAG
 991 GAA GCG CTG GCT CAC CCC TAC CTG GAG CAG TAC TAT GAC CCG ACG
1036 GAT GAG CCA GTG GCC GAG GAG CCC TTC ACC TTC GCC ATG GAG CTG
1081 GAT GAC CTA CCT AAG GAG CGG CTG AAG GAG CTC ATC TTC CAG GAG
1126 ACA GCA CGC TTC CAG CCC GGA GTG CTG GAG GCC CCC
```

Protein Expression of ERK1-K71R

ERK1-K71R was expressed as previously reported.[11]

Transformation: Plasmid encoding the mutant ERK1-K71R was transformed into BL21(DE3) *E. coli* chemically competent cells by heat shock. BL21(DE3) *E. coli* competent cells (20 µL) were thawed on ice and plasmid DNA (1 µL) added. The mixture was incubated on ice for 5 min before heat shock was performed at 42 °C for 30 s, immediately returning the cells onto ice after. After a further incubation

period of 5 min, the cells were fed with SOC medium (80 µL) before incubation with shaking at 37 °C, 250 rpm for 1 h.

Bacterial Culturing: The freshly transformed cells (5 µL with 20 µL SOC medium as carrier) were spread onto an LB/agar plate supplemented with ampicillin (100 µg/mL final concentration). The plate was incubated at 37 °C for 12 h. From this plate, 3 colonies were selected and cultured further in liquid LB medium (3×10 mL), again supplemented with ampicillin to give starter cultures.

Protein Expression: After suitable glycerol stocks had been made of the starter cultures, one of the cultures was cultured further. LB medium (6×800 mL), again supplemented with ampicillin was inoculated with the starter culture (6×1 mL). These new cultures were grown to $OD_{600} \sim 0.85$ in an incubator with shaking at 30 °C, 180 rpm. Protein expression was then induced by the addition of IPTG (0.4 mM final concentration) and incubation continued at the same conditions for 4 h. The cells were finally harvested by centrifugation (8000 rpm, JA10 rotor, 4×10 min, 4 °C), the pellets placed into bags, flash frozen in liquid nitrogen and stored at −80 °C.

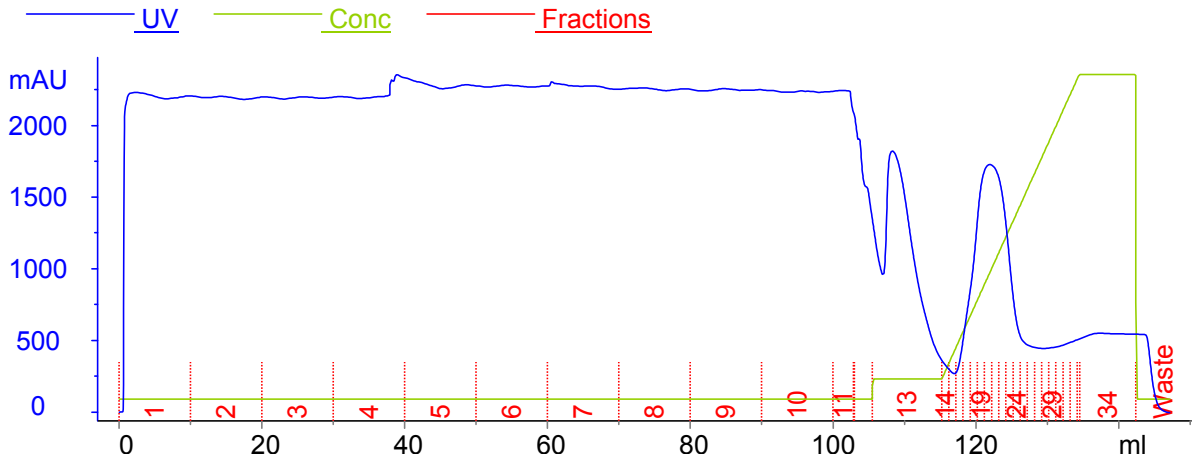
Protein Purification of ERK1-K71R

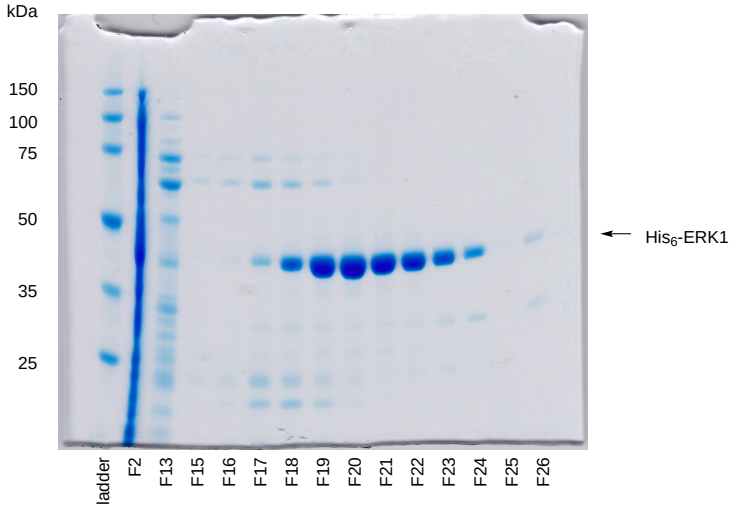
ERK1-K71R was purified using a protocol modified from that previously reported.[11]

Cell Lysis: The cell pellets (26 g) were combined, thawed on ice and re-suspended (100 mL) in ice cold ERK1 Ni^{2+} affinity lysis buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 20 mM imidazole, 5% glycerol, 0.5 mM TCEP) with stirring at 4 °C. Lysozyme (110 mg) was added and the mixture stirred for a further 2 h before sonication using a sonicator equipped with a microtip (25×2 s blasts, 60% amplitude, 2144 J total energy). DNase (1 mg) was added and the mixture stirred for a further 2 h before the lysate was clarified by centrifugation (20 000 rpm, JA25.50 rotor, 45 min, 4 °C). The lysate was further clarified by sequential filtration through 0.8 µm and 0.2 µm syringe filters.

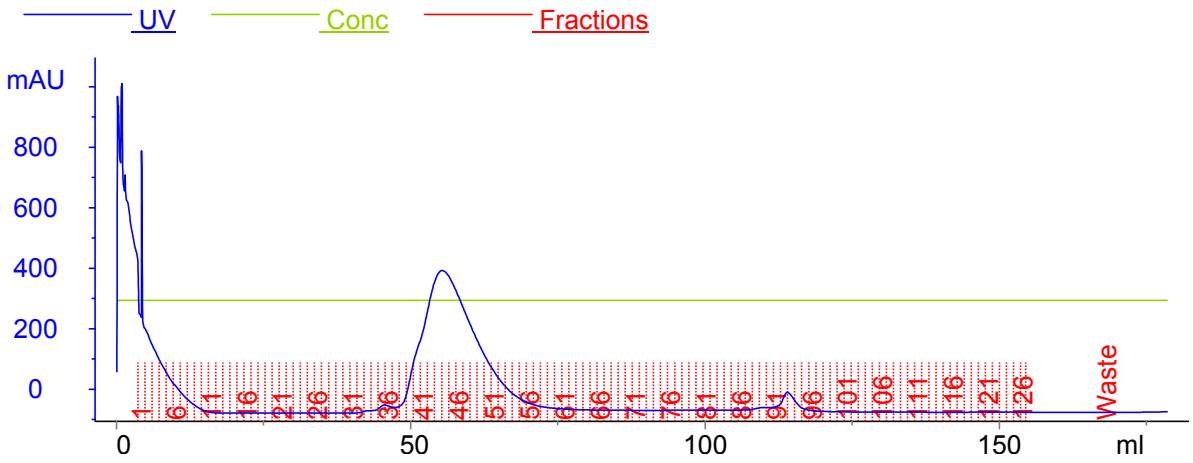
Ni^{2+} Affinity Chromatography: Using an ÄKTA Purifier™ FPLC system, the clarified lysate was loaded (105.5 mL) onto a pre-packed 1 mL HisTrap™ HP Ni^{2+} affinity column, chasing through (3.5 mL, 3.5 CV) with more lysis buffer. The column was washed (10 mL, 10 CV) with ERK1 Ni^{2+} affinity wash buffer (same as lysis buffer but with 50 mM imidazole) and the bound protein eluted using a linear gradient (20 mL, 20 CV, 1 mL fractions) from wash buffer to ERK1 Ni^{2+} affinity elution buffer (same as lysis buffer but with 500 mM imidazole). The fractions containing the protein of interest were eluted into

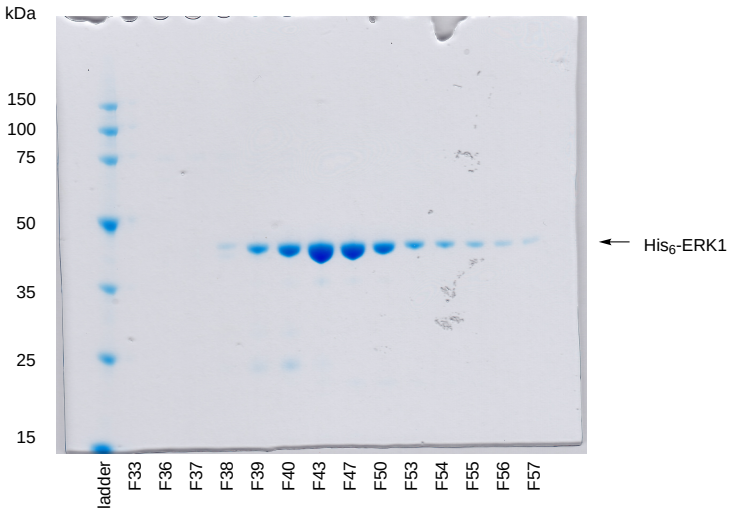
collection tubes already containing (2 mL each) lysis buffer. The fractions containing the desired protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and combined.

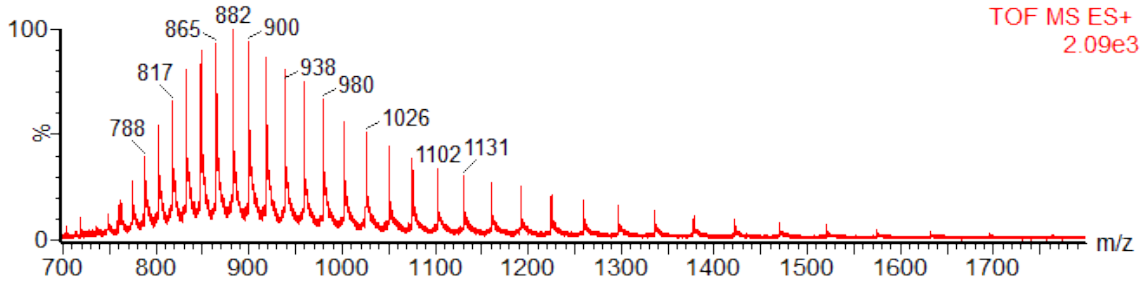


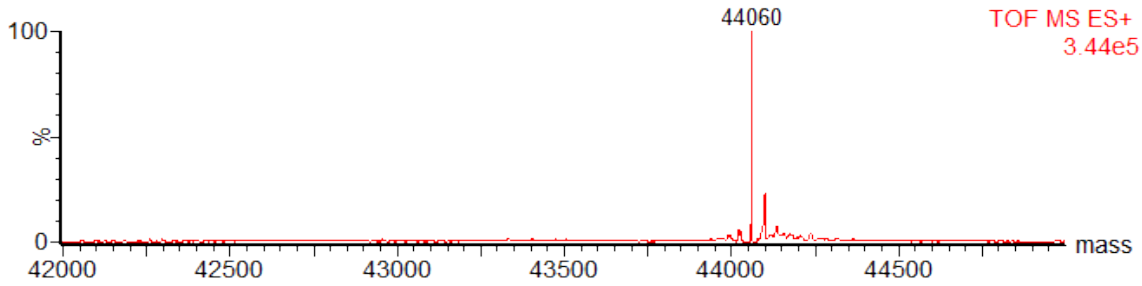


Size Exclusion Chromatography and Storage: The collected fractions were concentrated by Vivaspin (MWCO 10 000) to a final volume ∼1.5 mL before filtering through a syringe filter. Chasing through with ERK1 storage buffer (20 mM Tris pH 7.5, 10% glycerol, 1 mM EGTA, 0.5 mM TCEP), the concentrated proteins were loaded onto a HiLoad™ 16/60 Superdex 75 gel filtration column and filtered (180 mL, 1.5 CV, 1.2 mL fractions) into ERK1 storage buffer. The fractions containing the desired protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis, and collected. The collected fractions were re-concentrated by Vivaspin (MWCO 10 000, 2.77 mg/mL final concentration), divided into aliquots (90×40 µL), flash frozen in liquid nitrogen and stored at −80 °C: Yield: 10.4 mg in ERK1 storage buffer, (calculated mass: 44 058, observed mass: 44 060).


UV
Conc
Fractions
mAU
800
600
400
200
0
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
106
111
116
121
126
Waste
0
50
100
150
ml


kDa
150
100
75
50
35
25
15
ladder
F33
F36
F37
F38
F39
F40
F43
F47
F50
F53
F54
F55
F56
F57
His6-ERK1


TOF MS ES+
2.09e3
100
%
0
788
817
865
882
900
938
980
1026
1102
1131
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
m/z


TOF MS ES+
3.44e5
44060
100
%
0
42000
42500
43000
43500
44000
44500
mass

Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                                    MAHHHHHH
  1 MAAAAAQGGG GGEPRRTEGV GPGVPGEVEM VKGQPFDVGP RYTQLQYIGE
 51 GAYGMVSSAY DHVRKTRVAI RKISPFEHQT YCQRTLREIQ ILLRFRHENV
101 IGIRDILRAS TLEAMRDVYI VQDLMETDLY KLLKSQQLSN DHICYFLYQI
151 LRGLKYIHSA NVLHRDLKPS NLLINTTCDL KICDFGLARI ADPEHDHTGF
201 LTEYVATRWY RAPEIMLNSK GYTKSIDIWS VGCILAEMLS NRPIFPGKHY
251 LDQLNHILGI LGSPSQEDLN CIINMKARNY LQSLPSKTKV AWAKLFPKSD
301 SKALDLLDRM LTFNPNKRIT VEEALAHPYL EQYYDPTDEP VAEEPFTFAM
351 ELDDLPKERL KELIFQETAR FQPGVLEAP
```

5.7.3 Protein Expression and Purification of p38α

General Procedure for Mutagenesis of pGEX-2T-p38α-Cys172 Plasmid

Since the p38α plasmids possessed in-house originated from donated mutant plasmids (Section 2.11.1), mutagenesis had to be performed to regenerate the wild type sequence. The same mutagenesis procedure to the one previously described (Section 2.11.3) was used to mutate pGEX-2T-p38α-Cys172 plasmid to contain the wild type gene. The following primers were used:

- S119C forward: 5′-GCAGATCTGAATAATATTGTGAAATGCCAGAAACTGACCG-3′
- S119C reverse: 5′-CGGTCAGTTTCTGGCATTTCACAATATTATTCAGATCTGC-3′
- S162C forward: 5′-GCAATCTGGCAGTTAATGAAGATTGCGAACTGAAAATTCTGG-3′
- S162C reverse: 5′-CCAGAATTTTCAGTTCGCAATCTTCATTAACTGCCAGATTGC-3′

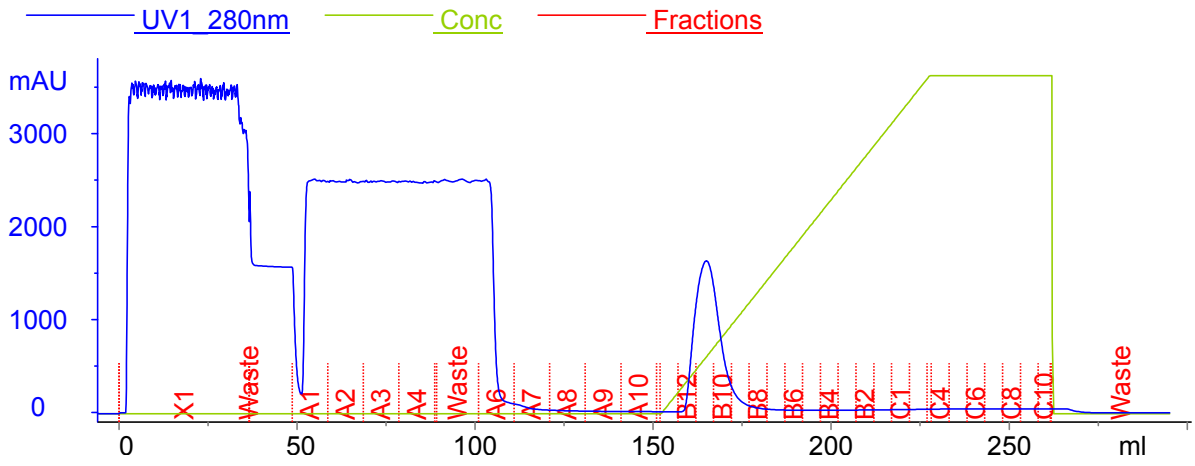
Sequence of pGEX-2T- p38α-WT:

