Exploring the mechanisms of Fibroblast Growth Factor Receptor activation and regulation using structural tools

Brendan Farrell

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology
&
The Astbury Centre of Structural Molecular Biology

September 2019

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors of this work has been explicitly indicated below. The candidate confirms that appropriate credit had been given within the thesis where reference has been made to the work of others.

Components of this thesis contain work from jointly authored publications.

Chapter 1 is partly based on the following review article:

Farrell B, Breeze A. (2018) Structure, activation and dysregulation of fibroblast growth factor receptor kinases: perspectives for clinical targeting. *Biochem Soc Trans*. **46**, 1753-1770.

Adaptations of Figures 1 to 4, Tables 1 to 3, and written content are presented in this thesis. The candidate devised and wrote the manuscript and prepared all figures with guidance and contributions from Breeze A.

Chapter 2 contains methodology adapted from the following research article:

Bunney TD, Inglis AJ, Sanfelice D, **Farrell B**, Kerr CJ, Thompson GS, Masson GR, Thiyagarajan N, Svergun DI, Williams RL, Breeze AL, Katan M. (2018) Disease variants of FGFR3 reveal molecular basis for the recognition and new roles for Cdc37 in Hsp90 chaperone system. *Structure*. **26**, 446-458.

M.K., T.D.B., D.I.S., R.L.W., and A.L.B. conceived the study. T.D.B. performed all protein expression, purification, biophysical and biochemical assays, and analysis. A.J.I., G.R.M., and R.L.W. performed and analysed all HDX-MS experiments. G.S.T. and B.F. performed all NMR experiments. D.S., G.S.T., B.F., and A.L.B. analysed all NMR data. N.T., D.S., C.J.K., and D.I.S. performed and analysed all SAXS experiments. M.K., A.L.B., and T.D.B. wrote the manuscript with contributions from all the authors.

The first draft of the NMR methodology of this article was written by the candidate with guidance by Breeze AL and is reproduced with adaptations in this chapter. The candidate also made comments and contributions to editing of the manuscript during its development. The methodology for expression and purification of Hsp90, Cdc37 and FGFR3 kinase domain, presented in this publication, were written by Bunney TD. These were adapted during expression of each of these components in this thesis. Additional elements of this publication were used elsewhere in the thesis, as detailed below.

Chapter 4 includes data and analyses from the following jointly-authored publications:

Sanfelice D, Koss H, Bunney TD, Thompson GS, **Farrell B**, Katan M, Breeze AL. (2018) NMR backbone assignments of the tyrosine kinase domain of human fibroblast growth factor receptor 3 in apo state and in complex with inhibitor PD173074. *Biomol NMR Assign*. **12**, 231-235.

NMR backbone assignments of FGFR3 kinase domains in *apo* and PD173074-bound states, and NMR spectra of PD173074-bound FGFR3 kinase domains were used in this thesis. These spectra were acquired by Koss H and Thompson GS. The candidate acquired spectra of selectively (un)labelled kinase samples to aid with backbone resonance assignments. These assignments were determined by Sanfelice D, Koss H and Thompson GS (BMRB entries 27082 and 27083). In this thesis, the candidate independently analysed these inhibitor-bound data and additional unpublished spectra showing the titration of PD173074 into kinase-Cdc37 complexes that were acquired by the candidate. All protein samples used in this chapter were prepared by Bunney TD.

Bunney TD, Inglis AJ, Sanfelice D, **Farrell B**, Kerr CJ, Thompson GS, Masson GR, Thiyagarajan N, Svergun DI, Williams RL, Breeze AL, Katan M. (2018) Disease variants of FGFR3 reveal molecular basis for the recognition and new roles for Cdc37 in Hsp90 chaperone system. *Structure*. **26**, 446-458.

NMR analyses of the binary complex of Hsp90 co-chaperone Cdc37 and FGFR3 kinase domain and an adaptation of Figure 5 (prepared by the candidate) are included in this thesis. In the chapter discussion, HDX-MS and SAXS data and their analyses are also presented (by Inglis AJ, Masson GR, Williams RL, Kerr CJ, Thiyagarajan N and Svergun DI). Bunney TD prepared all proteins used in this chapter, and some NMR data (the titrations of FGFR3 kinase domains with Cdc37) were acquired by Thompson GST. The candidate acquired all other spectra (partly with guidance from Thompson GST). Analysis of NMR data was completed by the candidate, Sanfelice D, and Thompson GST with contributions from Breeze AL. Specifically, the assignment of complex-dependent peaks was made through triple-resonance experiment analyses by Sanfelice D, by prediction of random coil chemical shifts by Thompson GST, and through analysis of selectively unlabelled spectra by the candidate. Chemical shift perturbation analysis of Cdc37 binding was completed by candidate alone, and of I538F substitution in the kinase domain by the candidate and Sanfelice D. All other analyses were completed by the candidate.

Acknowledgements

First of all, I would like to thank my supervisors Alex Breeze, Adrian Goldman and Stephen Muench for their support and guidance over the past few years, and for helping me to navigate the many challenges that this project has thrown at me. I would also like to thank the Wellcome Trust for their funding, without which none of this would have been possible.

Over the course of my PhD, I have been fortunate to work and collaborate with a great team of people. A big thank you to Gary Thompson for teaching me how to get going with NMR and laying the groundwork for the kinase chaperoning project, to Tom Bunney for providing me with everything from proteins for NMR, to plasmids and protocols, to Domenico Sanfelice for support in data analysis, and to Matilda Katan for the many thoughtful discussions. I also thank all those who have helped acquire and analyse data during my PhD and for letting me loose on their equipment, including but not limited to: Rachel George for acquiring and processing data all of my many (and often troublesome) mass spectrometry samples, Nikesh Patel and Maren Thomsen for their help with SEC-MALS, Arnout Kalverda for support in using (and not breaking) the NMR spectrometers, and to all of the EM facility staff for their training and help.

This is easily one of the hardest things that I have done, but it has been made all that bit easier by the friends I've made along the way. Thanks to everyone for the laughs, the chats, for enabling procrastination and for keeping me (relatively) sane. A big shout-out to Emma and Rachel who have been there with me through it all. Even though I often seem unimpressed by your shenanigans, know that you have been the most 'wellcome' distraction from science. Cheers to Jamie (my TBST and snacks thief, but sugoi best buddy), danke to Jannik (my fellow pal with superior colour vision and pillar of the trio), to Alex (my meme provider and Twitter accomplice), Claudia (the best desk neighbour I could ask for who has been very patient with my growing messiness) and Craig (my cider drinking companion). Thanks also to Julie, Maren, Jess, Maria, Francis, Steve, Jack, and to all the other members both past and present of the Breeze, Goldman and Muench groups, and of Astbury and Miall level 6, between which there are far too many people to list.

Last but certainly not least, I need to thank my parents, my brothers and my family. I know you might not understand very much of what I've been getting up to for the past 4 years, but you've supported me through it all nonetheless. You can pretend it's a picture book if you like.

Abstract

Fibroblast growth factor receptors (FGFRs) are receptor tyrosine kinases (RTKs) that regulate developmental pathways in response to fibroblast growth factor binding. When dysregulated, FGFRs drive a number of oncogenic and developmental pathologies. While strides have been made in understanding RTK biology, many aspects are still poorly described. In this thesis, structural tools were used to explore the regulation processes of FGFRs and of kinases more generally, focusing on two key aspects: (a) activity regulation in the context of full-length FGFR receptors, and (b) the regulation of kinases by the cellular chaperone Hsp90.

Currently, no full-length structure of FGFR nor any RTK is available. To address this lack of information, the expression of full-length FGFR3 and oncogenic variants thereof was established in insect cells. Purification of the receptor was optimised for preliminary negative-stain EM analyses of the receptor in *apo* and ligand-complexed forms. Though high-resolution structures of FGFR3 could not be determined during this project, the foundations were laid for future efforts.

In exploring the chaperoning of kinases by Hsp90, using NMR spectroscopy, it was found that the Hsp90 co-chaperone Cdc37, which recognises client kinases, induces a remodelling of the kinase N-lobe that results in an apparent local unfolding at the kinase N-terminus. Moreover, analyses of wild-type, I538F-substituted and inhibitor-bound FGFR3 kinases suggested that differences in the client strength of kinases relates to a network of residues within their N-lobes. To explore the process of kinase transfer to Hsp90, FGFR3-bound ternary complexes of Hsp90 were reconstituted and characterised by complementary structural and biochemical techniques. These analyses suggested that novel complexes which may describe kinase transfer to Hsp90 were captured. Building on the low-resolution maps obtained of these complexes, efforts were focused on sample optimisation in aim to obtain higher resolution information regarding kinase chaperoning using cryo-electron microscopy.

Table of Contents

Table of Tables

Table of Figures

Abbreviations

2D	2-dimensional
3D	3-dimensional
A-loop	Activation loop
ADP	Adenosine diphosphate
Aha1	Activator of Hsp90 ATPase protein 1
ATP	Adenosine triphosphate
ANZ3-8	Anzergent 3-8
AUC	Analytical ultracentrifugation
B_0	External magnetic field
BAIAP2L1	Brain-specific *a*ngiogenesis inhibitor 1-associated protein 2-like protein 1
BEVS	Baculovirus expression vector system
BIICs	Baculovirus infected insect cells
BSA	Bovine serum albumin
C-Cdc37	Carboxy-terminal domain of Cdc37
C12E8	Octaethylene glycol monodecyl ether
Cdk4	Cyclin dependent kinase 4
CFOS6	Cyclofos-6
CHS	Cholesteryl hemisuccinate
CnEm	Polyoxyethylene alkyl ether
CSP	Chemical shift perturbation
CTD	Carboxy-terminal domain of Hsp90
CTF	Contrast transfer function
CV	Column volumes
DDM	n-Dodecyl-β-D-maltoside
DIBMA	Diisobutylene-maleic acid
DM	n-Decyl-β-D-Maltoside
DMNG	Decyl maltose neopentyl glycol
DPA	Day following proliferation arrest
dRI	Change in refractive index
DTT	Dithiothreitol
ECD	Extracellular domain
EGFR	Epiderminal growth factor receptor
EM	Electron microscopy
EMDB	Electron microscopy data bank
FC12	Fos-choline-12
FGF	Fibroblast growth factor

FGFR	Fibroblast growth factor receptor
FID	Free-induction decay
FKBP51/2	51/2 kDa FK506-binding protein
FRET	Förster resonance energy transfer
FRS2α	FGFR substrate 2
FSC	Fourier shell correlatoin
FSEC	Fluorescence size exclusion chromatography
GDN	Glycol-dicosgenin
GFP	Green fluorescent protein
GR	Glucocorticoid receptor
GraFix	Gradient Fixation
Grb2	Growth factor receptor-bound protein 2
HDX	Hydrogen-deuterium exchange
HDX-MS	Hydrogen-deuterium exchange mass spectrometry
HOP	Hsp70-Hsp90 organising protein
HPN motif	His-Pro-Asn motif
HRD motif	His-Arg-Asp motif
Hsp70	Heat shock protein 70
Hsp90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactoside
IR	Insulin receptor
ITC	Isothermal titration calorimetry
JMD	Juxtamembrane domain
KD	Kinase domain
K_d	Dissociation binding constant
KLD mix	T4 Polynucleotide Kinase, T4 DNA Ligase and DpnI mix
LDAO	n-Dodecyl-N,N-Dimethylamine-N-Oxide
LMNG	Lauryl maltose neopentyl glycol
mAb	Monoclonal antibody
M_0	Net magnetisation
MAPK	Mitogen-activated protein kinase
M-Cdc37	Middle domain of Cdc37
MD	Middle domain of Hsp90
MEEVD motif	Met-Glu-Glu-Val-Asp motif of Hsp90
MOI	Multiplicity of infection
MS	Mass spectrometry