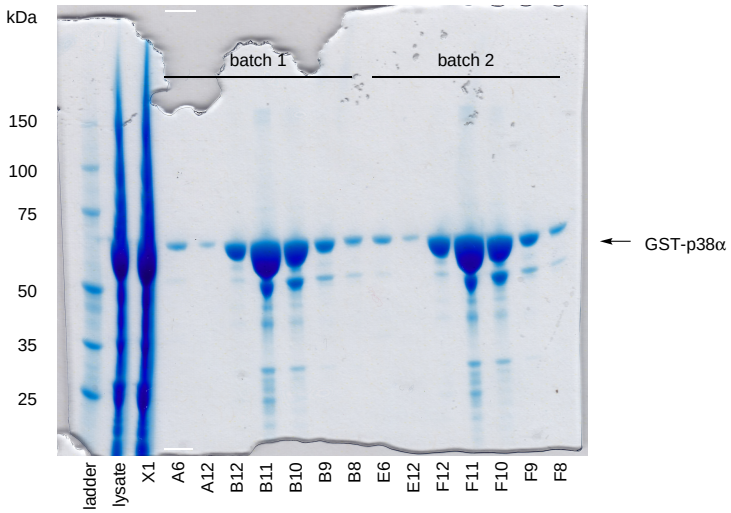
```
   1 ATG AGC CAG GAA CGT CCG ACC TTT TAT CGT CAG GAA CTG AAT AAA
  46 ACC ATT TGG GAA GTG CCG GAA CGT TAT CAG AAT CTG TCT CCG GTT
  91 GGT AGC GGT GCA TAT GGT AGC GTT TGT GCA GCA TTT GAT ACC AAA
 136 ACC GGT CTG CGT GTT GCA GTT AAA AAA CTG AGC CGT CCG TTT CAG
 181 AGC ATT ATT CAT GCC AAA CGT ACC TAT CGT GAA CTG CGT CTG CTG
 226 AAG CAT ATG AAA CAT GAA AAT GTG ATT GGT CTG CTG GAT GTT TTT
 271 ACA CCG GCA CGT AGC CTG GAA GAG TTT AAT GAT GTG TAT CTG GTG
 316 ACA CAT CTG ATG GGT GCA GAT CTG AAT AAT ATT GTG AAA TGC CAG
 361 AAA CTG ACC GAT GAT CAT GTG CAG TTC TTA ATC TAT CAG ATT CTG
 406 CGT GGC CTG AAA TAT ATT CAT AGC GCA GAT ATT ATT CAT CGT GAT
 451 CTG AAA CCG AGC AAT CTG GCA GTT AAT GAA GAT TGC GAA CTG AAA
 496 ATT CTG GAT TTT GGT CTG GCG CGT CAT ACC GAT GAT GAA ATG ACC
 541 GGT TAT GTT GCA ACC CGT TGG TAT CGT GCA CCG GAA ATT ATG CTG
 586 AAT TGG ATG CAT TAT AAT CAG ACC GTG GAT ATT TGG AGC GTT GGT
 631 TGT ATT ATG GCA GAA CTG CTG ACC GGT CGT ACC CTG TTT CCG GGT
 676 ACA GAT CAT ATT GAT CAG CTG AAA CTG ATT CTG CGT CTG GTT GGT
 721 ACA CCG GGT GCC GAA CTG CTG AAA AAA ATT AGC AGC GAA AGC GCA
 766 CGC AAT TAT ATT CAG AGC CTG ACC CAG ATG CCG AAA ATG AAT TTT
 811 GCC AAT GTG TTT ATT GGT GCA AAT CCG CTG GCA GTT GAT CTG CTG
 856 GAA AAA ATG CTG GTT CTG GAT AGC GAT AAA CGT ATT ACC GCA GCA
 901 CAG GCA CTG GCA CAT GCA TAT TTT GCC CAG TAT CAT GAT CCG GAT
 946 GAT GAA CCG GTT GCA GAT CCG TAT GAT CAG AGC TTT GAA AGC CGT
 991 GAT CTG CTG ATT GAT GAA TGG AAA AGC CTG ACC TAT GAT GAA GTG
1036 ATT AGC TTT GTT CCG CCT CCA CTG GAT CAA GAA GAA ATG GAA AGC
1081 TAA
```

Protein Expression and Purification of p38α-WT

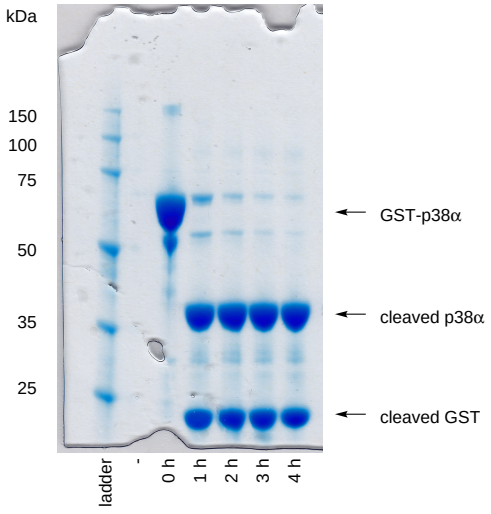
The same protein expression and purification procedure to the one previously described (Section 2.11.3) was used to express and purify p38α-WT. Fractions containing the target protein were determined from the UV trace of the FPLC report file and from SDS-PAGE analysis at each stage of purification. The desired fractions were then collected accordingly. (calculated mass: 41 437, observed mass: 41 436)

GST Affinity Chromatography:

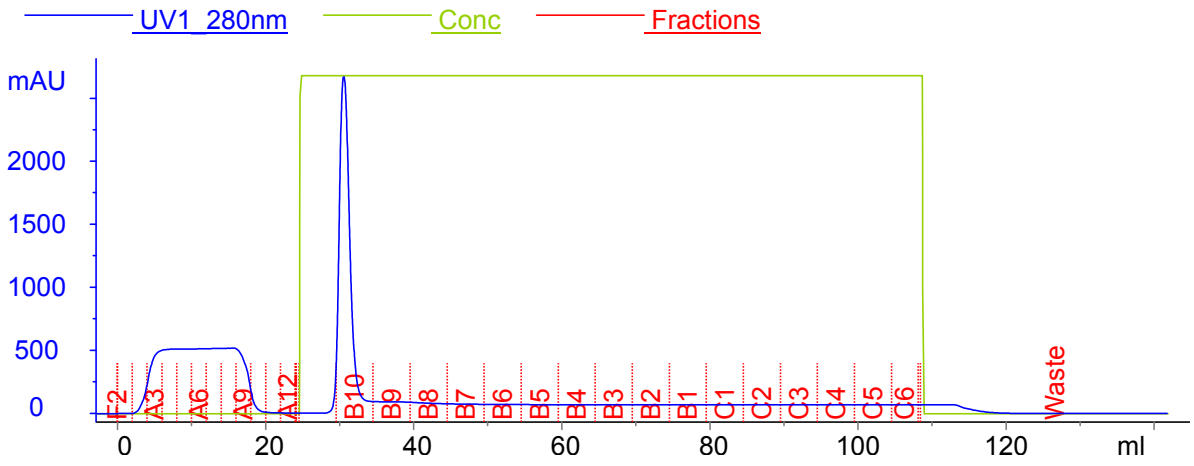


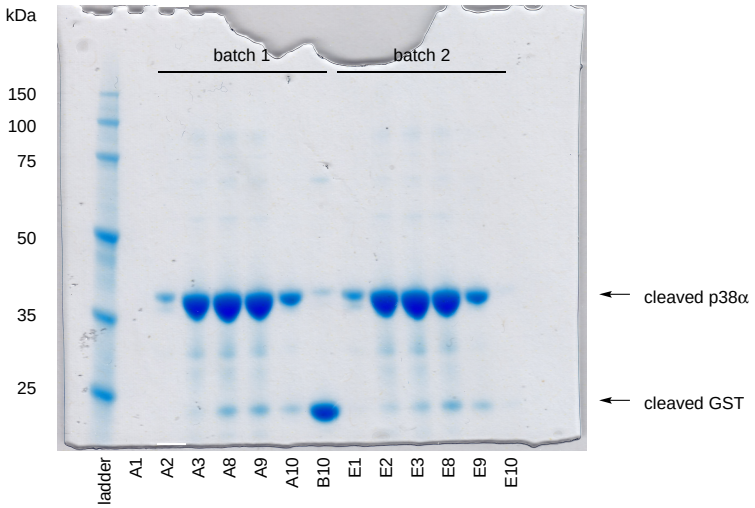


GST Tag Cleavage:

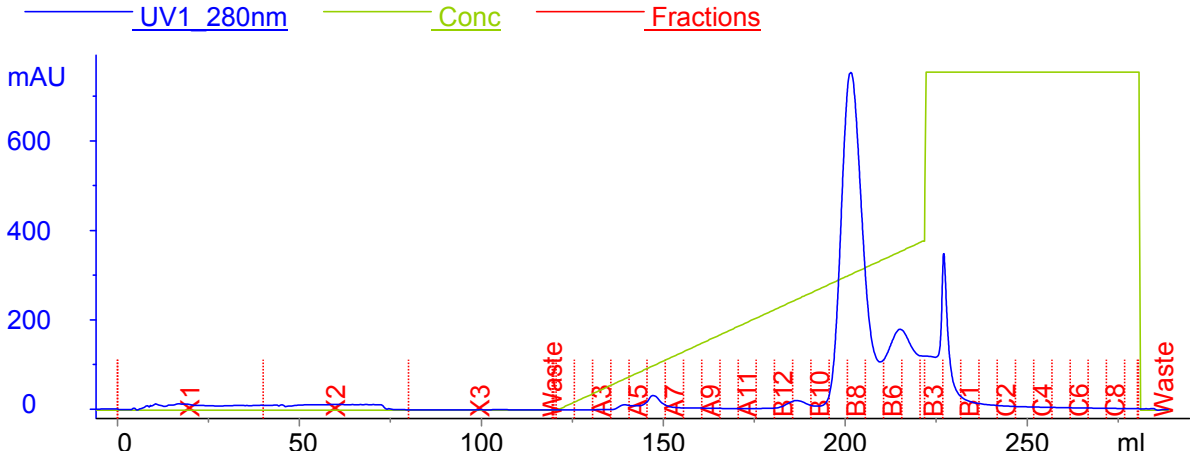


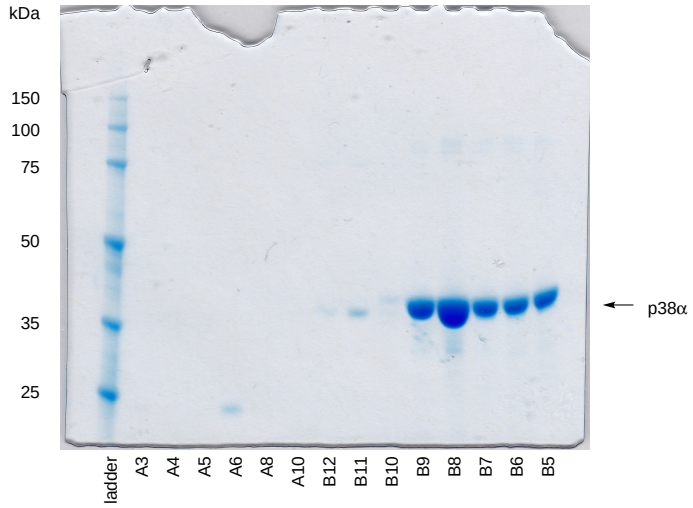
GST Tag Rebinding:



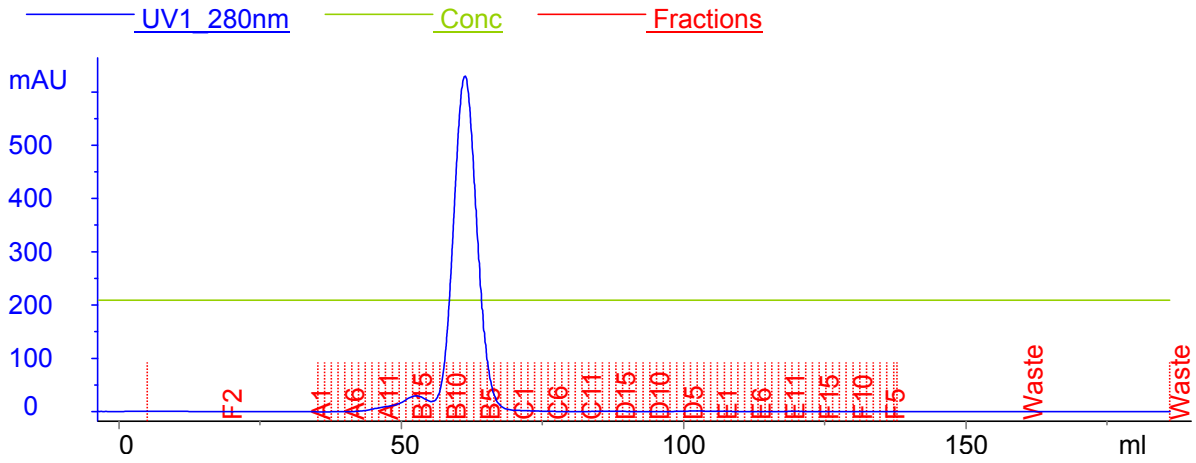


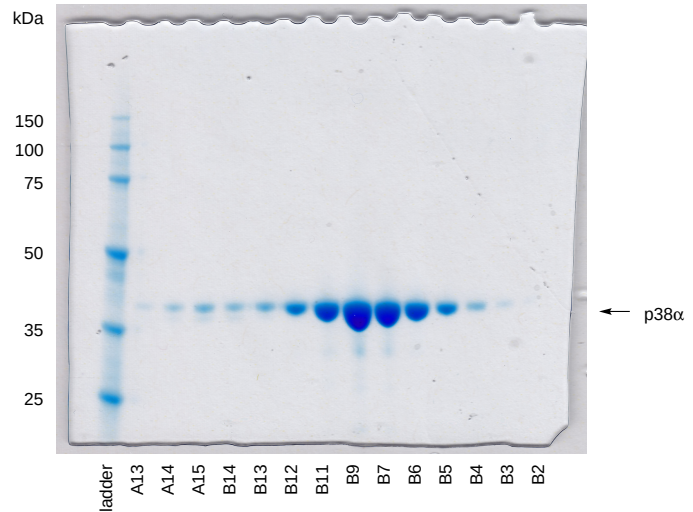
Anion Exchange Chromatography:

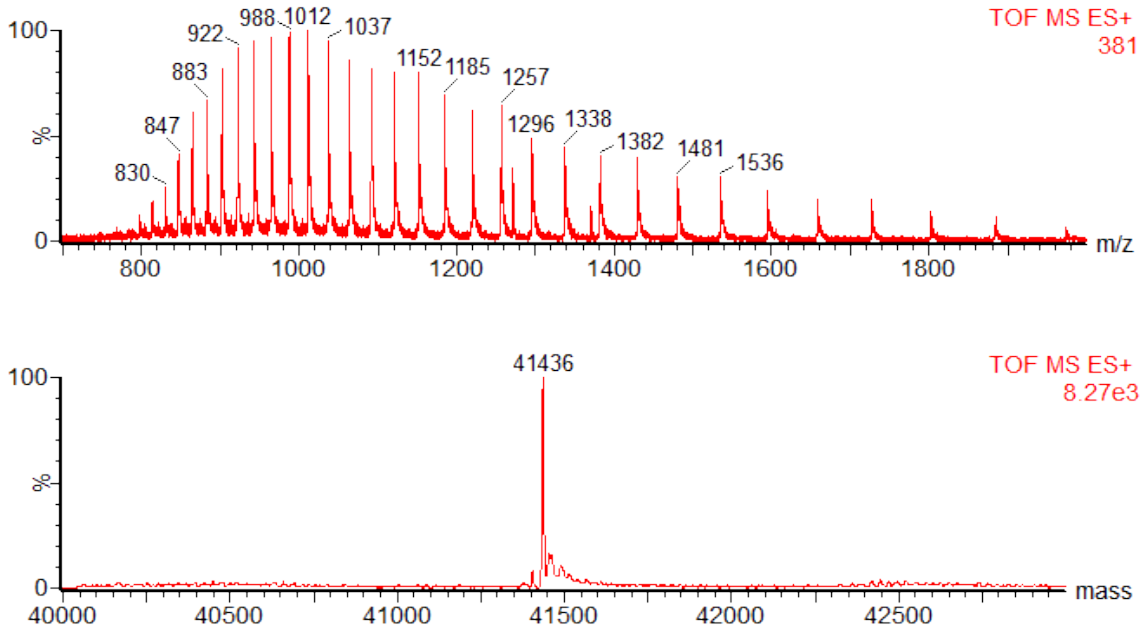




Size Exclusion Chromatography and Storage:







Protein sequence obtained by translation of the corresponding gene sequence:

```
tag                                                          GS
  1 MSQERPTFYR QELNKTIWEV PERYQNLSPV GSGAYGSVCA AFDTKTGLRV
 51 AVKKLSRPFQ SIIHAKRTYR ELRLLKHMKH ENVIGLLDVF TPARSLEEFN
101 DVYLVTHLMG ADLNNIVKCQ KLTDDHVQFL IYQILRGLKY IHSADIIHRD
151 LKPSNLAVNE DCELKILDFG LARHTDDEMT GYVATRWYRA PEIMLNWMHY
201 NQTVDIWSVG CIMAELLTGR TLFPGTDHID QLKLILRLVG TPGAELLKKI
251 SSESARNYIQ SLTQMPKMNF ANVFIGANPL AVDLLEKMLV LDSDKRITAA
301 QALAHAYFAQ YHDPDDEPVA DPYDQSFESR DLLIDEWKSL TYDEVISFVP
351 PPLDQEEMES
```

5.7.4 Measurement of the Activity of MEK1 (Chemical) Variants

The following procedure details the final, optimised procedure used. For the preliminary assays and during optimisation, the same final concentrations of all assay components was kept constant with only the MEK1 chemical variant concentration being varied.

Enzymatic Reaction Setup: Stocks of MEK1 variants (MEK1-Cys222, MEK1-Dha222, MEK1-pCys222, MEK1-S222E and MEK1-A19D), MEK1 substrates (ERK1-K71R and p38α-WT) and ATP (500 mM solution in water, pH 7.0) were thawed on ice. The concentrations of the stocks of MEK1 variants were determined by BCA assay. These stocks were diluted to their working concentrations with the appropriate buffer:

Component	Working Concentration	Buffer
MEK1 Variant	5 µM	MEK1 reaction buffer
MEK1 Substrate	50 µM	ERK1 storage buffer
ATP	2.5 mM	water

Reactions were then assembled with the components:

Addition Order	Component	Concentration	Volume/µL
1	Assay buffer stock	10×	15.0
2	Water	–	60.0
3	Substrate	10 µM	30.0
4	MEK1 variant	5 µM	15.0
Total	–	–	120.0

MEK1 kinase assay buffer (MEK1 KAB) 10× stock was prepared in advance and contained:

Component	Concentration/mM
MOPS	200
β-glycerophosphate	125
EGTA	25
Na_3VO_4	5
$MgCl_2$	100
DTT	10

Reaction Initialisation and Sample Collection: A sample (4 µL) of the reaction mixture was taken and quenched in pre-aliquoted guanidine solution (9 M solution in MEK1 reaction buffer, supersaturated). Additional water (1 µL) was added to this sample. The enzymatic reaction was then started by addition of ATP (29.0 µL of a 2.5 µM solution in water, pH 7.0) and incubated with shaking at 30 °C, 400 rpm. Further samples (5 µL) were taken at regular timepoints (5, 10, 20 and 30 min, then every 15 min for 6 h from reaction initialisation) and similarly quenched in guanidine solution. The samples were then analysed by LC-MS.

Data Analysis: The LC-MS data collected was analysed using MassLynx 4.1 as previously described (Section 2.11.1), with the addition of using an AutoIt script to drive the operator-controlled processes in processing the spectra. Data from the deconvoluted spectra were output from MassLynx as a spectrum list for quantitative analysis using MATLAB R2013b. Working examples of the scripts used are in Appendices F and B respectively.

In MATLAB, quantities of the three substrate species (unphosphorylated, mono-phosphorylated and bis-phosphorylated) were estimated from relative peak intensity, setting the baseline at the median of all the values in the spectrum list. The peaks were picked by taking the maximum intensities from the three ranges where the peaks of these species were expected. After plotting the protein of each species against time, the peak proportions were normalised to the starting substrate concentration. Initial rates were then estimated by linear regression (using the `polyfit` function) of the first 5–8 timepoints corresponding to each substrate phosphoform. The gradient was taken as the initial rate and the standard error on the rate was estimated by inputting the [S] vs. time data into Graphpad PRISM v.5.01.

Chapter 6

Method Outlook

The course of this work addressed the question of whether phosphocysteine, installed by means of chemical modification, would make suitable surrogates for the naturally occurring phosphothreonine and phosphoserine. The chemical modification methods were first applied to p38α, producing both p38α-pCys172 and p38α-pCys180. With the latter chemical variant being the mimic of the naturally occurring p38α-pThr180,[51,52] p38α-pCys180 was evaluated for its enzymatic ability. It was found that p38α-pCys180 had similar enzymatic activity to p38α-pThr180, thus demonstrating that the phosphocysteine installed by chemical modification is a sufficient mimic of phosphothreonine for the functional investigation on kinases. Closer analysis of the enzyme kinetic data also revealed that p38α-pCys180 behaved differently from active p38α-WT, having a different propensity in phosphorylating its substrate, ATF2. The exact consequences of this finding are unclear but they are likely to contribute to the alternative regulatory pathways that p38α has been implicated in.[51]

The chemical modification methodology was also applied to MEK1 to produce MEK1-pCys222 as a mimic of MEK1-pSer222, and its enzymatic competency was also evaluated. The evidence collected showed that phosphocysteine resulted in a more active MEK1 species than one activated by acidic residue substitution, MEK1-S222E. This suggested that phosphocysteine was a better substitute for phosphoserine than glutamate was, but no firm conclusion could be drawn since there is no suitable, native comparison to control against. However, to ensure that the method is acceptable as an alternative tool for investigating serine phosphorylation as well as threonine phosphorylation in kinases, phosphocysteine on MEK1 must be fully evaluated against phosphoserine. This could be achieved by comparing MEK1-pCys218pCys222

with MEK1-pSer218pSer222, provided phosphocysteine can be successfully installed at position 218 on MEK1 as well.

The lack of a suitable, native control for the enzymatic activity of MEK1-pCys222 in fact demonstrates the potential for protein chemical modification as an alternative investigative tool in the biological characterisation of kinase phosphoforms: no suitable native control can be produced since there is no available enzyme that can selectively form its native counterpart. MEK1-pCys222, with the other activation loop phosphorylation site intact as Ser218, is an arrangement that can only be obtained through synthetic means. The full evaluation of phosphocysteine against phosphoserine would of course simply broaden the scope with which this method can be used, from kinases containing threonine phosphorylation sites to ones containing Ser phosphorylation sites as well. Even without this validation, the implications of using chemical modification as an alternative tool for the study of protein kinase interactions can be considered.

Most immediately, provided the likely outcome that phosphocysteine is showed to be a sufficient mimic of phosphoserine, the independent installation of phosphocysteine in the MEK1 variants MEK1-Cys218 and MEK1-Cys222 will allow the individual contributions to function of the two phosphorylation sites to be discerned. Through the functional and enzymatic evaluation of the synthesised chemical variants, similar to the ones already undertaken, the results from furthering the chemical modification study could then be used to corroborate or reject the conclusions of previous studies that were done with acidic residue substitutes.[12,22]

The same format of studies involving cysteine substitution variant design, variant chemical modification and enzymatic evaluation as proposed for MEK1 could also be extended to other kinases, which natively and typically occur in phosphoforms containing multiple phosphorylations. A study on Aurora A kinase has already benefitted from these chemical modification methods in discerning the individual functions of its two activation loop phosphorylation sites,[48] thus demonstrating the generality of this format of study.

The more detailed biochemical studies that could result from the use of chemically generated kinase phosphoforms would also help in the parameterisation required for accurate modelling studies of kinase signalling cascades and networks.[144] Many of these models aim to rationalise the overall effect of a signalling cascade and rely on experimental measurements, such as those for reaction rates and binding affinities to give values for these parameters that are relevant to the species within the model. More faithful models would allow for a better holistic understanding of signalling networks.

Ultimately, it is the interactomic study of a kinase's different phosphoforms where the use of chemically modified recombinant kinases could make the largest impact, particularly in the realms of biological target validation, since it may be the case that different phosphoforms of the same kinase may interact differently with other endogenous proteins. A number of methods could be used to enact this. One method would be to use affinity pull-down, where the pre-prepared, affinity tagged kinase phosphoform is bound to the appropriate affinity resin and incubated with whole cell lysate to determine what binds to the phosphoform. Elution of the proteins and subsequent LC-MS/MS analysis would then reveal the binding partners. Another method would be to radiolabel the substrate interactions using [γ-^{32}P]-ATP. LC-MS/MS analysis of the radiolabelled phosphopeptides could then reveal the substrate interactions. This type of study could in turn help to elucidate more complex mechanistic behaviours.[145]

The reason for the envisaged wider applicability of the chemical modification method for making phosphorylated kinase mimics is due to the ability to rationally design the variants required for synthesising the kinase phosphoforms of interest. Building upon the advantage of scalability in methods used for producing recombinant proteins, protein chemical modification is also easily scalable, allowing for the production of large quantities of kinase chemical variants that may be required, particularly for structural and crystallographic studies. Key to the future success of all these studies is the chemical variant homogeneity that chemical modification is able to provide, thus giving the confidence that the only phosphoform that is present in the study is the one that has been designed.

References

[1] K. P. Chooi, S. R. G. Galan, R. Raj, J. McCullagh, S. Mohammed, L. H. Jones and B. G. Davis, *J. Am. Chem. Soc.*, 2014, **136**, 1698–1701.

[2] G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Sci. Signal.*, 2002, **298**, 1912–1934.

[3] M. A. Lemmon and J. Schlessinger, *Cell Cycle*, 2010, **141**, 1117–1134.

[4] L. N. Johnson, E. D. Lowe, M. E. M. Noble and D. J. Owen, *FEBS Lett.*, 1998, **430**, 1–11.

[5] T. O. Fischmann, C. K. Smith, T. W. Mayhood, J. E. Myers, P. Reichert, A. Mannarino, D. Carr, H. Zhu, J. Wong, R.-S. S. Yang, H. V. Le and V. S. Madison, *Biochemistry*, 2009, **48**, 2661–2674.

[6] J. Zhang, P. L. Yang and N. S. Gray, *Nat. Rev. Cancer*, 2009, **9**, 28–39.

[7] D. K. Treiber and N. P. Shah, *Chem. Biol.*, 2013, **20**, 745–746.

[8] B. Nagar, W. G. Bornmann, P. Pellicena, T. Schindler, D. R. Veach, W. T. Miller, B. Clarkson and J. Kuriyan, *Cancer Res.*, 2002, **62**, 4236–4243.

[9] G. Sun and R. J. A. Budde, *Biochemistry*, 1997, **36**, 2139–2146.

[10] W. Huang, D. S. Kessler and R. L. Erikson, *Mol. Biol. Cell*, 1995, **6**, 237–245.

[11] D. J. Robbins, E. Zhen, H. Owaki, C. A. Vanderbilt, D. Ebert, T. D. Geppert and M. H. Cobb, *J. Biol. Chem.*, 1993, **268**, 5097–5106.

[12] D. R. Alessi, Y. Saito, D. G. Campbell, P. Cohen, G. Sithanandam, U. Rapp, A. Ashworth, C. J. Marshall and S. Cowley, *EMBO J.*, 1994, **13**, 1610–1619.

[13] K. Gopalbhai, G. Jansen, G. Beauregard, M. Whiteway, F. Dumas, C. Wu and S. Meloche, *J. Biol. Chem.*, 2003, **278**, 8118–8125.

[14] T. Krell, J. Lacal, A. Busch, H. S. Jiménez, M. E. Guazzaroni and J. L. Ramos, *Annu. Rev. Microbiol.*, 2010, **64**, 539–559.

[15] T. C. Stadtman, *BioFactors*, 1994, **4**, 181–185.

[16] F. Sun, Y. Ding, Q. Ji, Z. Liang, X. Deng, C. C. L. Wong, C. Yi, L. Zhang, S. Xie, S. Alvarez, L. M. Hicks, C. Luo, H. Jiang, L. Lan and C. He, *Proc. Nat. Acad. Sci.*, 2012, **109**, 15461–15466.

[17] J. A. Adams, *Chem. Rev.*, 2001, **101**, 2271–2290.

[18] F. Diella, S. Cameron, C. Gemund, R. Linding, A. Via, B. Kuster, T. S. Ponten, N. Blom and T. Gibson, *BMC Bioinformatics*, 2004, **5**, 79+.

[19] Z. A. Knight, B. Schilling, R. H. Row, D. M. Kenski, B. W. Gibson and K. M. Shokat, *Nat. Biotechnol.*, 2003, **21**, 1047–1054.

[20] H. Molina, D. M. Horn, N. Tang, S. Mathivanan and A. Pandey, *Proc. Nat. Acad. Sci.*, 2007, **104**, 2199–2204.

[21] J. A. Adams, *Biochemistry*, 2003, **42**, 601–607.

[22] C. F. Zheng and K. L. Guan, *EMBO J.*, 1994, **13**, 1123–1131.

[23] A. Brunet, G. Pagès and J. Pouysségur, *Oncogene*, 1994, **9**, 3379–3387.

[24] K. A. Resing, S. J. Mansour, A. S. Hermann, R. S. Johnson, J. M. Candia, K. Fukasawa, G. F. Vande Woude and N. G. Ahn, *Biochemistry*, 1995, **34**, 2610–2620.

[25] S. P. Chambers, D. A. Austen, J. R. Fulghum and W. M. Kim, *Protein Expres. Purif.*, 2004, **36**, 40–47.

[26] J. P. Turowec, J. S. Duncan, A. C. French, L. Gyenis, N. A. St. Denis, G. Vilk and D. W. Litchfield, in *Protein Kinase CK2 is a Constitutively Active Enzyme that Promotes Cell Survival: Strategies to Identify CK2 Substrates and Manipulate its Activity in Mammalian Cells*, Elsevier, 2010, vol. 484, pp. 471–493.

[27] P. Lahiry, A. Torkamani, N. J. Schork and R. A. Hegele, *Nat. Rev. Genet.*, 2010, **11**, 60–74.

[28] R. Diskin, N. Askari, R. Capone, D. Engelberg and O. Livnah, *J. Biol. Chem.*, 2004, **279**, 47040–47049.

[29] N. Tzarum, R. Diskin, D. Engelberg and O. Livnah, *J. Mol. Biol.*, 2011, **405**, 1154–1169.

[30] M. Bell, R. Capone, I. Pashtan, A. Levitzki and D. Engelberg, *J. Biol. Chem.*, 2001, **276**, 25351–25358.

[31] W. Huang and R. L. Erikson, *Proc. Nat. Acad. Sci.*, 1994, **91**, 8960–8963.

[32] C. K. Smith, D. Carr, T. W. Mayhood, W. Jin, K. Gray and W. T. Windsor, *Protein Expres. Purif.*, 2007, **52**, 446–456.

[33] Y.-Y. Zhang, Z.-Q. Mei, J.-W. Wu and Z.-X. Wang, *J. Biol. Chem.*, 2008, **283**, 26591–26601.

[34] R. Akella, X. Min, Q. Wu, K. H. Gardner and E. J. Goldsmith, *Structure*, 2010, **18**, 1571–1578.

[35] Millipore, *p38 alpha/SAPK2a, active, Item No.: #14-587*.

[36] G. A. Keesler, J. Bray, J. Hunt, D. A. Johnson, T. Gleason, Z. Yao, S.-W. Wang, C. Parker, H. Yamane, C. Cole and H. S. Lichenstein, *Protein Expres. Purif.*, 1998, **14**, 221–228.

[37] P. A. Cole, A. D. Courtney, K. Shen, Z. Zhang, Y. Qiao, W. Lu and D. M. Williams, *Acc. Chem. Res.*, 2003, **36**, 444–452.

[38] T. L. Lamoureaux and D. H. Lee, *Chem. Commun.*, 2011, **47**, 8623–8625.

[39] W. W. C. Chan, *Biochemistry*, 1968, **7**, 4247–4254.

[40] T. Ruman, K. Długopolska, A. Jurkiewicz, D. Rut, T. Frczyk, J. Cieśla, A. Leś, Z. Szewczuk and W. Rode, *Bioorg. Chem.*, 2010, **38**, 74–80.

[41] J. M. Fujitaki, G. Fung, E. Y. Oh and R. A. Smith, *Biochemistry*, 1981, **20**, 3658–3664.

[42] G. S. Lukat, W. R. Mccleary, A. M. Stock and J. B. Stock, *Proc. Nat. Acad. Sci.*, 1992, **89**, 718–722.