M_{xy}	Net transverse (x-y) magnetisation
M_z	Net longitudal (z) magnetisation
N-Cdc37	Amino-terminal domain of Cdc37
NMR	Nuclear magnetic resonance
NTD	Amino-terminal domain of Hsp90
OD_{600}	Optical density at 600 nm
P-loop	Nucleotide binding loop
P13K	Phosphoinositide 3-kinase
pAb	Polyclonal antibody
PA	Proliferation arrest
PCR	Polymerase chain reaction
PDB	Protein data bank
PDGFR	Platelet-derived growth factor receptor
PES	Polyethersulfone
P_i	Inorganic phosphate
PP5	Serine/threonine protein phosphatase 5
ppm	Parts per million
RRID	Research resource identifier/antibody registry identifier
PTB	Phosphotyrosine binding
RTK	Receptor tyrosine kinase
Salipro	Saposin-lipoprotein
SAXS	Small-angle X-ray scattering
SCF	Stem cell factor
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography with multi-angle light scattering
SMA	Styrene maleic acid
STAT	Signal transducer and activator of transcription protein
T_1	Longitudal (z) relaxation time
T_2	Transverse (x-y) relaxation time
TACC3	Transforming acidic coiled-coil containing protein 3
Tah1	TPR-containing protein associated with Hsp90
TBST	Tris-buffered saline containing 0.1% v/v Tween®-20
TMD	Transmembrane domain
TPR	Tetratricopeptide repeat
TROSY	Tranverse relaxation optimised spectroscopy
YFP	Yellow fluorescent protein
γ	gyromagnetic ratio of a nucleus

Chapter 1. Introduction

1.1. Fibroblast growth factor receptors

1.1.1. Receptor tyrosine kinases

Through their role in signal transduction pathways, protein kinases participate in a plethora of cellular processes, ranging from cell growth and proliferation to differentiation and survival (1). One class of kinases, the receptor tyrosine kinases (RTKs), span the plasma membranes of cells at their surface. In total, there are 58 known RTKs belonging to 20 subfamilies in humans, each responding to unique exogenous growth signals. Despite this difference, all RTKs are fundamentally very similar, communicating the binding of external factors to their tyrosine kinase domains that stimulate intracellular signalling through their *trans*-autophosphorylation (2).

The fibroblast growth factor receptor (FGFR) family of RTKs contains four members, FGFRs 1-4. Sharing between 56 to 71% sequence identity (3), these receptors are highly homologous and share the same structural architecture: namely, an extracellular domain (ECD) comprising three immunoglobulin-like domains (D1, D2 and D3), a single membrane-spanning helix, and an intracellular tyrosine kinase domain (KD) (Figure 1.1). Activation of receptors occurs via ligand-induction of a conformational state at the ECD that is communicated to the KDs, stimulating their catalytic kinase activity. Crucially, this effect is mediated through dimers of RTKs; models of receptor activation can be separated into those describing activation through ligand-induced receptor dimerisation or through allosteric modes on pre-formed receptor dimers (Figure 1.1A). FGFRs have roles in embryonic development, tissue homeostasis, and metabolism (4–6).

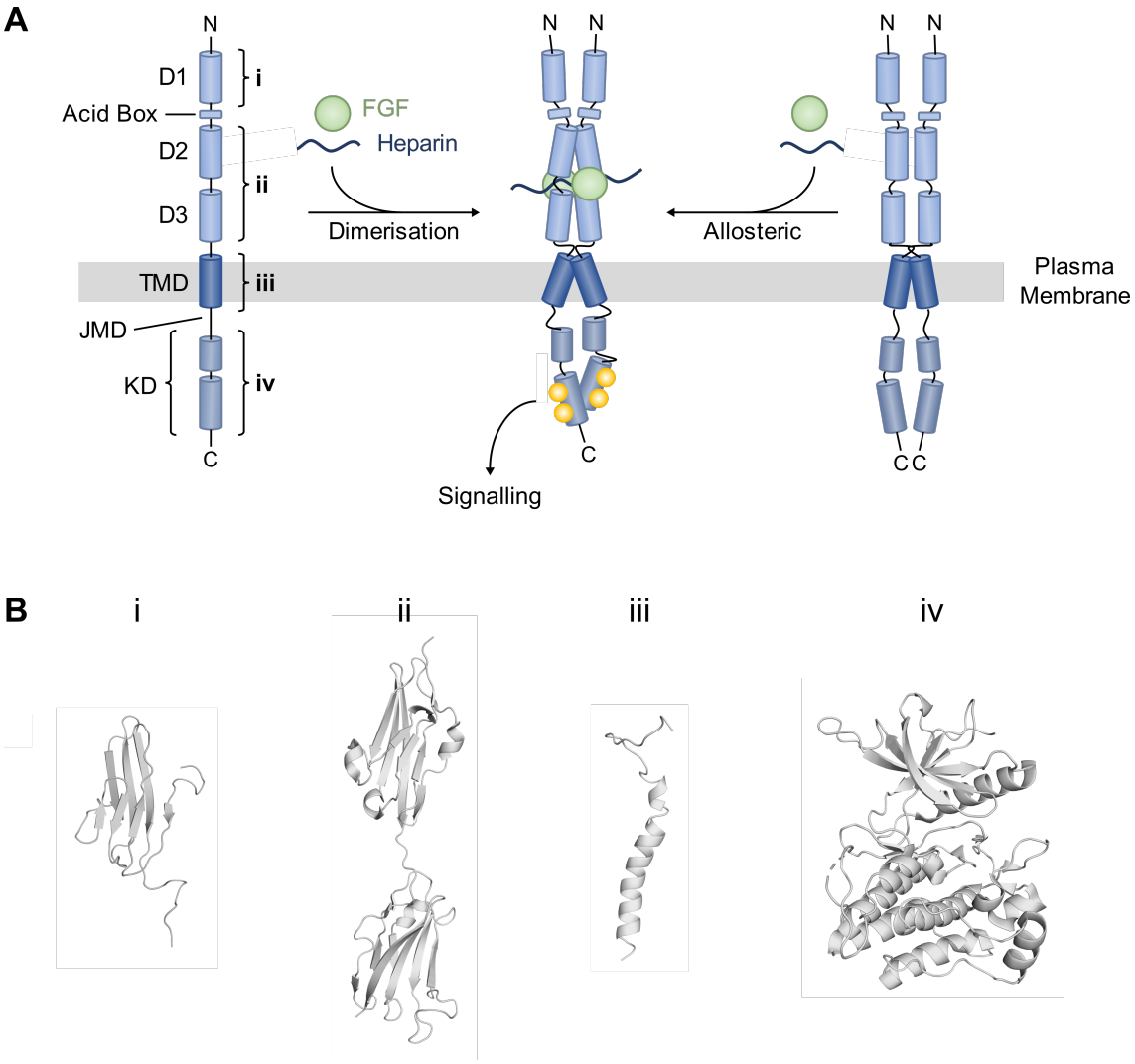


Figure 1.1. FGFR architecture and activation models.
(A) Schematic illustrating the architecture of fibroblast growth factor receptors (FGFRs) comprising an extracellular domain (ECD) containing domains D1 to D3 and the 'acid box', the transmembrane domain (TMD) and the kinase domain (KD). Preceding the kinase domain lies the juxtamembrane domain (JMD). During receptor activation, two copies of FGF ligand and heparin bind to the extracellular domain resulting in receptor activation and downstream signalling. Two models of receptor activation have been proposed: ligand-driven dimerisation and ligand-driven allosteric conformational changes. Labels i-iv correspond to the subpanels in panel B. (B) Structures of domain D1 (FGFR1, PDB entry 2CR3), domains D2-D3 (FGFR1, PDB entry 1FQ9), the transmembrane domain (FGFR3, PDB entry 2LZL), and the kinase domain (FGFR3, PDB entry 4K33). In each case, the N-terminus of each construct is at the top of the structure and the C-terminus at the bottom. Structures are not to scale. This figure is adapted from Farrell and Breeze (7).

1.1.2. Ligand-mediated activation of FGFRs

Despite their similarities, FGFR members have distinct roles in cell biology, partly attributed to their varied binding specificities to fibroblast growth factors (FGFs), 18 of which have been described in mammals (3,8). While some FGFs show binding redundancy among the FGF receptors (e.g. the universal binder FGF1 (acidic FGF)), others bind selectively to sole members of the family (9). FGFs can be subdivided into paracrine and endocrine ligands that differ in their mechanism of receptor activation. All FGFs share a beta-trefoil fold that presents a binding site for heparan sulfate on their surface, sequestering them close to the cell surface where receptor binding occurs (8,10). This effect is more pronounced for paracrine FGFs than for endocrine ones which exhibit lower binding affinity for heparan sulfate (8) and require additional Klotho co-receptors for FGFR activation (6,11–13). Generally, the action of paracrine FGFs will be described here.

While paracrine FGFs are able to bind independently to FGFRs in a 1:1 stoichiometry (14–20), heparan sulfate (or heparin) is necessary for receptor dimerisation and FGFR signalling *in vivo* (21–23). Crystal structures of FGFs bound to FGFR extracellular domains have been instrumental for our understanding of the ligand-induced activation mechanisms of FGFRs. In total, 15 structures of ligand-bound FGFR ECDs have been solved, in two of these cases with the co-factor heparin present also (24,25). In all structures, despite corresponding to different combinations of FGFR extracellular domains and FGFs (Table 1.1), the general binding mode of ligands is similar, bridging domains D2 and D3, as predicted from earlier biochemical experiments (26). In fact, when aligning all structures at the first of these two domains, they differ most in their positioning of domain D3 relative to D2, pivoting immediately after the latter (Figure 1.2). These differences in D3 arrangement have been proposed to have roles communicating ligand-specific activation signals across the membrane, as will be addressed later. These structures also explain how ligand binding specificities are achieved through alternative splicing of the amino-terminal part of domain D3, a process that results in ‘b’ and ‘c’ isoforms of FGFRs 1-3 that are differentially expressed in epithelial and mesenchymal tissues (8,17,26–29). This splicing event changes the binding interface for FGF ligands at the D3 domain, altering binding capabilities.