[43] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernandez-Gonzalez, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield and B. G. Davis, *Chem. Sci.*, 2011, **2**, 1666–1676.

[44] J. M. Chalker, *Ph.D. thesis*, University of Oxford, Department of Chemistry, 2009.

[45] R. Nathani, P. Moody, M. E. B. Smith, R. J. Fitzmaurice and S. Caddick, *ChemBioChem*, 2012, n/a.

[46] G. J. L. Bernardes, J. M. Chalker, J. C. Errey and B. G. Davis, *J. Am. Chem. Soc.*, 2008, **130**, 5052–5053.

[47] J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield and B. G. Davis, *Angew. Chem. Int. Ed.*, 2012, **51**, 1835–1839.

[48] F. C. Rowan, M. Richards, R. A. Bibby, A. Thompson, R. Bayliss and J. Blagg, *ACS Chem. Biol.*, 2013, **8**, 2184–2191.

[49] T. Shiraishi, S. Matsuyama and H. Kitano, *PLoS Comput. Biol.*, 2010, **6**, e1000851+.

[50] G. Pearson, F. Robinson, T. Beers Gibson, B.-e. Xu, M. Karandikar, K. Berman and M. H. Cobb, *Endocr. Rev.*, 2001, **22**, 153–183.

[51] P. R. Mittelstadt, H. Yamaguchi, E. Appella and J. D. Ashwell, *J. Biol. Chem.*, 2009, **284**, 15469–15474.

[52] N. Askari, J. Beenstock, O. Livnah and D. Engelberg, *Biochemistry*, 2009, **48**, 2497–2504.

[53] P. Cohen and D. R. Alessi, *ACS Chem. Biol.*, 2013, **8**, 96–104.

[54] J. Arrowsmith and P. Miller, *Nat. Rev. Drug Discov.*, 2013, **12**, 569.

[55] D. Vandamme, B. A. Minke, W. Fitzmaurice, B. N. Kholodenko and W. Kolch, *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 2014, **6**, 1–11.

[56] T. Mustelin, in *A brief introduction to the protein phosphatase families*, Humana Press, Clifton, N.J., 2007, vol. 365, pp. 9–22.

[57] M. J. Robinson and M. H. Cobb, *Curr. Opin. Cell Biol.*, 1997, **9**, 180–186.

[58] H. Rubinfeld and R. Seger, *Mol. Biotechnol.*, 2005, **31**, 151–174.

[59] Y. Keshet and R. Seger, *Methods Mol. Biol.*, 2010, **661**, 3–38.

[60] E. F. Wagner and A. R. Nebreda, *Nat. Rev. Cancer*, 2009, **9**, 537–549.

[61] A. G. Turjanski, J. P. Vaque and J. S. Gutkind, *Oncogene*, 2007, **26**, 3240–3253.

[62] G. L. Schieven, *Curr. Top. Med. Chem.*, 2009, **9**, 1038–1048.

[63] C. Pargellis, L. Tong, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss, S. Pav and J. Regan, *Nat. Struct. Biol.*, 2002, **9**, 268–272.

[64] M. Lee and C. Dominguez, *Curr. Med. Chem.*, 2005, **12**, 2979–2994.

[65] J. J. Ventura, S. Tenbaum, E. Perdiguero, M. Huth, C. Guerra, M. Barbacid, M. Pasparakis and A. R. Nebreda, *Nat. Genet.*, 2007, **39**, 750–758.

[66] J. Ren, S. Zhang, A. Kovacs, Y. Wang and A. Muslin, *J. Mol. Cell. Cardiol.*, 2005, **38**, 617–623.

[67] L. Xing, H. S. Shieh, S. R. Selness, R. V. Devraj, J. K. Walker, B. Devadas, H. R. Hope, R. P. Compton, J. F. Schindler, J. L. Hirsch, A. G. Benson, R. G. Kurumbail, R. A. Stegeman, J. M. Williams, R. M. Broadus, Z. Walden and J. B. Monahan, *Biochemistry*, 2009, **48**, 6402–6411.

[68] R. Seger, N. G. Ahn, J. Posada, E. S. Munar, A. M. Jensen, J. A. Cooper, M. H. Cobb and E. G. Krebs, *J. Biol. Chem.*, 1992, **267**, 14373–14381.

[69] P. J. Roberts and C. J. Der, *Oncogene*, 2007, **26**, 3291–3310.

[70] Y. Zhao and A. A. Adjei, *Nat. Rev. Clin. Oncol.*, 2014, **11**, 385–400.

[71] J. R. Simard, M. Getlik, C. Grutter, V. Pawar, S. Wulfert, M. Rabiller and D. Rauh, *J. Am. Chem. Soc.*, 2009, **131**, 13286–13296.

[72] M. Getlik, J. R. Simard, M. Termathe, C. Grütter, M. Rabiller, W. A. L. van Otterlo and D. Rauh, *PLoS ONE*, 2012, **7**, e39713+.

[73] S. B. Patel, P. M. Cameron, B. Frantz-Wattley, E. O'Neill, J. W. Becker and G. Scapin, *BBA-Mol. Cell Res.*, 2004, **1696**, 67–73.

[74] V. V. Mozhaev, Y. L. Khmelnitsky, M. V. Sergeeva, A. B. Belova, N. L. Klyachko, A. V. Levashov and K. Martinek, *Eur. J. Biochem.*, 1989, **184**, 597–602.

[75] V. Vagenende, M. G. S. Yap and B. L. Trout, *Biochemistry*, 2009, **48**, 11084–11096.

[76] K. Gekko and S. N. Timasheff, *Biochemistry*, 1981, **20**, 4667–4676.

[77] M. W. Washabaugh and K. D. Collins, *J. Biol. Chem.*, 1986, **261**, 12477–12485.

[78] G. L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70–77.

[79] A. F. S. A. Habeeb, in *Reaction of protein sulfhydryl groups with Ellman's reagent*, Elsevier, 1972, vol. 25, ch. 37, pp. 457–464.

[80] G. E. Begg and D. W. Speicher, *J. Biomol. Tech.*, 1999, **10**, 17–20.

[81] K. Talley and E. Alexov, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 2699–2706.

[82] N. J. Greenfield, *Nat. Protoc.*, 2007, **1**, 2527–2535.

[83] F. H. Niesen, H. Berglund and M. Vedadi, *Nat. Protoc.*, 2007, **2**, 2212–2221.

[84] M. N. Schulz, J. Landström and R. E. Hubbard, *Anal. Biochem.*, 2013, **433**, 43–47.

[85] I. R. León, V. Schwämmle, O. N. Jensen and R. R. Sprenger, *Mol. Cell. Proteomics*, 2013, **12**, 2992–3005.

[86] T. Scientific, *In-Solution Tryptic Digestion and Guanidination Kit*.

[87] A. M. Palumbo, J. J. Tepe and G. E. Reid, *J. Proteome Res.*, 2008, **7**, 771–779.

[88] L. Quan and M. Liu, *Mod. Chem. Appl.*, 2013, **01**, 1:e102+.

[89] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, in *Protein Identification and Analysis Tools on the ExPASy Server*, ed. J. M. Walker, Humana Press, 2005, pp. 571–607.

[90] S. C. Gill and P. H. von Hippel, *Anal. Biochem.*, 1989, **182**, 319–326.

[91] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, *Nucleic Acids Res.*, 2007, **35**, W522–W525.

[92] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32**, W665–W667.

[93] N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Nat. Acad. Sci.*, 2001, **98**, 10037–10041.

[94] S. J. Hubbard and J. M. Thornton, *Ph.D. thesis*, University College London, 1993.

[95] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nat. Biotechnol.*, 2007, **446**, 1105–1109.

[96] J. Cox and M. Mann, *Nat. Biotechnol.*, 2008, **26**, 1367–1372.

[97] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen and M. Mann, *J. Proteome Res.*, 2011, **10**, 1794–1805.

[98] D. G. Adams, Y. Wang, P. A. Mak, J. Chyba, O. Shalizi, J. Matzen, P. Anderson, T. R. Smith, M. Garcia, G. L. Welch, E. J. Claret, M. Fink, A. P. Orth, J. S. Caldwell and A. Brinker, *Curr. Chem. Genomics*, 2008, **1**, 54–64.

[99] G. Chen, M. D. Porter, J. R. Bristol, M. J. Fitzgibbon and S. Pazhanisamy, *Biochemistry*, 2000, **39**, 2079–2087.

[100] C. J. Hastie, H. J. McLauchlan and P. Cohen, *Nat. Protoc.*, 2006, **1**, 968–971.

[101] K. M. Sours, S. C. Kwok, T. Rachidi, T. Lee, A. Ring, A. N. Hoofnagle, K. A. Resing and N. G. Ahn, *J. Mol. Biol.*, 2008, **379**, 1075–1093.

[102] Millipore, *p38alpha/SARK2a (6His-tag), unactive*, 2008.

[103] P. Du, P. Loulakis, C. Luo, A. Mistry, S. P. Simons, P. K. LeMotte, F. Rajamohan, K. Rafidi, K. G. Coleman and K. F. Geoghegan, *Protein Expres. Purif.*, 2005, **44**, 121–129.

[104] Z. Yan, G. W. Caldwell and P. A. McDonell, *Biochem. Bioph. Res. Co.*, 1999, **262**, 793–800.

[105] S. Schnell and P. K. Maini, *Comments on Theoretical Biology*, 2003, 169–187.

[106] P. V. LoGrasso, B. Frantz, A. M. Rolando, S. J. O'Keefe, J. D. Hermes and E. A. O'Neill, *Biochemistry*, 1997, **36**, 10422–10427.

[107] W. F. Waas, H.-H. Lo and K. N. Dalby, *J. Biol. Chem.*, 2001, **276**, 5676–5684.

[108] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Portland Press, London, 2004.

[109] G. R. Stark, W. H. Stein and S. Moore, *J. Biol. Chem.*, 1960, **235**, 3177–3181.

[110] G. R. Stark and D. G. Smyth, *J. Biol. Chem.*, 1963, **238**, 214–226.

[111] J. P. Duffy, E. M. Harrington, F. G. Salituro, J. E. Cochran, J. Green, H. Gao, G. W. Bemis, G. Evindar, V. P. Galullo, P. J. Ford, U. A. Germann, K. P. Wilson, S. F. Bellon, G. Chen, P. Taslimi, P. Jones, C. Huang, S. Pazhanisamy, Y.-M. Wang, M. A. Murcko and M. S. S. Su, *ACS Med. Chem. Lett.*, 2011, **2**, 758–763.

[112] S. Miwatashi, Y. Arikawa, E. Kotani, M. Miyamoto, K.-I. Naruo, H. Kimura, T. Tanaka, S. Asahi and S. Ohkawa, *J. Med. Chem.*, 2005, **48**, 5966–5979.

[113] C. L. Manthey, S.-W. Wang, S. D. Kinney and Z. Yao, *J. Leukoc. Biol.*, 1998, **64**, 409–417.

[114] J. F. Braganza, M. A. Letavic and K. F. McClure, *Patent: WO2004072072*, 2004, 1–99.

[115] Y. Friedmann, A. Shriki, E. R. Bennett, S. Golos, R. Diskin, I. Marbach, E. Bengal and D. Engelberg, *Mol. Pharmacol.*, 2006, **70**, 1395–1405.

[116] R. Azevedo, M. van Zeeland, H. Raaijmakers, B. Kazemier, J. de Vlieg and A. Oubrie, *Acta Crystallogr. Sect. D*, 2012, **68**, 1041–1050.

[117] B. K. Shoichet, *J. Med. Chem.*, 2006, **49**, 7274–7277.

[118] D. J. Templeton, M.-S. Aye, J. Rady, F. Xu and J. V. Cross, *PLoS ONE*, 2010, **5**, e15012+.

[119] R. Diskin, M. Lebendiker, D. Engelberg and O. Livnah, *J. Mol. Biol.*, 2007, **365**, 66–76.

[120] J. Regan, C. A. Pargellis, P. F. Cirillo, T. Gilmore, E. R. Hickey, G. W. Peet, A. Proto, A. Swinamer and N. Moss, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 3101–3104.

[121] D. S. Millan, M. E. Bunnage, J. L. Burrows, K. J. Butcher, P. G. Dodd, T. J. Evans, D. A. Fairman, S. J. Hughes, I. C. Kilty, A. Lemaitre, R. A. Lewthwaite, A. Mahnke, J. P. Mathias, J. Philip, R. T. Smith, M. H. Stefaniak, M. Yeadon and C. Phillips, *J. Med. Chem.*, 2011, **54**, 7797–7814.

[122] R. M. Angell, T. D. Angell, P. Bamborough, M. J. Bamford, C.-w. Chung, S. G. Cockerill, S. S. Flack, K. L. Jones, D. I. Laine, T. Longstaff, S. Ludbrook, R. Pearson, K. J. Smith, P. A. Smee, D. O. Somers and A. L. Walker, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 4433–4437.

[123] J. R. Wiśniewski, A. Zougman, N. Nagaraj and M. Mann, *Nat. Methods*, 2009, **6**, 359–362.

[124] *doseResponse - File Exchange*, 2011, accessed Aug 2013, `http://www.mathworks.co.uk/matlabcentral/fileexchange/33604-doseresponse`.

[125] G. Pagès, A. Brunet, G. L'Allemain and J. Pouysségur, *EMBO J.*, 1994, **13**, 3003–3010.

[126] K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin and P. K. LeMotte, *Anal. Biochem.*, 1999, **267**, 169–184.

[127] A. Aitken and M. Learmonth, in *The Protein Protocols Handbook*, ed. J. M. Walker, Humana Press, New Jersey, 2002, ch. 59, pp. 455–456.

[128] L. Johansson, G. Gafvelin and E. S. J. Arnér, *BBA-Mol. Cell Res.*, 2005, **1726**, 1–13.

[129] J. Wang, S. M. Schiller and P. G. Schultz, *Angew. Chem. Int. Ed.*, 2007, **46**, 6849–6851.

[130] A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50–69.

[131] L. L. Y. Lee and J. C. Lee, *Biochemistry*, 1987, **26**, 7813–7819.

[132] B. M. Baynes, D. I. C. Wang and B. L. Trout, *Biochemistry*, 2005, **44**, 4919–4925.

[133] N. J. Greenfield, *Nat. Protocols*, 2007, **1**, 2876–2890.

[134] M. Scheepstra, *M.Sc. thesis*, University of Oxford, 2012.

[135] I. Wortzel and R. Seger, *Genes Cancer*, 2011, **2**, 195–209.

[136] T. A. J. Haystead, P. Dent, J. Wu, C. M. M. Haystead and T. W. Sturgill, *FEBS Lett.*, 1992, **306**, 17–22.

[137] J. E. Ferrell and R. R. Bhatt, *J. Biol. Chem.*, 1997, **272**, 19008–19016.

[138] M. J. Robinson, M. Cheng, A. Khokhlatchev, D. Ebert, N. Ahn, K.-L. Guan, B. Stein, E. Goldsmith and M. H. Cobb, *J. Biol. Chem.*, 1996, **271**, 29734–29739.

[139] C. N. Prowse and J. Lew, *J. Biol. Chem.*, 2001, **276**, 99–103.

[140] B.-e. Xu, J. L. Wilsbacher, T. Collisson and M. H. Cobb, *J. Biol. Chem.*, 1999, **274**, 34029–34035.

[141] J. Wu, A. J. Rossomando, J.-H. Her, R. Del Vecchio, M. J. Weber and T. W. Sturgill, *Proc. Nat. Acad. Sci.*, 1991, **88**, 9508–9512.

[142] H. Li, H. Xu, Y. Zhou, J. Zhang, C. Long, S. Li, S. Chen, J.-M. Zhou and F. Shao, *Sci. Signal.*, 2007, **315**, 1000–1003.

[143] W. S. VanScyoc, G. A. Holdgate, J. E. Sullivan and W. H. J. Ward, *Biochemistry*, 2008, **47**, 5017–5027.

[144] C.-Y. F. Huang and J. E. Ferrell, *Proc. Nat. Acad. Sci.*, 1996, **93**, 10078–10083.

[145] P. A. Lochhead, *Sci. Signal.*, 2009, **2**, pe4+.

Appendix A

MATLAB Scripts for Analysis and Manipulation of Gene Sequences

A.1 geneservice_analysis.m

```
% This script is for use with gene sequencing data that is received from
% gene sequencing. It is a shell script that loads both the forward and
% reverse sequences from sequencing, splices them together to give a
% consensus sequence and compares it base-by-base using the
% compare_sequence function

function [overlap,miscalls,mismatches]=geneservice_analysis(forward,...
    reverse,template)

% clear all
% clc

warning off
```

Deal with point mutations

This section is for dealing with point mutations that have been discovered in the sequence from gene sequencing. It is an override to see whether the middle of the gene is in fact being covered twice

```
% template(741)='T'; % needed as the gene seems to have been mutated for
% NDR2
```

produce the coding strand sequence from the reverse sequence

```
r_inverted=invert_sequence(gene_compliment(reverse));
```

sets the length of sequence that we look for for a match

```
match_length=50;
template_midpoint=length(template)/2;
```

Alignment 1

This module is for brute force alignment of the template and forward sequences

```
f_counter=-1;
f_start_find=[];

% grow forward sequence to at least the length of the template
for i=1:length(template)
    forward=strcat(forward,'N');
end

% find the index where the template sequence starts on the sequence from
% sequencing
while numel(f_start_find)==0 && f_counter<length(template)-match_length
    f_counter=f_counter+1;
    f_start_find=strfind(forward,template(f_counter+1:f_counter+...
        match_length));
end

% find the start position of the gene on the forward sequence
f_start_pos=f_start_find-f_counter; % position relative to raw
%forward sequence

f_cropped=forward(f_start_pos:f_start_pos+length(template)-1);
% sequence is same length as template
```

Alignment 2

This module is for brute force alignment of the template and reverse sequences

```
r_counter=-1;
r_start_find=[];

for i=1:length(template)
    r_inverted=strcat('N',r_inverted);
end % grow forward sequence to at least the length of the template

% find the index where the template sequence starts on the sequence from
% sequencing
while numel(r_start_find)==0 && r_counter<length(template)-match_length
    r_counter=r_counter+1;
    r_start_find=strfind(r_inverted,template(length(template)-...
        (match_length-1)-r_counter:length(template)-r_counter))+...
        (match_length-1);
end

% find the end position of the gene on the reverse sequence
r_end_pos=r_start_find+r_counter;

r_cropped=r_inverted(r_end_pos-length(template)+1:r_end_pos);
```

```
% sequence is same length as template
```

Print error messages if alignment isn't achieved

and make a dummy sequence of Ns

```
if isempty(f_start_pos)==1
    fprintf('%s\n',...
        'WARNING: Foward sequence could not be aligned to the template')
    for i=1:length(template)
        f_cropped=strcat('N',f_cropped);
    end

end

if isempty(r_end_pos)==1
    fprintf('%s\n',...
        'WARNING: Reverse sequence could not be aligned to the template')
    for i=1:length(template)
        r_cropped=strcat('N',r_cropped);
    end
end
```

Overlap

```
% Old code which used to be used for overlap checking
%overlap=-length(template)+concenus_f+concenus_r;

overlap=compare_sequence(f_cropped,r_cropped);
unmatched_start=sum((overlap-template_midpoint)<0);

if numel(overlap)==0
    overlap=length(template);
elseif numel(overlap)==1
    overlap=length(template)-1;
else
    overlap=overlap(unmatched_start+1)-overlap(unmatched_start);
end

overlap=overlap/length(template); % to give percentage overlap
```

checks whether the forward and reverse sequences are overlapping

```
overlap_strictness=0.2; % used to be 0.2

%if and(concenus_f>template_midpoint,concenus_r>template_midpoint)
if overlap>overlap_strictness
fprintf('%s\n','Forward and reverse sequences overlap')
else fprintf('%s\n',...
        'WARNING: Forward and reverse sequences DO NOT overlap')
end
```

splice together the consensus sequence from gene sequencing

```
if isempty(r_end_pos)==1
    fprintf('%s\n%',...
        'The whole forward sequence will be matched to the template')
    test=f_cropped;
elseif isempty(f_start_pos)==1
    fprintf('%s\n%',...
        'The whole reverse sequence will be matched to the template')
    test=r_cropped;
else
    test=strcat(forward(f_start_pos:f_start_pos+...
        floor(template_midpoint)),r_inverted(r_end_pos-ceil...
        (template_midpoint)+2:r_end_pos));
end
```

Outputs

```
tolerance=20;

miscalls=[strfind(test,'N') strfind(test,'M')];
miscalls=sort(miscalls);

if isempty(miscalls)==1
    fprintf('%s\n','Miscalls: None')
elseif length(miscalls)<tolerance
    fprintf('%s','Miscalls: ')
    fprintf('%d\t',miscalls)
    fprintf('\n')
else
    fprintf('%s','Miscalls: WARNING: There are over ')
    fprintf('%d\n',tolerance)
    fprintf('%s','Miscalls: ')
    fprintf('%d\t',miscalls)
    fprintf('\n')
end

% sprintf('%s',miscalls)
% sprintf('%i',miscalls)

% compare sequence from sequencing with reference sequence
mismatches=compare_sequence(template,test);
% sprintf('%i',mismatches)

if isempty(mismatches)==1
    fprintf('%s\n','Mismatches: None')
elseif length(mismatches)<tolerance
    fprintf('%s','Mismatches: ')
    fprintf('%d\t',mismatches)
    fprintf('\n')
else
```

```
    fprintf('%s','Mismatches: WARNING: There are over ')
    fprintf('%d\n',tolerance)
    fprintf('%s','Mismatches: ')
    fprintf('%d\t',mismatches)
    fprintf('\n')
end

% blank=zeros(4,4,4);
% template_protein=dna2pro(template,1,blank);
% test_protein=dna2pro(test,1,blank);

% protein_concensus=compare_sequence(template_protein,test_protein)

warning on
```

A.2 import_sequence.m

```
% given a file of text format (.txt, .seq, .fasta etc.), this function
% imports the text from the file into the MATLAB environment

function sequence=import_sequence(file_name)

% Open relavant file
fin = fopen(file_name); % input file

data_line=0;
data_string={};

while data_line~=-1

    data_line = fgets(fin);

    % Delete the heading from a .fasta sequence
    if strfind(data_line,'>')
    elseif data_line==-1
    else
        data_string=strcat(char(data_string),data_line);
    end

end

% change variable type to char
data_string=char(data_string);

% delete any spaces in the data
sequence=strrep(data_string,' ','');
sequence=CAPS_Sequence(sequence);
sequence=delete_numbers(sequence);

fclose(fin);
```

A.3 gene_compliment.m

```
% This script is an add-on to Gene_Processing.m. Gene_Processing.m is to
% be run as a pre-script. Running this function gives the complimentary
% base pairs of a DNA sequence (A replaced with T etc.)

function compliment=gene_compliment(gene)

gene_compliment=gene;

gene_compliment=strrep(gene_compliment,'A','t');
gene_compliment=strrep(gene_compliment,'T','a');
gene_compliment=strrep(gene_compliment,'C','g');
gene_compliment=strrep(gene_compliment,'G','c');

gene_compliment=strrep(gene_compliment,'t','T');
gene_compliment=strrep(gene_compliment,'a','A');
gene_compliment=strrep(gene_compliment,'g','G');
gene_compliment=strrep(gene_compliment,'c','C');

compliment=gene_compliment;
```

A.4 invert_sequence.m

```
% This function turns any sequence back-to-front so that the last
% letter is now the first

function inverted=invert_sequence(sequence)

% pre-allocation of an "empty" char array
inverted=cell(1,length(sequence));
inverted(:,:)={'%'};
inverted=char(inverted)';


for i=1:length(sequence)
    inverted(i)=sequence(length(sequence)-(i-1));
end
```

A.5 compare_sequence.m

```
% This function compares two strings of equal length, letter by letter
% and spitting out all the positions that a mis-match is found

function non_match=compare_sequence(ref_seq,test_seq)

warning off

%test_seq=strcat(f_cropped,'N');
%ref_seq=template;

test_seq=char(test_seq);
```

```
ref_seq=char(ref_seq);
test_seq=strrep(test_seq,' ','');
ref_seq=strrep(ref_seq,' ','');
```

ensure that the two sequences are the same length

```
length_ref=length(ref_seq);
length_test=length(test_seq);

% output error message if sequence lengths are not the same
if length_ref~=length_test
    fprintf('%s\n',...
        'WARNING: Length of sequences being compared don''t match')
end

% if not, assume that their starts are matched, and truncate the longer one
if length_ref>length_test
    ref_seq(end-abs(length_ref-length_test)+1:end)=[];
elseif length_ref<length_test
    test_seq(end-abs(length_ref-length_test)+1:end)=[];
end

% Detailed comparison: residue by residue analysis
comparison=(ref_seq==test_seq);
non_match=find(comparison==0);

warning on
```

A.6 dna2pro.m

```
% Despite the name of this function, due to the code being similar for
% the codon usage profile task, this function is written into the same
% function as additional inputs and outputs. i.e. this function is
% mainly for translating DNA to protein sequences but can also be used
% for generating code usage profiles and looking for rare codon usage.

% If you just want to translate DNA to protein sequence, input
% codon_usage as zeros(4,4,4) and have only one output

function [prot_seq,codon_profile]=dna2pro(DNA_seq,frame,codon_usage)

amino_acid=genetic_code;
DNA_seq_copy=DNA_seq;

bases={'T','C','A','G'};
bases_ref={'1','2','3','4'};

% pre-allocation
prot_seq='';
codon_profile=zeros(1,floor(length(DNA_seq)/3));
counter=1;
```

```
% adjust for the reading frame
if frame==2
    DNA_seq_copy(1)=[];
elseif frame==3
    DNA_seq_copy(1:2)=[];
else
end

while length(DNA_seq_copy)>2

    % take out first 3 bases as the codon being worked on
    codon=DNA_seq_copy(1:3);

    % remove any arithmetic operators from codon
    codon=strrep(codon,'-','X');
    codon=strrep(codon,'+','X');

    % replace the base letters with numbers for positioning in
    % translation matrix
    for i=1:4
        codon=strrep(codon,bases(i),bases_ref(i));
    end
    codon=char(codon);

    % find corresonding amino acid
    if isnan(str2double(codon))==1
        residue='-';
    else
        residue=amino_acid(str2double(codon(1)),str2double(codon(2)),...
            str2double(codon(3)));
    end

    % write residue to protein sequence
    prot_seq=strcat(prot_seq,residue);

    if isnan(str2double(codon))==1
        codon_profile(counter)=NaN;
    else
    codon_profile(counter)=codon_usage(str2double(codon(1)),...
        str2double(codon(2)),str2double(codon(3)));
    end

    % step to the next codon
    counter=counter+1;
    DNA_seq_copy(1:3)=[];
end

prot_seq=char(prot_seq);

not_translated=length(DNA_seq_copy);
```

```
if not_translated~=0
    fprintf('Frame of reference: %d\n',frame)
    fprintf('%s\n','WARNING: Some bases not translated')
else
end
```

A.7 genetic_code.m

```
% This function generates the genetic code in the MATLAB environment
% for use with the other scripts

function [amino_acid,amino_acid_num,code]=genetic_code

% T=1, C=2, A=3, G=4
```

Fill in amino acid table

```
amino_acid=cell(4,4,4);

amino_acid(:,:,1)={'F' 'S' 'Y' 'C';'L' 'P' 'H' 'R';'I' 'T' 'N' 'S';...
    'V' 'A' 'D' 'G'};
amino_acid(:,:,2)={'F' 'S' 'Y' 'C';'L' 'P' 'H' 'R';'I' 'T' 'N' 'S';...
    'V' 'A' 'D' 'G'};
amino_acid(:,:,3)={'L' 'S' 'X' 'X';'L' 'P' 'Q' 'R';'I' 'T' 'K' 'R';...
    'V' 'A' 'E' 'G'};
amino_acid(:,:,4)={'L' 'S' 'X' 'W';'L' 'P' 'Q' 'R';'M' 'T' 'K' 'R';...
    'V' 'A' 'E' 'G'};
```

Convert table string table into number table

```
caps={'A','B','C','D','E','F','G','H','I','J','K','L','M','N','O',...
    'P','Q','R','S','T','U','V','W','X','Y','Z'};
% numbers=num2cell(1:1:26);
numbers={'1','2','3','4','5','6','7','8','9','10','11','12','13',...
    '14','15','16','17','18','19','20','21','22','23','24','25','26'};

amino_acid_num=amino_acid;
for i=1:length(caps)
    amino_acid_num=strrep(amino_acid_num,caps(i),numbers(i));
end

amino_acid_num=str2double(amino_acid_num);
```

Fill in genetic code table

```
code_sparse=cell(4,4,4,3);
bases={'T','C','A','G'};
% T=1, C=2, A=3, G=4

for i=1:4
    code_sparse(i,:,:,1)=bases(i);
```

```
end

for i=1:4
    code_sparse(:,i,:,2)=bases(i);
end

for i=1:4
    code_sparse(:,:,i,3)=bases(i);
end

code=cell(4,4,4);
for i=1:4
    for j=1:4
        for k=1:4
            code(i,j,k)=strcat(code_sparse(i),code_sparse(j),...
                code_sparse(k));
        end
    end
end
```

A.8 rare_codons.m

```
% this script finds the rare codons in a gene sequence when
% expressed in E. coli

function rare_position=rare_codons(gene_sequence)

if strfind(gene_sequence,'.');
    gene=import_sequence(gene_sequence);
else
    gene=gene_sequence;
end

gene=reformat_sequence(gene,3);

rare={'AUA','AGG','AGA','CUA','CCC','GGA','CGG'};
rare=strrep(rare,'U','T');
rare_position=[];

for i=1:length(rare)
    rare_position=[rare_position strfind(gene,char(rare(i)))];
end

rare_position=(rare_position+3)./4;
```

Appendix B

MATLAB Scripts for Kinetic Assay Data Processing

B.1 KPC0418_423_kinetics.m

```
% This script is the top-level script for processing the kinetic data
% from all the experiments in the preliminary MEK1 kinetic assay.
```

clean-up

```
clear all
close all
clc
```

PROCESS

```
% parameters
samples=27;
reaction_number={'0418','0419','0420','0421','0422','0423'};
processing_options='';
sm=44057;
peaks=[sm,sm+80,sm+160]; % mass values for peaks of interest
times=[0,5,10,20,30,45,60];
times=times/60;
times=[times 1.25:0.25:6];
%start=[1,0,0];rad
max_time=6;
formatting='short'; % options are either 'short' or 'full'
product_ref='A';

concs_factors=10;%,1];
enzyme_conc=0.5; % \micro M
substrate_species={'ERK1','ERK1-P','ERK1-PP'};

% figure formatting
legend_pos={'west','north','east','east','east','east'};
rate_points=[8 6 6 5 7 3];
```

```
% open text file for output of processed data
output_file='1KPC0418_423_initial_rates.txt';
fout = fopen(output_file,'wt');
```

pre-allocation

```
initial_rates=zeros(length(reaction_number),length(peaks));

for i=1:length(reaction_number)
    no_samples=length(dir(strcat('1KPC',char(reaction_number(i)),'P',...
        char(product_ref),'*','_ENT01',char(processing_options),'.txt')));

    raw_peak_ints_mat=zeros(no_samples,length(peaks),...
        length(product_ref));
    all_pcnt_ints_mat=zeros(no_samples,length(peaks),...
        length(product_ref));
    sample_refs=zeros(no_samples,1,length(product_ref));

    h=figure('Position',[0 0 1000 650],'PaperPositionMode','auto');
    [raw_peak_ints_mat(:,:,i),all_pcnt_ints_mat(:,:,i),...
        sample_refs(:,:,i)]=reaction_kinetics_longassay(samples,...
        char(reaction_number(i)),processing_options,peaks,times,...
        max_time,formatting,product_ref);
    title(strcat('Kinetics of 1KPC',char(reaction_number(i)),'P',...
        product_ref))
    legend('ERK1','ERK1-P','ERK1-PP','Location',...
        [0.54 0.13 0.0813 0.130])
    if strcmp(formatting,'full')==1
        print(h,'-dpng','-r300',strcat('1KPC',...
            char(reaction_number(i)),'P',product_ref,'_kinetics.png'))
    end
```

Do a polyfit of the initial rates of consumption

```
    for j=1:length(peaks)
        linefit=polyfit(times(sample_refs(1:rate_points(i),:,i))',...
            concs_factors*all_pcnt_ints_mat(1:rate_points(i),j,i),1);
        initial_rates(i,j)=linefit(1);
    end
```

write the data used for the straight line fit to file

```
    % write name of experiment
    fprintf(fout,'%s',strcat('1KPC',char(reaction_number(i))));
    fprintf(fout,'\n');

    % write data
    for k=1:rate_points(i)
        fprintf(fout,'%3.4f\t',times(sample_refs(k,:,i))');
        for j=1:size(initial_rates,2)
            fprintf(fout,'%3.4f\t',concs_factors*...
```

```
                all_pcnt_ints_mat(k,j,i));
        end
        fprintf(fout,'\n');
    end

    fprintf(fout,'\n');
```

replot only the timecourse data as a separate graph

```
    h=figure('Position',[400 400 425 275],'PaperPositionMode','auto');
    plot(times(sample_refs(:,:,i)),all_pcnt_ints_mat(:,:,i),...
        '+');%,'-',times,noise,'--')
    hold on
    plot(times(sample_refs(:,:,i)),all_pcnt_ints_mat(:,:,i));
    %title(strcat('Kinetics of 1KPC',reaction_number(i),...
    %    processing_options))
    xlabel('Time/h')
    ylabel('Fractional conversion')
    axis([0 max_time 0 1])
    hold off
    %legend('ERK1','ERK1-P','ERK1-PP','location',char(legend_pos(i)))
    print(h,'-dpng','-r300',strcat('1KPC',char(reaction_number(i)),...
        'P',product_ref,'_kinetics_S.png'))

end
```