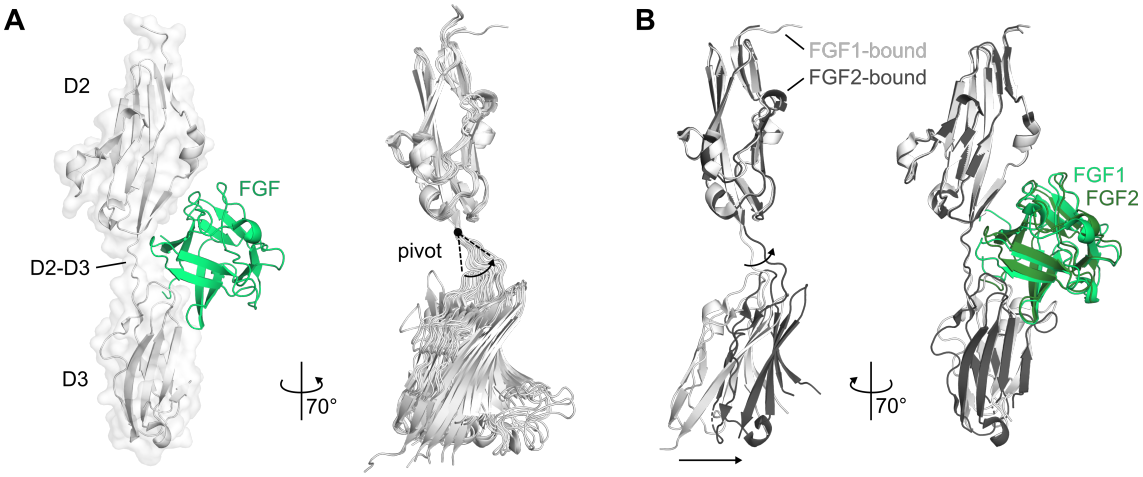


Figure 1.2. FGF ligand binding at the extracellular domain.
(A) FGFR1 extracellular domain bound to FGF1 (PDB entry 1EVT). The D2 and D3 domains of FGFR1 are shown in white with cartoon and surface representation, and FGF1 is shown in green with cartoon representation. Next to this, an alignment of the FGFR extracellular domains (D2 and D3) of 14 FGF-bound structures is shown in cartoon representation, aligned to the D2 domain, to illustrate pivoting at the D2-D3 linker (PDB entries 1EVT, 3OJV, 1CVS, 5W59, 3OJ2, 3OJM, 1NUN, 1DJS, 1EV2, 1IIL, 1II4, 2FDB, 1RY7 and 1FQ9). (B) An alignment of FGF1-bound and FGF2-bound FGFR2 extracellular domains, illustrating the pivoting between domains D2 and D3 induced by different ligands on the same receptor (PDB entries 1DJS and 1EV2).

Table 1.1. Ligand-bound extracellular complexes of FGFRs.

Receptor ECD	Ligand(s)	PDB entry	Reference
FGFR1	FGF1	1EVT	Plotnikov *et al*. (15)
FGFR1c	FGF1	3OJV	Beenken et al. (17)
FGFR1c	FGF2	1CVS	Plotnikov *et al*. (14)
FGFR1c	FGF9	5W59	Liu *et al.* (30)
FGFR2b with A172F	FGF1	3OJ2	Beenken *et al.* (17)
FGFR2b with P253R	FGF1	3OJM	Beenken *et al.* (17)
FGFR2b	FGF10	1NUN	Yeh et al. (31)
FGFR2c	FGF1	1DJS	Stauber *et al*. (16)
FGFR2c	FGF2	1EV2	Plotnikov *et al.* (15)
FGFR2c with P253R	FGF2	1IIL	Ibrahimi *et al.* (18)
FGFR2c with S252W	FGF2	1II4	Ibrahimi et al. (18)
FGFR2c	FGF8	2FDB	Olsen et al. (19)
FGFR3c	FGF1	1RY7	Olsen et al. (20)
FGFR1c	FGF2 & Heparin	1FQ9	Schlessinger et al. (24)
FGFR2c	FGF1 & Heparin	1E0O	Pellegrini *et al.* (25)

Two structures of FGFR extracellular domains bound to both FGF ligands and heparin have been solved, proposing rather different models of receptor dimerisation (24,25). Of these, the model proposed by Schlessinger *et al*. showing a ternary complex comprising two copies of heparin, FGF ligand and FGFR extracellular domain, is generally accepted. In this dimerisation model, two 1:1 FGF:FGFR complexes (as above) are arranged with two-fold symmetry such that their D2 domains interact with one another, and the two FGF ligands are located on opposite outside faces of the dimer (Figure 1.3). As observed previously in the absence of heparin (14), dimerisation is further stabilised by protein-protein interactions through FGF ligands interacting with both copies of the D2 domain. Finally, two heparin molecules bind to a positively-charged 'canyon' at the top of the dimer, each interacting primarily with one 1:1 FGF:FGFR complex, but also making contacts with the other (Figure 1.3). This structure predicts that the minimum unit length of heparin required for receptor dimerisation is a hexasaccharide, shown to also be the case in solution (24), though other studies have suggested that longer polysaccharides are required (32).

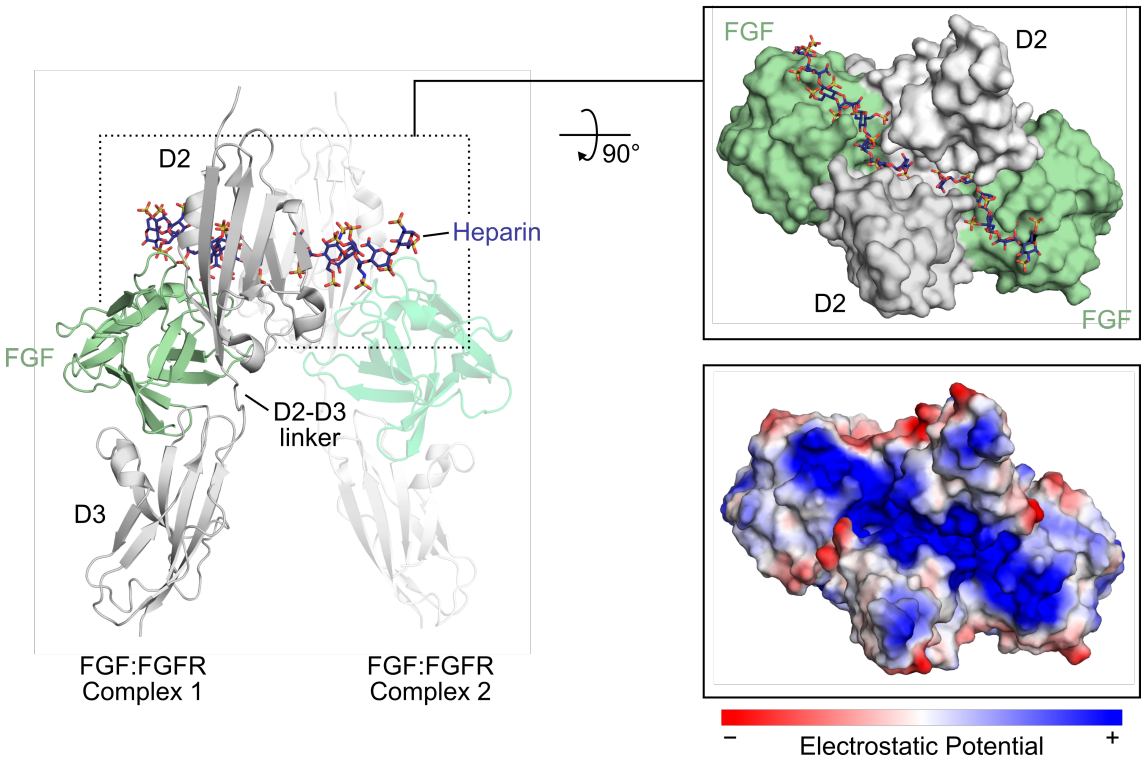


Figure 1.3. Dimerisation of the FGFR extracellular domain.
Crystal structure of FGF2- and heparin-bound FGFR1c extracellular domain in cartoon representation with heparin in stick representation (PDB entry 1FQ9). FGFR extracellular domain is shown in light grey and FGF ligand in green. One copy of the FGF:FGFR binary complex is partially transparent for clarity. The right panels show the top of the complex dimer in surface representation, coloured as in the left panel (top) and coloured according to electrostatic potential highlighting the presence of an electropositive canyon for heparin binding (bottom).

Notably, the D1 domain is missing from all solved ECD complex structures, and only two NMR structures of this domain (from human and mouse FGFR1) exist (PDB entries 2CR3 (unpublished) and 2CKN (33)). In companion with the highly conserved ‘acid box’ of FGFRs that lies between D1 and D2 (Figure 1.1), this first domain of the ECD is thought to contribute to autoinhibition mechanisms of the receptor (20,34–36), though this has not been demonstrated structurally.

1.1.3. How the transmembrane domain acts as a communication centre

To communicate ligand binding on the extracellular side to the intracellular tyrosine kinase domains, spatial and conformational information needs to be passed across the cellular membrane, a function fulfilled by the receptor transmembrane domain (TMD). Corresponding to a single helix that spans the membrane (Figure 1.1), the dimerisation of this domain is critical for kinase activation.

While canonical models of RTK activation describe ligand-induced dimerisation of receptors, there is evidence that FGFRs 1-3 may exist as dimers when unliganded to different extents (37–39), proposing that an allosteric activation model featuring ligand-induced conformational changes of a ‘pre-formed dimer’ may also be appropriate for FGFRs, similar to that of insulin receptor tyrosine kinase (40). The basis of such a model can be understood through analysis of a nuclear magnetic resonance (NMR)-determined structural ensemble of the FGFR3 TMD. In this structure, which is also the only structure available for the FGFR family, the TMD is dimeric, presented as a symmetric left-handed dimer of helices with 3_{10}- and alpha-helical character (Figure 1.4A) (41). Although FGFR transmembrane domains contain GxxxG-like motifs that are observed at the dimerisation interfaces of other receptor tyrosine kinases and single-pass transmembrane proteins (42–45), in the FGFR3 dimer, the helices cross one another approximately at the mid-point of their helices, immediately beneath these motifs (Figure 1.4A). From this observation and studies of other single-pass TMDs, it has been proposed that FGFR transmembrane helices could have multiple dimerisation states, with the use of one state over another being promoted by ligand binding at the receptor ECD.

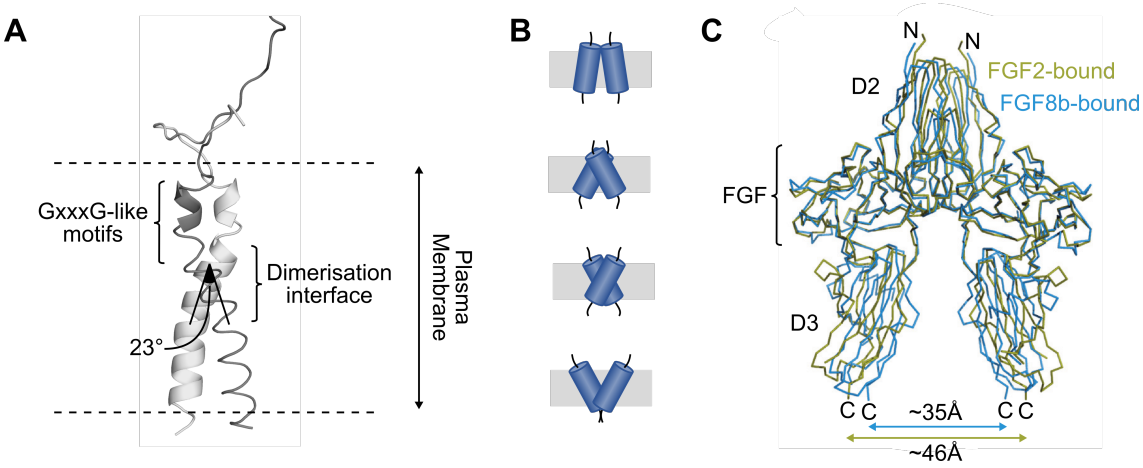


Figure 1.4. Communication of ligand binding and FGFR receptor activation by the transmembrane domain.
(A) One copy of the NMR structural ensemble of the FGFR3 transmembrane domain dimer in cartoon representation (PDB entry 2LZL), illustrating the dimerisation interface located beneath the GxxxG-like motifs and a helix crossover angle of 23 degrees. (B) Examples illustrating how different hypothetical arrangements of the transmembrane domain could result in the separation of their helix ends to different extents. (C) Overlay of FGFR2c bound to FGF2 and FGF8b in ribbon representation (PDB entries 1EV2 and 2FDB), illustrating different separation of the C-terminal end of the D3 domains. This panel adapted from Goetz *et al.* (8).