Output the initial rates as a text file

```
fprintf(fout,'%s','Initital Rates');
fprintf(fout,'\n');

for i=1:size(initial_rates,1)
    for j=1:size(initial_rates,2)
        fprintf(fout,'%3.4f\t',initial_rates(i,j));
    end
    fprintf(fout,'\n');
end

fclose(fout);
```

B.2 reaction_kinetics_longassay.m

```
% This function is for the final processing and display of the MS peak
% intensity data

function [raw_peak_intensities,all_percent_intensities,sample_refs]=...
    reaction_kinetics_longassay(samples,reaction_number,...
    processing_options,peaks,times,max_time,formatting,product_ref)

% where:
% - "samples" is the number of samples
% - "reaction_number" is the experiment number
```

```
% - "processing_options" is the options used in MassLynx to process the
% data
% - "peaks" are all the expected masses of the peaks
% - "start" are the quanities with which the masses are at t=0 ("1" for
% startng material, "0" for everything else)
% - "times" are the times at which the samples are taken
% - "max_time" is the time for the tim axis of the kinetics plot

[all_data,masses,sample_refs,baseline,noise]=...
    import_all_lists_longassay(samples,reaction_number,...
    processing_options,formatting,product_ref);

% get all the peak intensity data for all the peaks

raw_peak_intensities=zeros(length(sample_refs),length(peaks));

% given all the peak values, strip out all the intensities
for j=1:length(sample_refs)
    for i=1:length(peaks)
        raw_peak_intensities(j,i)=get_peak_intensity(masses,peaks(i),...
            all_data(:,j));
    end
end

% make the baseline values into a matrix suitable for transforming the
% matrix of peak intensity values
matrix_baseline=zeros(length(sample_refs),length(peaks));
for i=1:length(peaks)
    matrix_baseline(:,i)=baseline;
end

% correct the raw intensity values for the baseline
corrected_peak_intensities=raw_peak_intensities-matrix_baseline;

% calculate percentage intensity for each spectrum
peak_totals=sum(corrected_peak_intensities,2); % sum all intensity
                                               %for a given spectrum
matrix_peak_totals=zeros(length(sample_refs),length(peaks));
for i=1:length(peaks)
    matrix_peak_totals(:,i)=peak_totals;
end

all_percent_intensities=corrected_peak_intensities./matrix_peak_totals;

% convert all the peaks to string for making the legend
peaks_str=cell(1,length(peaks));
for i=1:length(peaks)
    peaks_str(i)={num2str(peaks(i))};
end

% determine the number of subplots required
```

```
dimensions=split_subplots(length(sample_refs));

if isempty(processing_options)==0
    processing_options=strcat('_',processing_options);
end

hold on
if strcmp(formatting,'full')==1
    subplot(dimensions(1),dimensions(2),dimensions(1)*dimensions(2))
end
plot(times(sample_refs),vertcat(all_percent_intensities),...
    '+');%,'-',times,noise,'--')
plot(times(sample_refs),vertcat(all_percent_intensities));
title(strcat('Kinetics of 1KPC',reaction_number,processing_options))
xlabel('time/h')
ylabel('percentage conversion')
axis([0 max_time 0 1])
legend(peaks_str)
hold off
```

B.3 import_all_lists_longassay.m

```
% This function is for importing the raw MS data, which is arranged in
% list format for biological LC-MS assay experiments

function [all_data,masses,sample_refs,baseline,...
    noise]=import_all_lists_longassay(samples,reaction_number,...
    processing_options,formatting,product_ref)

% where:
% - "samples" is the largest sample code being considered
% - "reaction_number" is the number in the experiment code as string

if isempty(processing_options)==1;
else
    processing_options=strcat('_',processing_options);
end
```

find all the files associated with the given reaction number

```
if isempty(product_ref)==1
    prod_type='@';
else
    prod_type='P';
end

sample_refs=dir(strcat('1KPC',char(reaction_number),prod_type,...
    char(product_ref),'*','_ENT01',char(processing_options),'.txt'));

% get all the names into an array, then only get the indices
sample_refs={sample_refs.name};
sample_refs=strrep(sample_refs,strcat('1KPC',...
```

```
    char(reaction_number),prod_type,char(product_ref)),'');
sample_refs=strrep(sample_refs,strcat('_ENT01',...
    char(processing_options),'.txt'),'');
sample_refs=sort(sample_refs);
sample_refs=str2num(char(sample_refs));
sample_refs=sample_refs(sample_refs<=samples);
```

convert the file indices back to string format with the 2 digit format

```
% import the first dataset to work out the size of matrix required
data=importdata(char(strcat('1KPC',reaction_number,prod_type,product_ref,...
    sprintf('%02d',sample_refs(1)),'_ENT01',processing_options,'.txt')));

% the if statement is needed to reduce the size of a high res dataset
if size(data,1)>3200
    all_data=zeros(size(data,1)/2,length(sample_refs));
    masses=data(:,1);
    masses=masses(1:2:end,:);
else
    all_data=zeros(size(data,1),length(sample_refs));
    masses=data(:,1);
end


% due to having a two digit format
for i=1:length(sample_refs)
    data=importdata(char(strcat('1KPC',reaction_number,prod_type,...
        product_ref,sprintf('%02d',sample_refs(i)),'_ENT01',...
        processing_options,'.txt')));
    if size(data,1)>3200
            % this if statement is needed to match mat dimensions
        data=data(1:2:end,:);
    end
    all_data(:,i)=data(:,2);
end

baseline=zeros(size(all_data,2),1);
noise=zeros(size(all_data,2),1);
baseline(:,1)=median(all_data);
noise(:,1)=iqr(all_data);

% determine the number of subplots required
dimensions=split_subplots(length(sample_refs));

if strcmp(formatting,'full')==1
    for i=1:length(sample_refs)
        subplot(dimensions(1),dimensions(2),i)
        hold on
        title(char(strcat('1KPC',reaction_number,prod_type,product_ref,...
            sprintf('%02d',sample_refs(i)),processing_options)))
        xlabel('mass')
```

```
        ylabel('intensity')
        plot(masses,all_data(:,i))
        plot(masses,baseline(i))
        plot(masses,baseline(i)+noise(i))
        plot(masses,baseline(i)-noise(i))
        hold off
        xlim([min(masses) max(masses)])
    end
elseif strcmp(formatting,'short')==1
end
```

B.4 get_peak_intensity.m

```
% This function is for extracting the height of peaks in an MS spectrum,
% given the mass search range

function peak_intensity=get_peak_intensity(masses,peak_mass,...
    spectrum_data,window)

% given the mass, these are the values of x that contain the given peak
if exist('window','var')==0
    window=8;
end
peak=(masses>peak_mass-window).*(masses<peak_mass+window);

peak_intensity=max(peak.*spectrum_data);
```

B.5 split_subplots.m

```
% This function is for determining the format of subplots, given the number
% of subplots, e.g. when n number of MS spectra need to be plotted

% given the number of subplots, determine the subplot format
function dimensions=split_subplots(samples)

% where:
% samples = the number of subplots required

% possible_dimensions=[1 2; 2 2; 2 2; 2 3; 2 3; 3 3; 3 3; 3 3;...
%     3 4; 3 4;3 4; 4 4; 4 4; 4 4; 4 4; 4 5; 4 5; 4 5; 4 5;...
%     5 5; 5 5; 5 5; 5 5; 5 5];

% a better way would be to calculate the requirement, rather than
% to use a LUT as is currently the case
dimensions=[ceil((samples+1)./ceil(sqrt(samples+1))) ...
    ceil(sqrt(samples+1))];
```

Appendix C

MATLAB Scripts for Analysis and Plotting of Circular Dichroism Spectra

C.1 KPC0199CD01_shell.m

paramters

```
% aa_import_cd_data parameters
sample_code='1KPC0199CD01';
repeats=5;

% ba_import_cd_blank_parameters
blank_sample_code='1KPC0154CD08';
blank_repeats=5;

% molardeg_values parameters
mass_concn=0.36; % from Bradford assay plate. Code: 1KPC0206PA1
FW=41373;
path_length=0.1;

% ea_plot_cd_data_3d parameters
angle_units='moldeg';

% fa_plot_temp_melt parameters
wavelength=225;
```

analysis

```
close all
% import blank data and CD data. Plot the raw data
[all_blank,temperatures_blank,averaged_blank,smoothed_blank]...
    =ba_import_cd_blank(blank_sample_code,blank_repeats);
[all_data,wavelengths,temperatures,averaged_data,smoothed_data,...
    residuals]=ca_cd_data_processing(sample_code,repeats);

% change the units of the data to molar elipticity
moldeg_averaged_blank=mdeg2moldeg(averaged_blank,mass_concn,FW,...
```

```
    path_length);
moldeg_smoothed_blank=mdeg2moldeg(smoothed_blank,mass_concn,FW,...
    path_length);
moldeg_averaged_data=mdeg2moldeg(averaged_data,mass_concn,FW,...
    path_length);
moldeg_smoothed_data=mdeg2moldeg(smoothed_data,mass_concn,FW,...
    path_length);

% plot CD data
[moldeg_averaged_data,moldeg_smoothed_data]=...
    da_remove_blank(moldeg_averaged_data,moldeg_smoothed_data,...
    moldeg_averaged_blank,moldeg_smoothed_blank);
ea_plot_cd_data_3d(moldeg_smoothed_data,wavelengths,sample_code,...
    angle_units,rendition);

% plot the temperature melt (for reference)
fa_plot_temp_melt(wavelength,wavelengths,temperatures,...
    moldeg_smoothed_data,sample_code,angle_units);
melt_leg = 'return';
fa_plot_temp_melt(wavelength,wavelengths,temperatures,...
    moldeg_smoothed_data,sample_code,angle_units,melt_leg);
```

C.2 ba_import_cd_blank.m

```
function [all_blank,temperatures_blank,averaged_blank,smoothed_blank]...
    =ba_import_cd_blank(blank_sample_code,blank_repeats)

% Where:
% - "blank_sample_code" is in format 1KPC0XXXCDYY
% - "repeats" is an integer number with the number of repeat files

[all_blank,wavelengths_blank,temperatures_blank]=...
    aa_import_cd_data(blank_sample_code,blank_repeats);

[averaged_blank,smoothed_blank,residuals]=ab_process_cd_data(all_blank);

figure
for i=1:17
    subplot(4,5,i)
    plot(wavelengths_blank,residuals(:,i))
end
suplabel('Residuals Between Raw and Smoothed Data for Blank','t')

figure
plot(wavelengths_blank,averaged_blank)
title('Averaged Blank Data')
xlabel('wavelength/nm')
ylabel('angle of rotation/m^o')

figure
plot(wavelengths_blank,smoothed_blank)
```

```
title('Smoothed Blank Data')
xlabel('wavelength/nm')
ylabel('angle of rotation/m^o')
```

C.3 ca_cd_data_processing.m

```
function [all_data,wavelengths,temperatures,averaged_data,...
    smoothed_data,residuals]=ca_cd_data_processing(sample_code,repeats)

[all_data,wavelengths,temperatures]=aa_import_cd_data(sample_code,...
    repeats);

[averaged_data,smoothed_data,residuals]=ab_process_cd_data(all_data);

figure
for i=1:length(temperatures)
    subplot(4,5,i)
    plot(wavelengths,residuals(:,i))
end
suplabel(strcat('Residuals Between Raw and Smoothed Data for :',...
    sample_code),'t')

figure
plot(wavelengths,averaged_data)
title(strrep(strcat('Averaged Data for@',sample_code),'@',' '))
xlabel('wavelength/nm')
ylabel('angle of rotation/m^o')

figure
plot(wavelengths,smoothed_data)
title(strrep(strcat('Smoothed Data for@',sample_code),'@',' '))
xlabel('wavelength/nm')
ylabel('angle of rotation/m^o')
```

C.4 mdeg2moldeg.m

```
function moldeg_values=mdeg2moldeg(mdeg_values,mass_concn,FW,path_length)

% Where:

% "mdeg_values" are circular dichromism values in mdeg
% "mass_concn" is the protein concentration in mg/mL (equivilant to g/L)
% "FW" is the formula weight of the protein
% "path_length" is the thickness of the sample in cm

molar_concn=mass_concn/FW;
%molar_concn=8.70133E-06; % test value from experiment 1KPC0199PA1

moldeg_values=(mdeg_values)/(molar_concn*path_length*10);
% The factor of 10 converts path_length to m and the final value from
% mmolar to molardeg
```

```
% [theta] = 100xtheta/(Cxl) where C is molar_concn and l is path length
% equation from http://www.photophysics.com/tutorials/
% circular-dichroism-cd-spectroscopy/7-cd-units-conversions
```

C.5 da_remove_blank.m

```
function [averaged_data,smoothed_data]=da_remove_blank(averaged_data,...
    smoothed_data,averaged_blank,smoothed_blank)

averaged_data=averaged_data-averaged_blank;
smoothed_data=smoothed_data-smoothed_blank;
```

C.6 ea_plot_cd_data_3d.m

```
function ea_plot_cd_data_3d(smoothed_data,wavelengths,sample_code,...
    angle_units,rendition)

% these truncations are for the figures for the paper
%wavelengths(length(wavelengths)-29:length(wavelengths))=[];
%smoothed_data(length(smoothed_data)-29:length(smoothed_data),:)=[];

% parameters
f_size=16;

mock_temp_data=20:10:180;

h=figure('Position',[200 200 560 420],'PaperPositionMode','auto');
% set(h,'position',[0 0 525 375])

% line rendition of CD temperature melt
if strcmp(rendition,'lines')==1
    for i=1:size(smoothed_data,2)
        temp_data=smoothed_data(:,i);
        temps=ones(size(smoothed_data,1),1)*mock_temp_data(i);
        hold on
        plot3(wavelengths,temps,temp_data)
        hold off
    end

    xlabel('Wavelength/nm')

    set(gca, 'YTick', 1:5, 'YTickLabel',{'-10' '40' '90' '40' '-10'});
    set(gca, 'YTickMode','auto');
    ylabel('Temperature/ ^oC')

    %view([1,-1,1])
    %view([34,23])
    view([19,23])
    option='_L';

    % surface rendition of CD temperature melt
```

```
elseif strcmp(rendition,'surf')==1

    hsur=surf(mock_temp_data,wavelengths,smoothed_data);
    set(gca,'ydir','reverse')

    ylabel('Wavelength/nm','fontsize',f_size)

    % NB: There are two different viewing angles
    view([-66,23])
    %k=xlabel('temperature/^oC','fontsize',f_size);
    set(gca, 'xtick',0:50:200)
    set(gca, 'XTickLabel',{'-10' '40' '90' '40' '-10'});
    %set(gca, 'XTickMode','auto');

    %pos=get(k,'position');
    %set(k,'position',[pos(1) pos(2) pos(3)])

    set(hsur,'facealpha',0.7,'edgecolor',[0.3,0.3,0.3])
    % view([-66,9])
            % this rendition will make it more obvious for the return
    option='_S';

    set(gca,'zlim',[-3.5e6,1.5e6])

end

% set common figure parameters for both renditions
grid on
if strcmp(angle_units,'mdeg')==1
    zlabel('Angle of rotation/m^o','fontsize',f_size)
elseif strcmp(angle_units,'moldeg')==1
    zlabel('Molar ellipticity/M^-^1m^-^1','fontsize',f_size)
end
%title(strcat('CD Temperature Melt/Reanneal (sample:',sample_code,')'))
%view([1,-0,1]) % for the melting curve

set(gca,'fontsize',f_size,'fontname','helvetica')

% NB: Try to make this linked to the axis
annotation('textbox',[0.86 0.16 0.1 0.05],'string','Return',...
    'fontname','helvetica','fontsize',f_size,'EdgeColor','none');
annotation('textbox',[0.78 0.1 0.1 0.05],'string','Out',...
    'fontname','helvetica','fontsize',f_size,'EdgeColor','none');
annotation('textbox',[0.72 0.04 0.2 0.05],'string','Temperature/^oC',...
    'fontname','helvetica','fontsize',f_size,'EdgeColor','none');
% ea_plot_cd_data_3d(smoothed_data,wavelengths,sample_code,...
%     angle_units,rendition)

% print to png
print(h,'-dpng','-r300',strcat(sample_code,'_ALL',option,'.png'))
```

C.7 fa_plot_temp_melt.m

```
function fa_plot_temp_melt(wavelength,wavelengths,temperatures,...
    smoothed_data,sample_code,angle_units,melt_leg)
% 225nm is the minimum at 10degC

if exist('melt_leg','var')==0
    melt_leg='out';
end

if strcmp(melt_leg,'out')==1
    temp_range=1:9;
elseif strcmp(melt_leg,'return')==1
    temp_range=9:17;
end

ind=find(wavelengths==wavelength);

h=figure('Position',[400 400 425 275],'PaperPositionMode','auto');
plot(temperatures(temp_range),smoothed_data(ind,temp_range))
hold on
plot(temperatures(temp_range),smoothed_data(ind,temp_range),'+')
hold off
xlabel('Temperature/^oC')
xlim([0 90])
if strcmp(melt_leg,'return')==1
    set (gca,'Xdir','reverse')
end
ylim([-3.5e6 0.5e6])
if strcmp(angle_units,'mdeg')==1
    ylabel('Angle of rotation/m^o')
elseif strcmp(angle_units,'moldeg')==1
    ylabel('Molar ellipticity/M^-^1m^-^1')
end
% title(strrep(strcat('Melting curve from experiment:',sample_code,...
%     ':@',num2str(wavelength),'nm'),':',' '))

print(h,'-dpng','-r300',strcat(sample_code,'_MELT',...
    num2str(wavelength),'_',upper(melt_leg(1)),'.png'));
```

C.8 aa_import_cd_data.m

```
function [all_data,wavelengths,temperatures]=aa_import_cd_data...
    (sample_code,repeats)
% where "all_data" is in mdeg_values

% import the first file to get the right size for the matrix
data=importdata(strcat(sample_code,'_R1_C.csv'));

% feed in all the data into a 3D matrix
all_data=zeros([size(data),repeats]);
```

```
for i=1:repeats
    all_data(:,:,i)=importdata(strcat(sample_code,'_R',...
        num2str(i),'_C.csv'));
end

% extract the column and row headings
wavelengths=all_data(:,1,1);
wavelengths(1)=[];

temperatures=all_data(1,:,1);
            % runs could be repeats or different temperatures
temperatures(1)=[];

% delete the column and row headings
all_data(:,1,:)=[];
all_data(1,:,:)=[];
```

C.9 ab_process_cd_data.m

```
function [averaged_data,smoothed_data,residuals]=...
    ab_process_cd_data(all_data)

% average all the repeats
averaged_data=mean(all_data,3);

smoothed_data=sgolayfilt(averaged_data,1,5,[],1);

% check that all the residuals are OK
residuals=smoothed_data-averaged_data;
```

C.10 simple_cd_analysis.m

clean-up

```
clear all
close all
clc
```

VARIABLES

```
% filenames
input_files={'1KPC0407CD02_ALL_C.csv','1KPC0408CD01_ALL_C.csv',...
    '1KPC0409CD01_ALL_C.csv'};
blank_file='1KPC0400_blank0001_C.csv';

% parameters
mass_concn=[0.51,0.50,0.46];
FW=[45676,45642,45756];
path_length=0.1;

%formatting style
```

```
format_color={'r','g','b'};
```

Open a figure for plotting stuff

```
h=figure('Position',[400 400 425 275],'PaperPositionMode','auto');

for i=1:length(input_files)
% start analysis
cd_data=importdata(char(input_files(i)));

wavelengths=cd_data(:,1);
spectra=cd_data(:,2:end);

% take mean of the spectra
spectrum=mean(spectra,2);

% import blank
cd_data_blank=importdata(blank_file);

wavelengths_blank=cd_data_blank(:,1);
spectra_blank=cd_data_blank(:,2:end);

% take mean of the spectra
spectrum_blank=mean(spectra_blank,2);

% make the dimensions of the blank and the spectrum match
if wavelengths(1)==wavelengths_blank(1)
    if length(spectrum_blank)>length(spectrum)
        spectrum_blank(length(spectrum)+1:end)=[];
    elseif length(spectrum)>length(spectrum_blank)
        spectrum(length(spectrum_blank)+1:end)=[];
    end
end

% subtract the blank from the spectrum and smooth
spectrum=spectrum-spectrum_blank;
[~,spectrum,~]=ab_process_cd_data(spectrum);

% convert to molar ellipticity
spectrum=mdeg2moldeg(spectrum,mass_concn(i),FW(i),path_length);

% plot the spectrum
hold on
plot(wavelengths,spectrum,char(format_color(i)))
hold off

end

% label the plot
xlabel('Wavelength/nm')
ylabel('Molar ellipticity/M^-^1m^-^1')
```

```
legend('MEK1-Cys222','MEK1-Dha222','MEK1-pCys222')
title('CD spectra of all chemically modified variants of MEK1-Cys222')

% print spectra
print(h,'-dpng','-r300','1KPC0407-409CD01_ALL.png')
```

Appendix D

MATLAB Scripts for Differential Scanning Fluorimetry Data Processing

D.1 KPC0172_temp_melt_anal.m

```
% clean-up
close all
clear all
clc
printing='off';

% import and reformat the data
data=importdata('1KPC0172_20130118_111110.csv');
data=data.data;
temps=data(:,1);
data(:,1)=[];

% pre-allocation
melt_temps=zeros(6,8);
min_melts=zeros(6,8);
max_melts=zeros(6,8);
window=3;

% plot data into the same format as the original assay plate
figure
for i=1:size(data,2)-12
    % select subplot
    subplot(8,6,i)

    % select only sigmoidal section of curve
    data_temp=data(:,i);
    if window<length(data_temp)
        data_temp((length(data_temp)-10):end)=[];
        [~,min_ind]=min(data_temp);
        [~,max_ind]=max(data_temp);
        min_melts(i)=temps(min_ind);
        max_melts(i)=temps(max_ind);
```

```
        data_temp=data_temp(min_ind-window:max_ind+window);
        temps_temp=temps(min_ind-window:max_ind+window);
    elseif window>length(data_temp)
        temps_temp=temps;
    end

    % plot and fit the data
    plot(temps_temp,data_temp)
    [~,melt_temps(i)]=fit_4PL(temps_temp,data_temp);

    if strcmp(printing,'on')==1
        % plot for printing
        h=figure('Position',[400 400 425 275],'PaperPositionMode','auto');
        plot(temps_temp,data_temp,'+')
        [~,melt_temps(i)]=fit_4PL(temps_temp,data_temp);
        xlabel('Temperature/^oC')
        ylabel('Fluorescence')
        axis([30 60 0 2])
        hold off
        print(h,'-dpng','-r300',strcat('1KPC0172_MELT_',sprintf('%02d',i)))
        close 2
    end

end

% repair the data orientation to match that of plate
melt_temps=real(melt_temps)';
min_melts=real(min_melts)';
max_melts=real(max_melts)';

means_melt_temps=[mean(melt_temps(:,1:3),2) mean(melt_temps(:,4:6),2)];
means_min_melts=[mean(min_melts(:,1:3),2) mean(min_melts(:,4:6),2)];
means_max_melts=[mean(max_melts(:,1:3),2) mean(max_melts(:,4:6),2)];

% output the melting temperatures
output_file='1KPC0172_melt_summary.txt';
fout = fopen(output_file,'wt');

for i=1:size(means_melt_temps,1)
    fprintf(fout, '%4.1f\t%4.1f\t', means_melt_temps(i,1:2));
    fprintf(fout, '\n');
end
fprintf(fout, '\n');

for i=1:size(means_min_melts,1)
    fprintf(fout, '%4.1f\t%4.1f\t', means_min_melts(i,1:2));
    fprintf(fout, '\n');
end
fprintf(fout, '\n');

for i=1:size(means_max_melts,1)
```

```
    fprintf(fout, '%4.1f\t%4.1f\t', means_max_melts(i,1:2));
    fprintf(fout, '\n');
end
fprintf(fout, '\n');

fclose(fout);
%open(output_file)
```

D.2 fit_4PL.m

```
function [hillCoeff,ec50]=fit_4PL(dose,Response)
```

fit the 4-PL curve

```
%hill equation 4 parameter sigmoid
sigmoid=@(beta,T)beta(1)+(beta(2)-beta(1))./(1+exp((beta(3)-T)./beta(4)));

%description of paramters:
% beta(1)=min
% beta(2)=max
% beta(3)=Tm
% beta(4)=a (Hill slope)

%calculate some rough guesses for initial parameters
minResponse=min(Response);
maxResponse=max(Response);
midResponse=mean(dose);
mindose=min(dose);
maxdose=max(dose);

%fit the curve and compute the values
[coeffs,r,J]=nlinfit(dose,Response,sigmoid,[minResponse maxResponse...
    midResponse 1]);

ec50=coeffs(3);
hillCoeff=coeffs(4);

%plot the fitted sigmoid
xpoints=linspace(mindose,maxdose,1000);
hold on
plot(xpoints,sigmoid(coeffs,xpoints),'b')
hold off
```

Appendix E

MATLAB Scripts for Gel Densitometry of SDS-PAGE Gels

E.1 KPC0328_332_quant.m

```
% This script is used for densitometry of and SDS-PAGE gel
```

clean-up

```
clear all
close all
clc
```

import image and make it negative

```
im1=imread('1KPC0328-332G1_0003.tif');
im1=imcomplement(im1);
figure
imshow(im1)
```

get all top left co-ordinates of the spots

```
spot_x=[364 591 707 824 1060 1287 1403 1522];
spot_y=[820 831 828 831 822  817  813  811];
spot_h=937-831;
spot_w=928-824;
```

excise all spots and get intensities

```
spot_int=zeros(1,length(spot_x));
figure

for i=1:length(spot_x)
    spot_temp=im1(spot_y(i):spot_y(i)+spot_h,spot_x(i):spot_x(i)+spot_w);
    spot_int(i)=sum(sum(spot_temp));
    subplot(2,4,i)
    imshow(spot_temp)
end
```

take ratios of the spot intensities

```
ratio=spot_int(1:4)./spot_int(5:8);
```

Appendix F

AutoIt Scripts for Controlling MassLynx

F.1 kpc_ms_openallchromatogram_r.au3

```
; This script is for opening of the chromatograms of all spectra in the
; auto_input folder in MassLynx

; Need to:
; - be able to control the keyboard. It keeps switching to German for
; some reason
; - make better input for the variables

Func WinActivateWait($arg)
EndFunc   ;==>WinActivateWait

; Variables
; $MSFile = ""
$OutputFolder = "V:\Lab8\Chooi\temp_for_transfer\processed_lcms\" ;Victoria
$HomeFolder01 = "R:\Phin\" ; BGD-MassStore
$CalFile = "2014-06-28_myo_cal_01_a.scl" ; "2014-01-24_myo_cal_02_k.scl"
$param_file_name = "experiment_parameters_r.txt"
$AnalRun = "01" ; This number is the suffix for filenames, e.g. ENT01
$CombineRange = "433:461" ; "264:352"
$MassRange01 = "700"
$MassRange02 = "2000"
$MassRange03 = "42000:45000"
$MassRange04 = "20000:50000"
$display_style = 'overlay'

; Get all file names in auto_input
#include <File.au3>
$AllMSFiles = _FileListToArray($HomeFolder01 & "Auto_input\")

; Pre-cleanup
WinClose("[CLASS:SpecWClass]")
WinClose("[CLASS:ChroWClass]")

; Open MassLynx
```

```
Run("mlynx4.exe")
WinWait("[CLASS:MassLynxUIWClass; REGEXPTITLE:MassLynx -*]", "")
Sleep(500)
WinActivate("[CLASS:MassLynxUIWClass]", "")

; Find what the HomeDir is
$HomeDir = StringSplit($HomeFolder01, ":")
$HomeDir = $HomeDir[1]

; Navigate to folder with the spectra, open first file
Sleep(2000)
Send("{ALT}fd")
WinWait("Data Browser")
WinActivate("Data Browser")
ControlFocus("Data Browser", "", 1213)
Send($HomeDir)
WinWait("Data Browser", $HomeDir & ":\")
ControlFocus("Data Browser", "", 1207)
Send("ph")
WinWait("Data Browser", "\Phin\")
Send("{RIGHT}au")
WinWait("Data Browser", "\Phin\Auto_input\")
Send("{RIGHT}")
ControlFocus("Data Browser", "", 1212)
Send("{UP}")
ControlFocus("Data Browser", "", 107)
Send("{+}")
ControlFocus("Data Browser", "", 109)
Send("{-}")
Send("{ENTER}")
WinWaitClose("Data Browser")
WinWait("[CLASS:ChroWClass]")

; Make sure that new chromatograms are added to the view, and
; not just replaced
WinActivate("[CLASS:ChroWClass]")
Send("{Alt}wn")
WinWait("New Chromatogram")
WinActivate("New Chromatogram")
Send("a{Enter}")
WinWaitClose("New Chromatogram")
Sleep(200)

If $display_style = 'overlay' Then
; make all opened chromatograms overlap each other
WinActivate("[CLASS:ChroWClass]")
Send("{Alt}dv")
WinWait("Chromatogram Display View")
WinActivate("Chromatogram Display View")
Sleep(200)
Send("{Tab 8}{+}{Enter}")
```

```
WinWaitClose("Chromatogram Display View")
Sleep(200)
EndIf

WinClose("[CLASS:ChroWClass]")

For $i = 1 To UBound($AllMSFiles)

; Open spectrum as a mock
WinActivate("[CLASS:MassLynxUIWClass]", "")
Sleep(1000)
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(300)
Send($AllMSFiles[$i])
ControlFocus("Data Browser", "", 107)
Send("{-}")
ControlFocus("Data Browser", "", 109)
Send("{+}")
Send("{Enter}")
WinWait("[CLASS:SpecWClass]")

; Save experiment parameters to get SampleCode
WinActivate("[CLASS:MassLynxUIWClass]", "")
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(400)
Send("{Tab 3}e")
WinWait("Experimental Record")
WinActivate("Experimental Record")
Sleep(400)
Send("{ALTDOWN}f{ALTUP}s")
WinWait("Save As")
WinActivate("Save As")
Sleep(1200)
Send($HomeFolder01 & "auto_output\" & $param_file_name)
Send("{ENTER}")
WinWaitClose("Save As")
WinClose("Experimental Record")
Send("{ENTER}")
WinWait("[CLASS:SpecWClass]")

; Extract SampleCode from text file
$SampleCode = FileReadLine($HomeFolder01 & "auto_output\"
     & $param_file_name, 17)
$SampleCode = StringSplit($SampleCode, ":")
$SampleCode = $SampleCode[2]
$SampleCode = StringReplace($SampleCode, " ", "")
```

```
; ignore wash and myo files
If StringInStr($SampleCode, "wash") <> 0
Or StringInStr($SampleCode, "ipa") <> 0
Or StringInStr($SampleCode, "myo") <> 0 Then
Else

; Open chromatogram
WinActivate("[CLASS:MassLynxUIWClass]", "")
Sleep(1000)
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(300)
Send($AllMSFiles[$i])

ControlFocus("Data Browser", "", 107)
Send("{+}")
ControlFocus("Data Browser", "", 109)
Send("{-}")
Send("{Enter}")
WinWait("[CLASS:ChroWClass]")
EndIf

WinClose("[CLASS:SpecWClass]")
FileDelete($HomeFolder01 & "auto_output\" & $param_file_name)

Next

; WinClose("[CLASS:MassLynxUIWClass]")
```