In this model of activation, the use of different dimerisation interfaces at the TMD would result in helices being arranged differently, most notably showing different degrees of helix splaying (Figure 1.4B). In the simplest form of this model, the transmembrane domain would have two states: ‘off’ and ‘on’ states. Ligand binding at the receptor ECD would induce transitions between these states, presumably mediated through precise positioning of the D3 domains that immediately precede the TMD (8). As FGFRs can bind to more than one FGF ligand, a binary ‘on-off’ description is unlikely to fully describe receptor activation mechanisms; the existence of multiple dimerisation interfaces at the TMD which correlate with specific ligand-binding scenarios may be more faithful. With respect to this model, as the NMR structure of FGFR3 TMD was solved in isolation of the rest of the receptor and indeed any ligands, this structure was proposed to correspond to a basal dimer state. Dimerisation at the GxxxG-like motif interface was suggested to correlate with full activation of the receptor, with the use of interfaces between these extremes to occur under partial activation (41).

This interpretation has been brought into question following recent Förster resonance energy transfer (FRET)-based studies investigating the effect of FGF1 and FGF2 ligand binding on FGFRs 1 to 3, where both helix separation and receptor phosphorylation were analysed (37). While these supported a model wherein FGF ligands induce varied amounts of helix separation at the TMD, and that this degree of separation correlates

very strongly with receptor phosphorylation, they argued that the dimerisation state observed by NMR corresponds to a higher activity state than would be achieved by using the GxxxG-like motifs. Additionally, through the observation that FGF1 and FGF2 induce different levels of receptor phosphorylation despite inducing apparently similar separation of TMD helices, these experiments and others highlighted that helix rotation likely also plays a role in the degree of receptor activation; i.e. to communicate the appropriate activity level, the transmembrane domain controls both the orientation *and* the proximity of kinase domains (27,37). It is worth noting here that it has been demonstrated that FGF ligands induce very different biological outcomes (46); these differences are likely mediated by inducing these different states at the TMD. The crystal structures of FGFR extracellular domains bound to different ligands (e.g. FGFR2c bound to FGF2 and FGF8b) which show differences in the distance between the C-terminal ends of their D3 domains (~46 Å for FGF2 and ~35 Å for FGF8b), demonstrate how this may be achieved (Figure 1.4D) (8).

1.1.4. The FGFR tyrosine kinase domain: the catalytic heart of the action

FGF ligand binding at the extracellular surface of FGFR receptors ultimately serves to activate the catalytic activity of their tyrosine kinase domains located within the cell, communicated, as detailed above, by the transmembrane domains. The tyrosine kinase domain (KD) of RTKs are at the heart of the signalling outputs from receptor activation, achieved through kinase-mediated phosphorylation reactions at tyrosine residues.

Crystal structures of FGFR kinase domains show that they share the prototypical bi-lobed kinase fold of eukaryotic kinases, comprising a predominantly beta-sheet N-lobe (bar the functionally critical helix αC) and a larger alpha-helical C-lobe (Figure 1.5A) (47–50). To perform phospho-transfer, both adenosine triphosphate (ATP) and substrate bind at the cleft between the two lobes. Structural elements involved in nucleotide binding are generally located in the N-lobe, namely facilitated by residues of the hinge region, the glycine-rich P-loop (or nucleotide binding loop) which folds over and encloses ATP, and by a conserved lysine residue that forms a critical salt bridge interaction with a glutamate residue of the αC helix (K514 and E531 in FGFR1) (47). Substrate binding, on the other hand, is orchestrated by the C-lobe. While ATP binds to the kinase N-lobe, the invariant aspartate residue of the His-Arg-Asp (HRD) motif (D623 in FGFR1), which is conserved among protein kinases, is located in the catalytic loop of the C-lobe. Phosphorylation of substrates therefore requires the two lobes to rotate towards one another (Figure 1.5B).

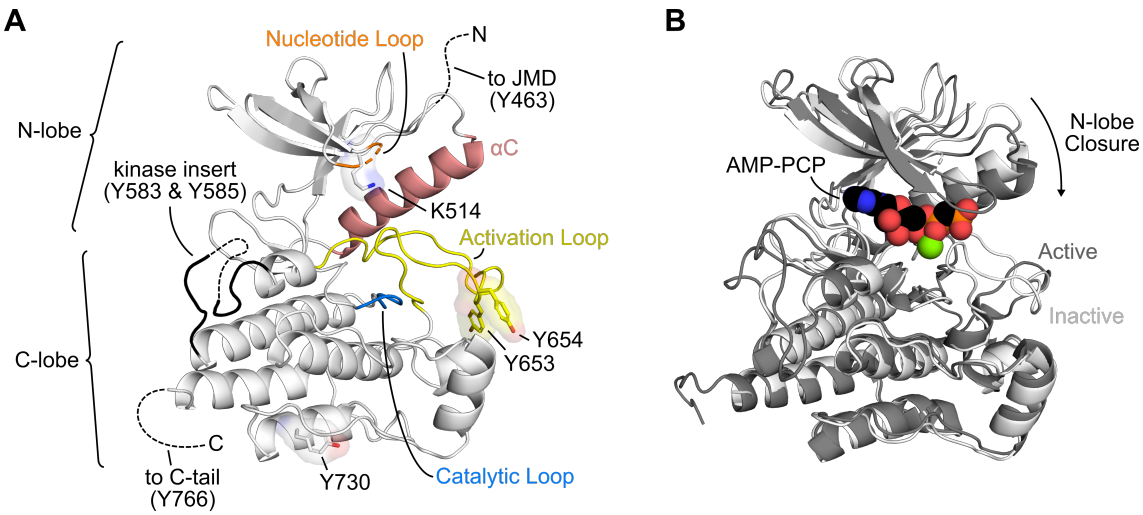


Figure 1.5. The FGFR tyrosine kinase domain.
(A) FGFR1 kinase domain (PDB entry 1FGK) in cartoon representation with its functionally-critical regions highlighted, including: the αC helix (salmon), the nucleotide binding loop (the P-loop, orange), the catalytic loop containing the HRD motif (blue), the kinase insert (black), and the activation loop (yellow). Residue K514 that is involved in ATP binding, and the location of tyrosine residues phosphorylated during receptor activation (Y463, Y583, Y585, Y653, Y654, Y730 and Y766), are shown in stick and surface representation where possible, or their locations indicated when present in unresolved regions (dashed lines). (B) FGFR1 kinase domain in inactive (light grey) and active conformations (dark grey) illustrating N-lobe closure upon binding of nucleotide (AMP-PCP in this instance, shown in spheres) (PDB entries 1FGK and 3GQI).

Following kinase activation, kinase domains in the receptor dimer autophosphorylate one another; *in vitro*, this occurs initially at Y653 (in FGFR1) of the YYKK motif in the kinase activation loop (A-loop) (51). In FGFR1, seven phosphorylatable tyrosine residues have been identified (Y463, Y583, Y585, Y653, Y654, Y730 and Y766) (52,53), five of which are phosphorylated in a sequential order *in vitro* (Y653>Y583>Y463>Y585>Y654) (51). Tyrosine residues Y653 and Y654 of the activation loop are essential for kinase activity, and their phosphorylation increases catalytic activity 50-100 and 500-1000 fold respectively (51). Other phosphorylated tyrosine residues serve as docking sites for Src homology 2 (SH2) domain-containing adaptor proteins for the stimulation of downstream signalling cascades. Phospho-Y766 of FGFR1, for example, serves as a binding site for phospholipase Cγ (54,55). Likewise, Y724 of FGFR3 (equivalent to Y730 of FGFR1) appears to play a central role in FGFR3-mediated signalling, affecting activation of phosphoinositide 3-kinase (PI3K), signal transducer and activator of transcription protein (STAT) and mitogen-activated protein kinase (MAPK) pathways (56,57).

Immediately upstream of the kinase domain, the juxtamembrane domain (JMD) (Figure 1.3) serves as a further site for the coupling of receptor activation to signalling cascades. Here, FGFR substrate 2 (FRS2α) binds constitutively to FGFR1 in a non-canonical, phosphotyrosine-independent manner, and upon FGFR-dependent phosphorylation,

acts as a scaffold for Grb2 (growth factor receptor-bound protein 2) adaptor protein for MAPK signalling (58–60).

1.1.5. Mechanisms of activity regulation of the kinase domain

Kinases are said to exist in a 'two-state' dynamic equilibrium, a model where a kinase exhibits ensembles of rigid, catalytically-inactive (i.e. inhibited) states, or dynamic, conformationally-heterogeneous active states. As kinases play integral roles in cell biology, it is crucial that their activities are tightly regulated. This can achieved at the level of the kinase domain through various means (2), retaining the kinase in an autoinhibited state with a minimal basal activity until stimulation.

In FGFRs, kinase autoinhibition is primarily achieved by the steric blocking of substrate binding, mediated by a tyrosine kinase-invariant proline residue (P663 in FGFR1) located at the C-terminal end of the activation loop (47). Autoinhibition is further established by the so-called 'molecular brake' of FGFR kinases, corresponding to a cluster of residues at the kinase hinge that, as the name suggests, places a brake on kinase activity that needs to be relieved to enable kinase activation (48). First described for FGFR2, this cluster of residues forms a hydrogen bonding network between residues H544 and N549 (of the αC-β4 loop), E565 (of the kinase hinge), and K641 (of C-lobe strand β8) in the inactive kinase state (Figure 1.6). The release of the molecular brake through breaking of this hydrogen bonding network occurs concomitantly with local conformational changes in the activation loop, and a global conformational rotation of the N-lobe towards the C-lobe, establishing a catalytically-competent state. In this state, the activation loop exists in an extended conformation no longer blocking substrate binding, and is stabilised upon phosphorylation of the YYKK motif through salt-bridge interactions with a conserved arginine residue, also of the activation loop (residues Y657 and R649 in FGFR2) (Figure 1.6) (48).