F.2 kpc_ms_manualwrite_win7_r.au3

```
; This script is for the automated processing of mass spectra
; using MassLynx

; Variables
; $MSFile = ""
$OutputFolder = "V:\Lab8\Chooi\temp_for_transfer\processed_lcms\"
$HomeFolder01 = "R:\Phin\" ; BGD-MassStore
$CalFile = "2014-07-05_myo_cal_09_r.scl"
$param_file_name = "experiment_parameters_r.txt"
$AnalRun = "01"
$CombineRange = "431:457";432:459"
$MassRange01 = "700"
$MassRange02 = "2000"
$MassRange03 = "42000:45000"
$MassRange04 = "20000:50000"
$analysis_type = "reanal" ; "fresh", "fix" or "reanal"

; Get all file names in auto_input
```

```
#include <File.au3>
$AllMSFiles = _FileListToArray($HomeFolder01 & "auto_input\")

; Pre-cleanup
WinClose("[CLASS:SpecWClass]")
WinClose("[CLASS:ChroWClass]")

; Open MassLynx
Run("mlynx4.exe")
WinWait("[CLASS:MassLynxUIWClass; REGEXPTITLE:MassLynx -*]", "")
Sleep(500)
WinActivate("[CLASS:MassLynxUIWClass]", "")

; Find what the HomeDir is
$HomeDir = StringSplit($HomeFolder01, ":")
$HomeDir = $HomeDir[1]

; Navigate to folder with the spectra, open first file
Sleep(500)
Send("{ALT}fd")
WinWait("Data Browser")
WinActivate("Data Browser")
ControlFocus("Data Browser", "", 1213)
Send($HomeDir)
WinWait("Data Browser", $HomeDir & ":\")
ControlFocus("Data Browser", "", 1207)
Send("ph")
WinWait("Data Browser", "\Phin\")
Send("{RIGHT}au")
WinWait("Data Browser", "\Phin\Auto_input\")
Send("{RIGHT}")
ControlFocus("Data Browser", "", 1212)
Send("{UP}")
If $analysis_type = "fresh" Then
ControlFocus("Data Browser", "", 107)
Send("{+}")
ControlFocus("Data Browser", "", 109)
Send("{-}")
ElseIf $analysis_type = "reanal" Or $analysis_type = "fix" Then
ControlFocus("Data Browser", "", 107)
Send("{-}")
ControlFocus("Data Browser", "", 109)
Send("{+}")
EndIf
Send("{ENTER}")
WinWaitClose("Data Browser")
If $analysis_type = "fresh" Then
WinWait("[CLASS:ChroWClass]")
ElseIf $analysis_type = "reanal" Or $analysis_type = "fix" Then
WinWait("[CLASS:SpecWClass]")
EndIf
```

```
If $analysis_type = "fresh" Then
; Expand chromatogram to full range
WinActivate("[CLASS:ChroWClass]")
Send("{ALT}dnd")
WinWait("Default Chromatogram Range")
Send("l{ENTER}")
WinWaitClose("Default Chromatogram Range")

; Set the display of the chromatogram to how I like it
WinActivate("[CLASS:ChroWClass]")
Send("{Alt}dv")
WinWait("Chromatogram Display View")
WinActivate("Chromatogram Display View")
Sleep(400)
Send("h")
WinWait("Header Editor ( ChrHeader )")
WinActivate("Header Editor ( ChrHeader )")
Sleep(400)
Send("c{Tab 2}r{Tab}ssss{Tab}a{Tab}{Enter}")
WinWaitClose("Header Editor ( ChrHeader )")
Send("o{Enter}")
WinWaitClose("Chromatogram Display View")

; make a mock spectrum so that spectrum settings can change
WinActivate("[CLASS:ChroWClass]")
Send("{ALT}pc")
WinWait("Combine Spectrum")
Send("100:101" & "{ENTER}")
WinWaitClose("Combine Spectrum")
EndIf

; Make sure all the peak annotation spectrum display settings
; are the ones I want
WinWait("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
Sleep(800)
Send("{Alt}dp")
WinWait("Spectrum Peak Annotation")
Send("0{Tab 4}{-}{Tab 8}{UP}{Tab}10{Enter}")
WinWaitClose("Spectrum Peak Annotation")

; Set spectrum display settings
WinWait("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
Sleep(800)
Send("{Alt}dv")
WinWait("Spectrum Display")
Send("0{Tab 10}{+}{Tab}{-}{Tab}{+}{Tab 6}{Enter}")
WinWait("Header Editor ( SpeHeader )")
Send("c{Enter}")
```

```
WinWaitClose("Header Editor ( SpeHeader )")
WinActivate("Spectrum Display")
Send("+{TAB 2}{Enter}")
WinWaitClose("Spectrum Display")

; Make the spectrum range settings the ones I want
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
Send("{Alt}dnd")
WinWait("Default Spectrum Range")
WinActivate("Default Spectrum Range")
Send("alu{-}{ENTER}")
WinWaitClose("Default Spectrum Range")

; close chromatogram and spectrum windows
WinClose("[CLASS:SpecWClass]")
WinClose("[CLASS:ChroWClass]")

For $i = 1 To UBound($AllMSFiles)

; Open first chromatogram
WinActivate("[CLASS:MassLynxUIWClass]", "")
Sleep(1000)
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(300)
Send($AllMSFiles[$i] & "{ENTER}")
If $analysis_type = "fresh" Then
WinWait("[CLASS:ChroWClass]")
ElseIf $analysis_type = "reanal" Or $analysis_type = "fix" Then
WinWait("[CLASS:SpecWClass]")
WinClose("[CLASS:SpecWClass]")
EndIf

If $analysis_type = "reanal" Or $analysis_type = "fix" Then
; Open combined spectrum
WinActivate("[CLASS:MassLynxUIWClass]", "")
Sleep(1000)
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(300)
Send("{Tab 3}")
Sleep(200)
Send("i")
Sleep(300)
WinWait("History Selector")
WinActivate("History Selector")
Sleep(300)
; if reanalysis, delete all previous deconvoluted spectra
If $analysis_type = "reanal" Then
```

```
Send("{DOWN}{DOWN}{TAB}")
For $repeats = 1 To 20
Send("d")
Sleep(200)
Next
Send("{SHIFTDOWN}{TAB}{SHIFTUP}")
ControlFocus("History Selector", "", 80)
EndIf ;;;;
If $analysis_type = "fix" Then
Send("{DOWN}{ENTER}")
ElseIf $analysis_type = "reanal" Then
Send("{DOWN}{DOWN}{ENTER}")
EndIf
WinWaitClose("History Selector")
WinActivate("Data Browser")
Send("{ENTER}")
WinWaitClose("Data Browser")
WinWait("[CLASS:SpecWClass]")
EndIf

; cleanup old experiment parameter file
FileDelete($HomeFolder01 & "auto_output\" & $param_file_name)

; Save experiment parameters to get SampleCode
WinActivate("[CLASS:MassLynxUIWClass]", "")
Send("{ALTDOWN}f{ALTUP}d")
WinWait("Data Browser")
WinActivate("Data Browser")
Sleep(400)
Send("{Tab 3}e")
WinWait("Experimental Record")
WinActivate("Experimental Record")
Sleep(400)
Send("{ALTDOWN}f{ALTUP}s")
WinWait("Save As")
WinActivate("Save As")
Sleep(1200)
Send($HomeFolder01 & "auto_output\" & $param_file_name)
Send("{ENTER}")
WinWaitClose("Save As")
WinClose("Experimental Record")
WinClose("Data Browser")

; Extract SampleCode from text file
$SampleCode = FileReadLine($HomeFolder01 & "auto_output\"
     & $param_file_name, 17)
$SampleCode = StringSplit($SampleCode, ":")
$SampleCode = $SampleCode[2]
$SampleCode = StringReplace($SampleCode, " ", "")

; ignore wash and myo files
```

```
If StringInStr($SampleCode, "wash") <> 0
     Or StringInStr($SampleCode, "ipa") <> 0
     Or StringInStr($SampleCode, "myo") <> 0 Then
Else

; delete old files associated with the SampleCode
If $analysis_type = "fresh" Then
FileDelete($OutputFolder & $SampleCode & "*"
 & $AnalRun & "*")
EndIf

If $analysis_type = "fresh" Then
; Save image of chromatogram
WinActivate("[CLASS:ChroWClass]")
Sleep(400)
WinMove("[CLASS:ChroWClass]", "", 50, 50, 680, 291)
Sleep(200)
Send("{ALTDOWN}e{ALTUP}c")
Sleep(200)
Run("mspaint.exe")
WinWait("[CLASS:MSPaintApp]")
Sleep(300)
Send("{ALTDOWN}f{ALTUP}e")
WinWait("Image Properties")
Send("1{TAB}1{ENTER}")
WinWaitClose("Image Properties")
Send("{ALTDOWN}h{ALTUP}vp")
Send("{ALTDOWN}f{ALTUP}a")
WinWait("Save As")
Sleep(1500)
Send($OutputFolder & $SampleCode & "_CHROM" & $AnalRun
              & "{TAB}p{ENTER}")
WinWaitClose("Save As")
WinClose("[CLASS:MSPaintApp]")

; Combine MS if a CombineRange is given
If $CombineRange = "" Then
WinActivate("[CLASS:ChroWClass]")
Else
WinActivate("[CLASS:ChroWClass]")
Send("{ALT}pc")
WinWait("Combine Spectrum")
Send($CombineRange & "{ENTER}")
WinWaitClose("Combine Spectrum")
EndIf

;;;; switches keyboard settings here

; Save spectrum
WinWait("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
```

```
Send("{ALTDOWN}f{ALTUP}v")
WinWait("[TITLE:Spectrum Save; CLASS:#32770]", "Combine")
WinActivate("[TITLE:Spectrum Save]")
Send("{ENTER}")
WinWaitClose("Spectrum Save")
Sleep(200)

; Calibrate Spectrum if a calibration file name is given
If $CalFile = "" Then
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
Else
WinActivate("[CLASS:SpecWClass; REGEXPTITLE:Spectrum -*]")
Send("{ALTDOWN}t{ALTUP}a")
WinWait("Apply Calibration")
Send("{Tab 3}")
Send($CalFile)
Send("{TAB}")
Send($HomeFolder01)
Send("{ENTER}")
WinWaitClose("Apply Calibration")
WinWait("[TITLE:Calibrate; CLASS:#32770]")
Send("{ENTER}")
WinWaitClose("Calibrate")
EndIf
EndIf

; Change the mass range of the spectrum
WinActivate("[CLASS:SpecWClass]")
Send("{ALTDOWN}d{ALTUP}nf")
WinWait("Display Range", "", 1)
Send($MassRange01)
Send("{TAB}")
Send($MassRange02)
Send("{ENTER}")

; Deconvolute the spectrum
WinActivate("[CLASS:SpecWClass]")
Send("{ALTDOWN}p{ALTUP}1")
WinWait("MaxEnt", "", 2)
Send($MassRange03)
Send("{TAB}1{Tab 2}0.55{TAB}33{TAB}33{ENTER}")
WinWaitClose("MaxEnt", "", 2)
WinWait("MaxEnt", "MaxEnt may experience dif")
WinActivate("MaxEnt", "MaxEnt may experience dif")
Send("{ENTER}")
WinWaitClose("MaxEnt", "MaxEnt may experience dif")
WinActivate("[CLASS:MaxEntClass;
         REGEXPTITLE:Electrospray Maximum Entropy *]")
WinWait("[CLASS:MaxEntClass; REGEXPTITLE:
         Electrospray Maximum Entropy *]", "Converged")
Send("{ALT}{Space}x")
```

```
Sleep(200)
MouseClick("Left", 45, 35, 1, 0)
WinWaitClose("[CLASS:MaxEntClass;
         REGEXPTITLE:Electrospray Maximum Entropy *]", "Converged")
WinWait("Spectrum")
WinActivate("Spectrum")
Send("{ALTDOWN}w{ALTUP}s")
Sleep(800)

; Close unwaited child windows
WinActivate("[CLASS:SpecWClass]")
Send("{ALT}w2{ALT}{-}c")
Send("{ALT}w2{ALT}{-}c")

; Deconvolute the spectrum again, since the first time usually fails
WinActivate("[CLASS:SpecWClass]")
Send("{ALTDOWN}p{ALTUP}1")
WinWait("MaxEnt", "", 2)
Send("{Enter}")
WinWaitClose("MaxEnt", "", 2)
WinWait("MaxEnt", "MaxEnt may experience dif")
WinActivate("MaxEnt", "MaxEnt may experience dif")
Send("{ENTER}")
WinWaitClose("MaxEnt", "MaxEnt may experience dif")
WinActivate("[CLASS:MaxEntClass; REGEXPTITLE:
Electrospray Maximum Entropy *]")
WinWait("[CLASS:MaxEntClass; REGEXPTITLE:
Electrospray Maximum Entropy *]",
         "Converged")
Send("{ALT}{Space}x")
Sleep(200)
MouseClick("Left", 45, 35, 1, 0)
WinWaitClose("[CLASS:MaxEntClass; REGEXPTITLE:
         Electrospray Maximum Entropy *]", "Converged")
WinWait("Spectrum")
WinActivate("Spectrum")
Send("{ALTDOWN}w{ALTUP}s")
Sleep(800)

If $analysis_type = "fresh" Then
; Save Image of IS
WinActivate("[CLASS:SpecWClass]")
WinMove("[CLASS:SpecWClass]", "", 50, 50, 700, 779)
Send("{ALTDOWN}w{ALTUP}s")
Sleep(200)
Send("{ALTDOWN}w{ALTUP}1")
Sleep(200)
Send("{ALT}ec")
Run("mspaint.exe")
WinWait("[CLASS:MSPaintApp]")
Send("{ALTDOWN}h{ALTUP}vp")
```

```
Send("{ALT}fa")
WinWait("Save As")
Sleep(1000)
Send($OutputFolder & $SampleCode & "_IS" &
$AnalRun & "{TAB}p{ENTER}")
WinWaitClose("Save As")
WinClose("[CLASS:MSPaintApp]")
EndIf

; Save image of ENT
WinActivate("[CLASS:SpecWClass]")
Send("{ALT}w2")
Sleep(200)
Send("{ALT}ec")
Run("mspaint.exe")
WinWait("[CLASS:MSPaintApp]")
WinActivate("[CLASS:MSPaintApp]")
Send("{ALTDOWN}h{ALTUP}vp")
Send("{ALT}fa")
WinWait("Save As")
Sleep(1000)
Send($OutputFolder & $SampleCode & "_ENT" & $AnalRun
& "{TAB}p{ENTER}")
If $analysis_type = "reanal" Or $analysis_type = "fix" Then
WinWait("Confirm Save As")
WinActivate("Confirm Save As")
Send("y")
WinWaitClose("Confirm Save As")
EndIf
WinWaitClose("Save As")
WinClose("[CLASS:MSPaintApp]")

; Save list of ENT
WinActivate("[CLASS:SpecWClass]")
Send("{ALT}el")
Run("notepad.exe")
Sleep(200)
WinWait("Untitled - Notepad", "")
Send("{ALT}ep")
Send("{ALT}fa")
WinWait("Save As")
Sleep(1000)
Send($OutputFolder & $SampleCode & "_ENT" & $AnalRun & "{ENTER}")
If $analysis_type = "reanal" Or $analysis_type = "fix" Then
WinWait("Confirm Save As")
WinActivate("Confirm Save As")
Send("y")
WinWaitClose("Confirm Save As")
EndIf
WinWaitClose("Save As")
WinClose("[CLASS:Notepad]")
```

```
If $analysis_type = "reanal" Or $analysis_type = "fresh" Then
; Close unwanted child windows
WinActivate("[CLASS:SpecWClass]")
Send("{ALT}w2{ALT}{-}c")
Send("{ALT}w2{ALT}{-}c")

; Deconvolute in wider mass range
WinActivate("[CLASS:SpecWClass]")
Send("{ALT}w1")
Send("{ALT}p1")
WinWait("MaxEnt", "", 2)
Send($MassRange04)
Send("{TAB}1{Tab 2}0.55{TAB}33{TAB}33{ENTER}")
WinWaitClose("MaxEnt", "", 2)
WinWait("MaxEnt", "MaxEnt may experience dif")
WinActivate("MaxEnt", "MaxEnt may experience dif")
Send("{ENTER}")
WinWaitClose("MaxEnt", "MaxEnt may experience dif")
WinActivate("[CLASS:MaxEntClass; REGEXPTITLE:
             Electrospray Maximum Entropy *]")
WinWait("[CLASS:MaxEntClass; REGEXPTITLE:
             Electrospray Maximum Entropy *]", "Converged")
Send("{ALT}{Space}x")
Sleep(200)
MouseClick("Left", 45, 35, 1, 0)
WinWaitClose("[CLASS:MaxEntClass; REGEXPTITLE:
             Electrospray Maximum Entropy *]", "Converged")
WinWait("Spectrum")
WinActivate("Spectrum")
Send("{ALT}ws")
Sleep(800)

; Save Image of ENT_W
WinActivate("[CLASS:SpecWClass]")
WinMove("[CLASS:SpecWClass]", "", 50, 50, 700, 779)
Send("{ALT}ws")
Sleep(200)
Send("{ALT}w2")
Sleep(200)
Send("{ALT}ec")
Run("mspaint.exe")
WinWait("[CLASS:MSPaintApp]")
Send("{ALTDOWN}h{ALTUP}vp")
Send("{ALT}fa")
WinWait("Save As")
Sleep(1000)
Send($OutputFolder & $SampleCode & "_ENT" &
$AnalRun & "_W{TAB}p{ENTER}")
If $analysis_type = "reanal" Or $analysis_type = "fix" Then
WinWait("Confirm Save As")
```

```
WinActivate("Confirm Save As")
Send("y")
WinWaitClose("Confirm Save As")
EndIf
WinWaitClose("Save As")
WinClose("[CLASS:MSPaintApp]")

EndIf
EndIf

; Clean up windows
WinClose("[CLASS:SpecWClass]")
WinClose("[CLASS:ChroWClass]")
FileDelete($HomeFolder01 & "auto_output\" & $param_file_name)

Next

; WinClose("[CLASS:MassLynxUIWClass]")
```