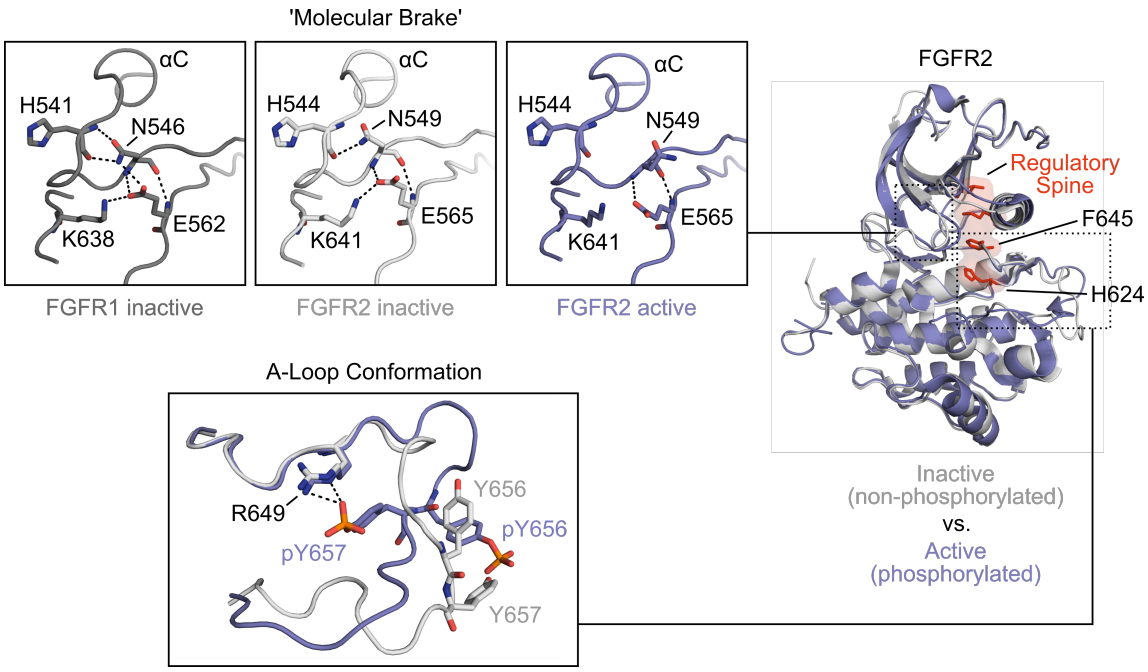


Figure 1.6. Structural hallmarks of FGFR kinase activation.
Alignment (right panel) of FGFR2 kinase domain in inactive (light grey) and active (blue) conformations in cartoon representation (PDB entries 2PSQ and 2PVF), with the hydrophobic regulatory spine and its contributing residues (e.g. H624 of the HRD motif and F645 of the DFG motif) shown in red sticks with surface representation. The expanded panels show breaking of the 'molecular brake' hydrogen bonding network in activated FGFR2 (top) and conformational changes of the activation loop upon phosphorylation of residues Y656 and Y657 of the YYKK motif (bottom). The 'molecular brake' of inactive FGFR1 kinase domain (dark grey, PDB entry 4V01) is also shown to illustrate how this feature is conserved among FGFRs. Figure adapted from Farrell and Breeze (7).

The establishment of an active kinase state is also associated with further structural hallmarks. The conformational state of the Asp-Phe-Gly (DFG) motif, located at the start of the activation loop, is one of these. Classically, this motif can exist in one of two states: the 'DFG-in' or 'DFG-out' state. When in the former of these two states, the aspartate residue of this motif plays an essential role in ATP binding through coordination of all three of its phosphate groups (either directly or via magnesium ions); similar interactions are not possible in the 'DFG-out' state resulting in catalytic incompetency (61,62). The 'DFG-out' state also breaks a cluster of hydrophobic residues called the regulatory spine that spans the two kinase lobes. The assembly of this 'spine' is essential for kinase activity, and has been exploited in the development of small-molecule kinase inhibitors (63,64).

Recently, the analysis of inhibited, partially and fully activated FGFR kinase domains revealed the presence of an extended allosteric network at the N- and C-lobe interface. Composed of the 'molecular brake', an 'A-loop plug' (which holds the activation loop in the autoinhibited conformation), and hydrophobic patches of residues around the DFG-motif and the αC helix (termed the 'DFG-latch' and 'αC-tether' respectively), these

elements mediate long-range communication between the ‘molecular brake’ and activation loop (65).

1.1.6. Roles of FGFR signalling in physiology and pathology

As discussed previously, the activation of FGFRs through the binding of its growth factor ligands leads to receptor *trans*-autophosphorylation. Ultimately, this tyrosine phosphorylation of the receptor kinase domains acts to both further activate the kinase and to form a platform for the activation of intracellular signalling cascades. It is through the action of these cascades that FGFRs incite cellular changes and play out their roles in biology. Physiologically, FGFRs serve key roles in embryonic development and organogenesis, and in adults they mediate functions in tissue repair and metabolism (66).

Evidenced by the embryonic lethality of their gene knockouts, FGFRs 1 and 2 are essential for embryonic development (67,68). FGFR-driven development and organogenesis can be largely described by a classical reciprocal signalling mechanism between epithelia and the mesenchyme through consequence of the reciprocal and tissue-specific expression of FGFR ‘b’ and ‘c’ splice variants and the FGF ligands that activate them (66). Among other roles, these mechanisms of FGFR-mediated signalling are essential in limb budding and lung development, with the latter being driven partly through reciprocal FGF9 and FGF10 ligand signalling, though largely through variants of FGFR1 and 2 alone (66). Intriguingly, the classical separation of ‘b’ and ‘c’ splice variants of receptors in epithelial and mesenchymal tissues does not always hold true, as is demonstrated by the observation of both variants of FGFR3 in the epithelium (69). This highlights differences in the roles of FGFR3 (and FGFR4) with respect to FGFRs 1 and 2: knockout of FGFR3 is not lethal in mice models but is instead associated largely with skeletal development abnormalities (70,71).

Though fulfilling different roles, FGFRs mediate their phenotypic outcomes through stimulation of the same four major intracellular signalling cascades: the MAPK, PI3K, PLCγ and STAT pathways (66). These cascades, which also exhibit considerable crosstalk, act to alter protein function and cellular expression profiles to drive cellular processes associated with cellular growth, survival, proliferation and differentiation. This can be mediated, for instance, through MAPK-driven activation of transcription factors (the factors Etv4 and Etv5 are often stimulated by FGFRs (72)) or through PI3K-driven inactivation of the pro-apoptotic transcription factor FOXO (73). Phenotypic outcome of FGFR activation appears to be related to the strength and longevity of its signalling in the cell, a factor that can be regulated at numerous levels including the availability of ligands, receptor internalisation and degradation and access to intracellular signalling partners (66). The latter can be demonstrated through the action of the adaptor protein

Grb2 that is essential in connecting FRS2α mediated signalling to the MAPK and PI3K pathways, but has also been observed to negatively regulate FGFR2 through binding to its C-terminal tail (74).

1.1.7. Dysregulation of FGFRs

Given the role of FGFRs in embryonic development and their influence on intracellular signalling cascades, it is unsurprising that their dysregulation causes a plethora of pathologies including developmental dysplasia and cancer (75,76). This subversion of regulatory mechanisms is achieved through various manners, from the overexpression of FGFRs and their ligands, to the acquisition of one or several point mutations, or through oncogenic gene fusion events (76). While doing so in different ways, these aberrations generally result in receptor hyperactivity.

1.1.7.1. Point mutations in the extracellular and transmembrane domains

Components of the FGF signalling pathways are the most frequently mutated kinases carrying non-synonymous somatic mutations in human cancers (77). Though found in all FGFRs, point mutations occur most commonly in FGFR3, and are located in all three receptor domains, achieving receptor dysregulation through different manners.

Point mutations in the extracellular domain of FGFR3 such as R248C/S249C (of the D2-D3 linker as in thanatophoric dysplasia and keratosis (78,79)) and C228R (of domain D2 as in carcinoma (39,77)) are thought to act through recapitulating receptor stimulation in the absence of FGF ligands. This is accomplished by introducing free cysteine residues at locations in the extracellular domain which can drive obligate receptor dimerisation into a conformational state that mimics what would be achieved through the action of ligand binding (Figure 1.7A).

Mutations at the transmembrane domain likely act in a similar manner, inducing conformational changes that mimic FGF binding at the extracellular domain. This could be through the introduction of free cysteine residues such as G370C/G375C (located at the N-terminal end of the TMD as in achondroplasia and keratosis (79,80)), or through the introduction of large polar residues such as with the substitutions G380R (achondroplasia and hypochondroplasia (81,82)) and A391E (Crouzon syndrome with acanthosis nigricans (83)). Though achieving it through different means, these substitutions likely alter the dimerisation interface at the transmembrane domain, inducing particular conformational states that can activate tyrosine kinase activity, independent of FGF binding; this independence has been demonstrated for A391E *in vitro* (37,39).

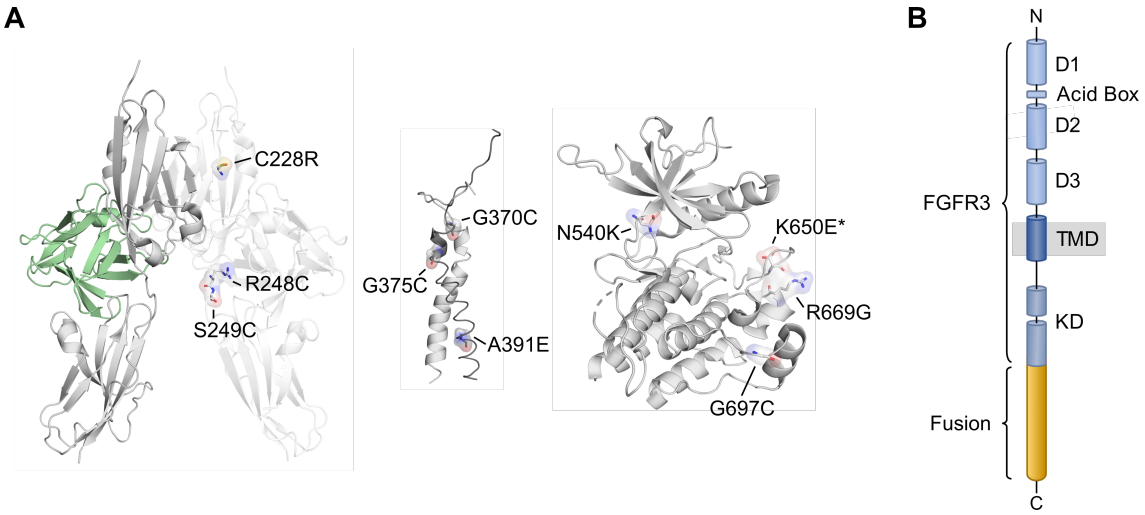


Figure 1.7. Pathological aberrations of FGFRs.
(A) The location of a selection of developmental disease and cancer-associated mutations of FGFR3 in the extracellular domain (left, PDB entry 1FQ9), the transmembrane domain (middle, PDB entry 2LZL), and the kinase domain (right, PDB entry 4K33). Mutations are shown with stick and transparent surface representation. Note that the FGFR3 kinase domain of PDB entry 4K33 contained the kinase-activating K650E mutation, highlighted here with an asterisk. (B) Schematic illustration of example FGFR3 gene fusions, showing the fusion of a second protein in frame at the C-terminal end of the receptor. Figure adapted from Farrell and Breeze (7).

1.1.7.2. Point mutations of FGFR at the kinase domain

The kinase domain is especially abundant in point mutations from various FGFR pathologies. These mutations are typically found in regulatory elements such as the 'molecular brake', the A-loop, the kinase hinge, the DFG-latch and others, affecting the activity status of the kinase *in vitro* (65). Strikingly, *in vitro* studies also show that mutations in different elements when combined have additive effects on kinase activity, demonstrating how mutations shift kinases along the two-state dynamic equilibrium (65).

Recently, a study on the prevalence and effect of point mutations in FGFR3 identified N540 (of the 'molecular brake') and K650 (of the activation loop) as mutational hotspots in FGFRs, with both mutations also eliciting significant stimulation of FGFR3 kinase autophosphorylation *in vitro* (84). Both of these mutation sites have been characterised: mutations in the 'molecular brake' are thought to facilitate the transition of a kinase to the active state by overcoming autoinhibition imposed by the brake (48), while the mutation K650E stabilities the active, extended conformation of the activation loop by mimicking phosphorylation of the YYKK motif (49). Surprisingly, the clinical prevalence of kinase domain substitutions does not directly correlate with the magnitude of their stimulatory effect. For example, the substitution R669G, while resulting in the greatest stimulation of FGFR3 kinase autophosphorylation *in vitro*, is not a mutation hotspot in FGFRs (84). Moreover, while G697C is a prevalent mutation, it has no stimulatory effect on kinase activity (though this prevalence is based on a single study) (84). Curiously, numerous

deleterious mutations in FGFR kinase domains have also been identified (84,85). Currently, the role of deleterious, neutral and inhibitory mutations found in pathologies of FGFRs is unclear but may play out at the level of the wider interactome of the kinase.

1.1.7.3. Oncogenic gene fusions of FGFRs

Although of relatively low clinical incidence, oncogenic gene fusions of FGFRs have recently come to light in a variety of cancer types (86,87). Typically, these fuse self-associating structural elements of a second protein in frame with the receptor, often near the C-terminal end, but also less frequently at the N-terminal end. Though supporting structural evidence is currently lacking, FGFR gene fusions are expected to enable ligand-independent signalling through obligate dimerisation of the receptor, in a similar manner to that reported for the TPR-MET kinase fusion (88).

While a variety of fusion partners have been identified for FGFR2 (76), reported gene fusions of FGFR3 almost exclusively occur with a protein called ‘transforming acidic coiled-coil containing protein 3’ (TACC3) (76,89). These fusions and FGFR3-BAIAP2L1 (a fusion with ‘brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1’) are exquisitely sensitive to FGFR-selective inhibitors (87,89).

1.1.8. Targeting FGFR aberrations

Spurred by the finding that FGFR signalling drives a number of cancers, there have been many efforts to develop medicines targeting FGFR aberrations. These have mainly come in the form of small-molecule inhibitors targeting the kinase domain (90–92).

Inhibitors of FGFR kinases can be characterised by their differences in kinase selectivity and inhibitor type. Most FGFR inhibitors are non-covalent and act in an ATP-competitive manner, and range from those which are non-selective (e.g. dovitinib, ponatinib and lucitanib) to those which are FGFR-selective (e.g. AZD4547, Debio-1347, BGJ398 and PD173074) (Table 1.2). With the exception of ponatinib, these inhibitors are all type I class inhibitors, binding to the kinase with a ‘DFG-in’ state. While type II inhibitors of kinases that bind to the ‘DFG-out’ state (like ponatinib) have generally proven to be more selective than their type I counterparts (93,94), no FGFR-selective type II inhibitors have been successfully developed to date. Type I inhibitors have nonetheless shown positive results in clinical trials; for example, AZD4547, a potent inhibitor of FGFRs 1 to 3, has shown promising responses in preclinical and phase I clinical trials, particularly towards tumours with FGFR amplifications (95–97), though resistance can be conferred by the gatekeeper mutation V555M in FGFR3 (98).

Table 1.2. Small molecule inhibitors of FGFRs.

Name ***Company***	**Target** ***Type***	**Measured IC_{50}** **(nM, in vitro)**	**Reference**
Dovitinib *Novartis*	Kinase domain *Multi-kinase, Type I*	FGFR1: 8 FGFR3: 9	Trudel *et al.* (99)
Ponatinib *ARIAD Pharmaceuticals*	Kinase domain *Multi-kinase, Type II*	FGFR1: 2.2 FGFR2: 1.6 FGFR3: 18.2 FGFR4: 7.7	O'Hare *et al.* (100)
Lucitanib *Clovis Oncology*	Kinase domain *Multi-kinase, Type I*	FGFR1: 17.5 FGFR2: 82.5 FGFR3: 237.5 FGFR4: >1000	Bello *et al.* (101)
AZD4547 *AstraZeneca*	Kinase domain *FGFR-selective, Type I*	FGFR1: 0.2 FGFR2: 2.5 FGFR3: 1.8 FGFR4: 165	Gavine *et al.* (102)
Debio-1347 *Debiopharm International*	Kinase domain *FGFR-selective, Type I*	FGFR1: 9.3 FGFR2: 7.6 FGFR3: 22 FGFR4: 290	Nakanishi *et al*. (103)
BGJ398 *Novartis*	Kinase domain *FGFR-selective, Type I*	FGFR1: 0.9 FGFR2: 1.4 FGFR3: 1 $FGFR3^{K650E}$: 4.9 FGFR4: 60	Guagnano *et al.* (104)
PD173074 *Pfizer*	Kinase domain *FGFR-selective, Type I*	FGFR1: 22-25 FGFR3: 29	Lamont *et al.* (105), Mohammadi *et al*. (106)
PRN1371 *Principa Biopharma*	Kinase domain *FGFR-selective, Covalent*	FGFR1: 0.6 FGFR2: 1.3 FGFR3: 4.1 FGFR4: 19.3	Brameld *et al*. (107)
TAS-120 *Tahio Oncology*	Kinase domain *FGFR-selective, Covalent*	FGFR1: 3.9 FGFR2: 1.3 FGFR3: 1.6 FGFR4: 8.3	Meric-Bernstam *et al*. (108)
BLU-554 *Blueprint Medicines Corporation*	Kinase domain *FGFR-selective, Covalent*	FGFR1: 624 FGFR2: 1202 FGFR3: 2203 FGFR4: 5	Hoeflich *et al.* (109), Kim *et al.* (110)
BLU-9931 *Blueprint Medicines Corporation*	Kinase domain *FGFR-selective, Covalent*	FGFR1: 591 FGFR2: 493 FGFR3: 150 FGFR4: 3	Hagel *et al.* (111)
H3B-6527 *H3 Biomedicine, Eisai Incorporation*	Kinase domain *FGFR-selective, Covalent*	FGFR1: 320 FGFR2: 1290 FGFR3: 1060 FGFR4: <1.2	Joshi *et al.* (112)
SSR128129E *N/A*	FGFR1 ECD *Reversible*	FGFR1: 1900	Bono *et al.* (113)

Note: ECD (extracellular domain). The following inhibitors have multiple names: Dovitinib (CHIR258 and TKI258); ponatinib (AP24534); lucitanib (E-3810); Debio-1347 (CH5183284); BGJ398 (NVP-BGJ398 and infigtatinib). This table is adapted from Farrell and Breeze (7).

Covalent inhibition of FGFR kinases presents an alternative method for selective kinase inhibition, and has found particular success in the development of FGFR isoform-selective inhibitors. These inhibitors also occupy the ATP binding cleft between the two kinase lobes, but do so irreversibly through covalent linkage to cysteine residues near the nucleotide binding site. Several covalent inhibitors of FGFRs have been developed including those with pan-FGFR activities (e.g. PRN1371 and TAS-120) or FGFR4-selective activities (e.g. BLU-554, BLU-9931 and H3B-6527) (Table 1.2).

Opportunities for isoform-selective inhibition of FGF receptors are also presented through targeting their extracellular domains (Table 1.3). Research efforts here are best exemplified by anti-FGFR2 and anti-FGFR3 monoclonal antibodies or antibody-drug conjugates (114–117), but also include FP-1039 (a FGF-ligand trap which sequesters FGFs preventing their binding to FGFRs (118)), and the novel small molecular inhibitor SSR128129E which binds selectively to FGFR1 extracellular domain (Table 1.2) (113,118,119).

Table 1.3. Alternative approaches for targeting FGFR signalling.

Name ***Company***	**Target**	**Type**	**Reference**
Bemarituzumab *Five Prime Therapeutics*	FGFR2b ECD	Monoclonal antibody	Pierce *et al.* (114)
BAY1179470 *Bayer*	FGFR2 ECD	Monoclonal antibody	Schatz *et al.* (116)
BAY1187982 *Bayer*	FRFR2 ECD	Antibody-drug conjugate	Sommer *et al.* (115)
LY3076226 *Eli Lilly*	FGFR3 ECD	Antibody-drug conjugate	Eli Lilly and Company (117)
MFGR1877S *Genentech*	FGFR3 ECD	Monoclonal antibody	Trudel *et al.* (120)
FP-1039 *Five Prime Therapeutics*	FGF2 and others	FGF ligand trap	Blackwell *et al.* (118)

Note: ECD (extracellular domain). The following inhibitors have multiple names: bemarituzumab (FPA144); FP-1039 (GSK3052230). This table is adapted from Farrell and Breeze (7).

1.2. Understanding activation of full-length RTKs

While studies on isolated domains of FGF receptors and other RTKs have provided invaluable insights into the roles of individual domains in the receptor activation process, in the absence of high-resolution structures of intact, full-length receptors, we cannot truly understand how this highly-important class of proteins are activated in their physiological context. This lack of structural information for intact RTKs is a crucial limitation to understanding in the field and is a clear target to be resolved. While efforts in the past have failed thus far, pioneering studies in the past decade have provided glimpses of full-length receptors, and as more membrane protein structures are being solved year on year (121), there is promise that high-resolution structures of RTKs should soon be within our reach.

1.2.1. Current structural insights into the activation of full-length RTKs

To date, only low-resolution negative-stain electron microscopy (EM) structures of full-length RTKs have been determined. These structures and snapshots of the insulin receptor (IR) (122), of epidermal growth factor receptor (EGFR) (123,124), platelet-derived growth factor receptor beta (PDGFRβ, 27 Å resolution) (125) and KIT receptor (25 and 28 Å resolution) (126), indicate the arrangement of extracellular and intracellular domains in full-length receptor dimers (Figure 1.8). However, while previously-solved crystal structures can be docked into these low-resolution maps, they are unable to inform on why these arrangements are important and exactly how information is communicated across the membrane between these two regions.

Arguably, our best understanding of full-length RTK activation is currently provided by the insulin receptor. As an obligate covalent dimer in both its active and inactive states (40), rather than inducing changes in oligomeric state, the binding of insulin results in a concerted conformational rearrangement of the receptor that promotes intracellular kinase activation. In a recent 2018 study, low-resolution structural snapshots were captured of this activation event by negative-stain electron microscopy (EM) (Figure 1.8D) (122). These suggested that the extracellular domain plays an autoinhibitory role in kinase activity by arranging the transmembrane helices of the dimer such that they are separated in the inactive state, preventing the two tyrosine kinase domains from contacting one another. This is overcome by insulin binding, causing extracellular domain rearrangements that allow the transmembrane domain and kinase domains to dimerise, or at least allow them to be in closer proximity to one another (122,127,128). This model is supported by recent cryo-EM structures of the IR extracellular domain (129,130) that provide higher resolution insights into how these rearrangements may take place at the extracellular domain.

These studies and others highlight the critical role that the transmembrane domain plays across the RTK family in communicating activation stimulus of ligand binding to kinase activation. While there are contradictions on the scale of the conformational rearrangements induced by ligand binding (e.g. IR transmembrane helices are proposed to be monomeric when the receptor is unliganded while FGFRs are proposed to already be dimerised, as discussed previously), there is no doubt that the transmembrane helices play a central role in the activation of full-length RTKs. This is emphasised by the fact that pathological mutations are often found in and around this domain of many RTKs including FGFRs (41,45,131).

1.2.2. An outlook for full-length FGFR and RTK structural biology

Although RTKs share a conserved domain architecture (i.e. extracellular ligand-binding domains, a single-pass transmembrane helix, a juxtamembrane domain, intracellular tyrosine kinase domain and C-terminal tail), the exact activation and autoinhibitory mechanisms between families differ (2). Thus, while a single high-resolution RTK structure will be pivotal for our knowledge of RTK biology, many structures of the same receptor in multiple states and of different receptors from different families will be necessary to obtain a comprehensive understanding of their complexity. Although the reductionist approach has and is providing very useful insights into the biology of the individual domains of receptors, the precise details of their activation, such as whether large or smaller and more subtle conformational changes in domain arrangement are critical, will likely remain unclear until structures of intact receptors are available for each RTK subfamily.

While representatives of the extracellular (17,24,119), transmembrane (41), and intracellular tyrosine kinase domains (47,93,94,132) of the FGFR family have been structurally elucidated in isolation, no full-length structure at any resolution for any family member has been determined to date. Moreover, as no FGFR family member has had each of its individual domains solved thus far, models of activation rely on piecing together information from different family members; this approach, though valuable, does not allow for subtleties between members to be appreciated. Obtaining full-length structures of FGFR family members would therefore be useful not only in expanding our understanding of differences and similarities between RTK families, but also in fine-tuning our models of FGFR activation.

It is notable that all structural information available for full-length RTKs to date has been generated from EM. While there are additional factors that have prevented their high-resolution structural determination (as with other membrane proteins, detailed in 1.4.1), the inability thus far of X-ray crystallography to determine structures of full-length receptors while showing success with individual domains, may in part be because intact

RTKs possess too much inherent conformational heterogeneity that they are not crystallisable. Negative-stain EM images of liganded EGFR, where two distinct dimer states were identifiable in the same sample, support this premise (Figure 1.8A), highlighting the challenges that will be faced in high-resolution structural determination. These challenges are further highlighted in the first published effort of full-length RTK structural determination by cryo-EM, released in August 2019 (133). In this study, once more of the insulin receptor, neither the kinase domains nor (most of) the transmembrane domains were resolved, even though the full-length receptor was used in this study, and despite that the receptor was also in a ligand-bound active state (133). This highlights that there is considerable heterogeneity at the intracellular side of the receptor even when it is activated, a feature that is functionally logical as their kinase domains can have multiple phosphorylation states and various signalling outputs, but that will also be challenging to structurally resolve.

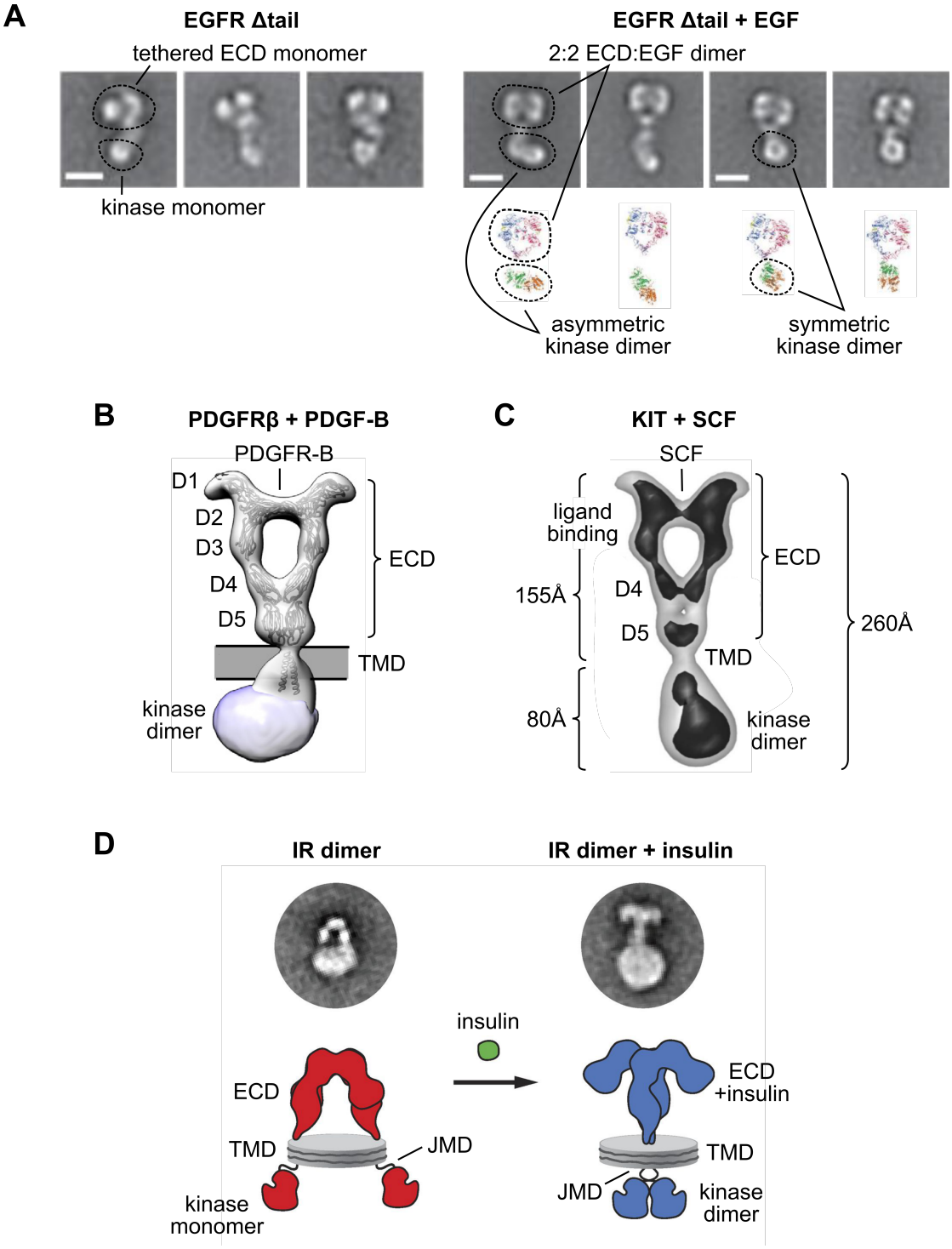


Figure 1.8. Low-resolution structural insights into RTK biology.
(A) Negative-stain 2D classes of EGFR Δtail in unliganded and liganded forms showing visualisation of the extracellular domains (ECD) and intracellular kinase domains, and conformational variation in each instance. Proposed domain arrangements using solved crystal structures are shown beneath in cartoon. Figures have been adapted from Li-Zi *et al.* (123) and Diwanji *et al.* (128). (B) Negative-stain EM volume of PDGFRβ bound to PDGF-B with known structures of the ECD and transmembrane domain (TMD) docked. Adapted from Chen *et al.* (125). (C) Negative-stain EM volume of KIT receptor bound to SCF (stem cell factor). Adapted from Opatowsky *et al.* (126). (D) Negative-stain EM 2D classes of unliganded and liganded insulin receptor (IR) with illustrative interpretations beneath. Adapted from Gutmann *et al.* (122).

1.3. The role of the chaperone Hsp90 in kinase regulation

As previously introduced, it is essential that kinases are tightly regulated such that they are active only when required. This regulation can be achieved in many ways, from those which are kinase-specific to those relevant to kinases more generally. Examples of the former can be seen in RTKs where kinase activation is controlled to occur only following ligand stimulation, and the latter in the actions of the cellular chaperone heat shock protein 90 (Hsp90).

1.3.1. The ubiquitous cellular chaperone Hsp90

Hsp90 is a highly-conserved, ATP-dependent molecular chaperone which facilitates the folding and maturation of a diverse set of client proteins in eukaryotic cells. Acting late in the cytosolic protein homeostasis network, Hsp90 plays a central role in the correct folding of unstable, partially unfolded or misfolded clients, preventing their aggregation in the crowded cellular environment (134). While originally discovered as one of several proteins with elevated expression levels under stress conditions (such as ‘heat shock’) (135), Hsp90 has roles in many cellular processes under physiological conditions, a consequence of its varied client base. In fact, Hsp90 is functionally very flexible, being involved not only in protein folding and activation, but also in the assembly of protein complexes, and the enabling of ligand binding to various proteins (134).

The importance of Hsp90 in cellular function is highlighted by its conservation among many bacteria and all eukaryotes (136), where several isoforms of Hsp90, each with individual roles, exist. In humans, at least four isoforms of Hsp90 have been described: the two major cytosolic isoforms, Hsp90α and Hsp90β, which together are the most abundant proteins in the cell; the endoplasmic reticulum isoform, Grp94; and the mitochondrial isoform, TRAP1 (134,137). Sharing 86% sequence identity and 93% sequence similarity at the protein level (137), the alpha and beta isoforms of Hsp90 - the major isoforms in the cell - are highly similar, differing most in their expression profiles; while Hsp90β is constitutively expressed, Hsp90α is induced to express when cells are under stress (134,138). Generally, these two cytosolic isoforms will be discussed here; however, many insights into Hsp90 structure and function have been yielded from studies on their bacterial and yeast homologues.

1.3.2. The conformational cycle of Hsp90

Hsp90 functions as a homodimer, composed of two promoters each containing an amino-terminal ATP-binding domain (NTD), a middle domain (MD) responsible for client binding and ATP hydrolysis, connected to the NTD via a charged linker domain, and a carboxy-terminal domain (CTD) responsible for dimerisation (Figure 1.9). The ‘chaperoning’ of client proteins by Hsp90 is dependent on its ATPase activity, which, due

to the splitting of the ATPase components in both the NTD (responsible for ATP binding) and MD (containing the catalytic loop), requires large-scale conformational changes to cycle through.

Summarising many studies, the Hsp90 cycle can be broken down into four major conformational states and changes between these: the open state, the intermediate state, and two distinct closed states formed from dimerisation of the NTDs (Figure 1.9). Progress through these states correlates with the nucleotide binding state of the homodimer, coinciding with hydrolysis of ATP. While ATP hydrolysis is essential for function of Hsp90, the rate of hydrolysis is very slow, with a turnover of only 0.10 or 0.08 units of ATP per minute for human Hsp90α and Hsp90β, respectively (139). Structural and biophysical analyses indicate that the formation of a conformational state that permits ATP hydrolysis is the rate-limiting step of ATP turnover, rather than the binding and actual hydrolysis of ATP, both of which are relatively rapid (134,140,141). This rate limiting step reflects the large-scale changes required to form this hydrolysis-competent state that is composed of both interactions within and between protomers. The long residence of Hsp90 in the open state is proposed to be a critical aspect of its 'chaperoning' action on clients.

The conformational cycle of Hsp90 begins with an open 'V-shaped' conformation, where the NTDs can exist in either a nucleotide-free or ATP-bound state (Figure 1.9). While it has been reported that ATP binding to human Hsp90 does not commit the chaperone to ATP hydrolysis (139), the binding of ATP represents the first step in a series of coordinated conformational transitions that lead to this fate. While conformational changes of human Hsp90 are not driven by nucleotides *per se*, explaining its cycle in this context is useful in the first instance. Crucially, to allow NTD dimerisation and attain a closed state necessary for hydrolysis, prior structural rearrangement of a 'lid' region in the NTD is required. This change can occur in response to ATP binding to the nucleotide cleft of the NTD, inducing the 'lid' region – a helix-turn-helix motif – to move rigidly through almost 180° to enclose ATP in its binding pocket (Figure 1.10) (142,143). In doing so, a hydrophobic patch is exposed that becomes buried upon NTD dimerisation, a process which also leads to domain swapping of the N-terminal strands of each NTD (Figure 1.10) (143,144). Following their dimerisation, the NTDs next associate with the MDs, a crucial step that rearranges the 'catalytic loop' located in the MD such that a competent ATP hydrolytic centre is formed between these two domains (145). Following ATP hydrolysis, Hsp90 returns to the open state, facilitated by NTD dissociation and release of adenosine diphosphate (ADP) and inorganic phosphate (Figure 1.9).

While human Hsp90 is able to cycle through ATP hydrolysis alone, and seems to be able to do so with or without bound clients (134), the cycle is highly regulated *in vivo*, primarily

by a host of co-chaperones which interact with Hsp90 at different stages of its catalytic cycle. Regulation is also achieved at the expression level and through an extensive array of post-translational modifications that can affect both local binding interfaces and act as longer-range allosteric communication switches (134).

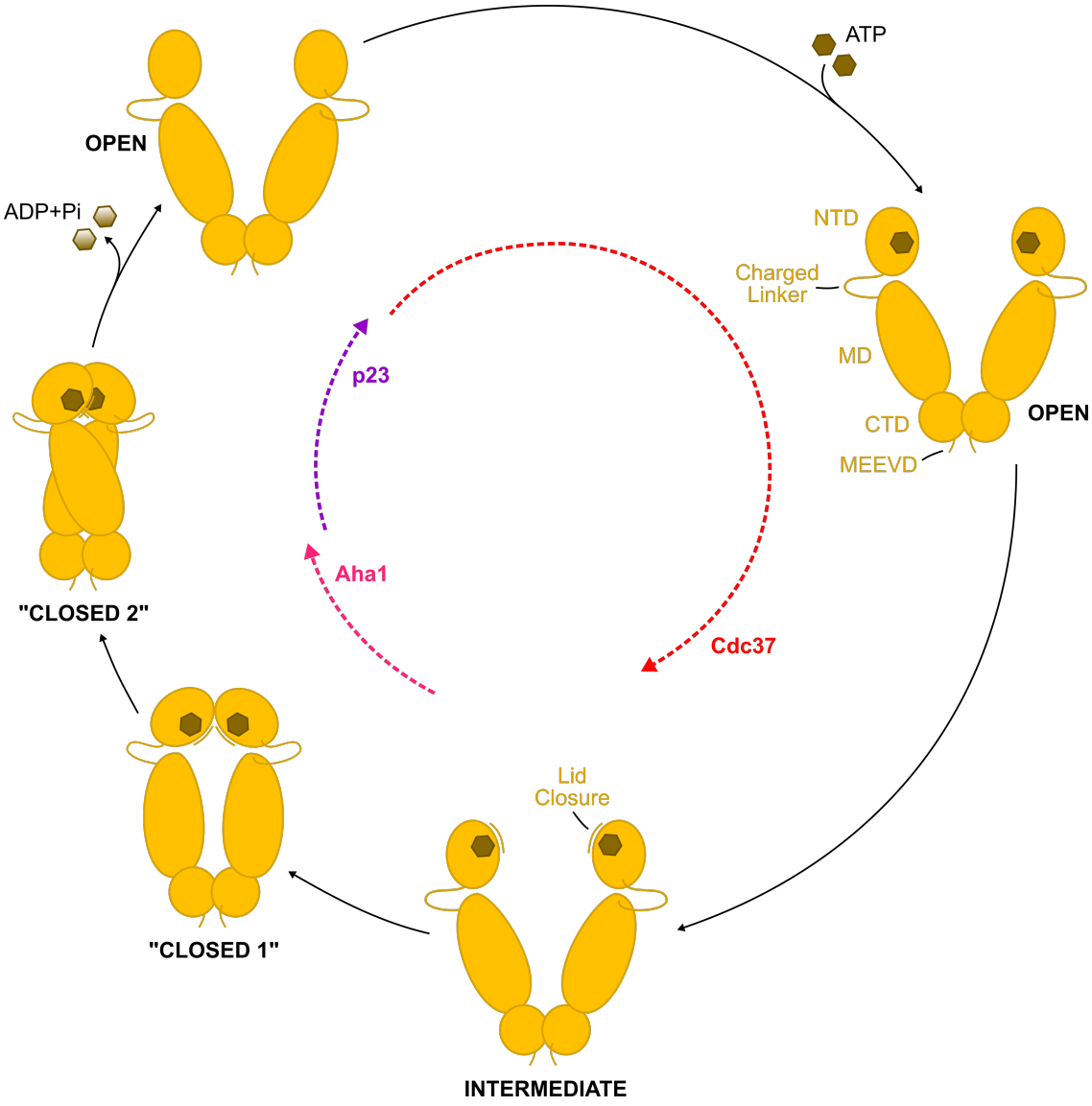


Figure 1.9. The conformational cycle of Hsp90.
A schematic representation of the conformational cycle of Hsp90, illustrating the different states, the approximate time spent in each state, and the action of example co-chaperones (Cdc37, Aha1 and p23) in driving the cycle. Each protomer of dimeric Hsp90 is composed of an amino-terminal domain (NTD), a middle domain (MD) and a carboxy-terminal domain (CTD). A charged linker connects the NTD and MD, and a Met-Glu-Glu-Val-Asp (MEEVD) motif follows the CTD. Following binding of ATP to the NTD, conformational changes lead to 'lid' closure, allowing dimerisation of the NTDs of each protomer. Further conformational rearrangements establish interactions between MD and NTDs, forming an ATP hydrolysis competent state. The release of ADP and inorganic phosphate (Pi) follows, returning Hsp90 to its open nucleotide-free. For clarity, clients nor co-chaperones are shown. This figure is adapted from Schopf *et al.* (134).

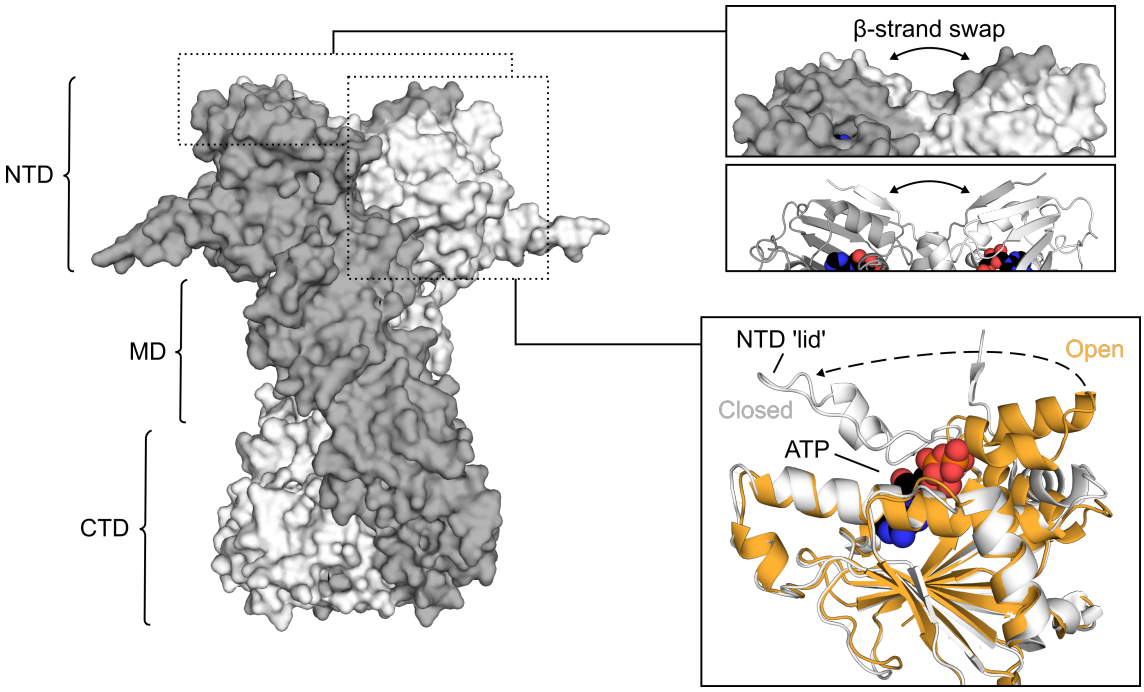


Figure 1.10. ATP-induced conformational changes of Hsp90.
Surface representation of a closed Hsp82 dimer (the yeast homologue of human Hsp90) with the two protomers coloured in light and dark grey, respectively (PDB entry 2CG9). The approximate location of the NTD, MD and CTD of each protomer are shown. Expanded panels show the domain swapping of a β-strand at the NTDs in a closed dimer (top), and an overlay of Hsp90 NTDs in the presence (white) and absence (orange) of ATP (PDB entries 2CG9 and 1AH6), illustrating closure of the 'lid' motif (bottom). In these panels, structures are shown in cartoon representation and ATP as spheres.

1.3.3. Co-chaperones of Hsp90

So far, more than 20 co-chaperones of Hsp90 have been found in eukaryotes, ranging from those with general roles in Hsp90 regulation such as p23, FKBP51/2, Aha1 (activator of Hsp90 ATPase protein 1), HOP (Hsp70-Hsp90 organising protein), and PP5 (serine/threonine-protein phosphatase 5), to those implied to have roles specific to a subset of clients such as Cdc37, Unc45, and Tah1 (TPR-containing protein associated with Hsp90) (Table 1.4) (134,146). The actions of co-chaperones also vary, from roles in client recruitment and transfer (e.g. HOP and Cdc37), modulation of the cycle through activation or inhibition of ATPase activity (e.g. Aha1 and p23, respectively), and through handling the introduction and removal of post-translational modifications of Hsp90 and other co-chaperones (e.g. PP5). In many cases, particularly during client recruitment to Hsp90, co-chaperones fulfil multiple roles, often also inhibiting the ATPase activity of Hsp90 to enable client transfer.

While many co-chaperones interact with Hsp90 through binding via their tetratricopeptide repeat (TPR) domains to its Met-Glu-Glu-Val-Asp (MEEVD) motif which follows the CTD,

others are found to interact with Hsp90 at the CTD more generally, and at the MD or the NTD (134,143,147–149) (Figure 1.11A). As the binding interfaces between Hsp90 and many co-chaperones are shared, the binding of some co-chaperones is mutually exclusive. For example, Aha1 competes with p23, Cdc37 and HOP for Hsp90 binding (150), but not the co-chaperone Cpr6 (151). These binding exclusivities reflect the roles of each co-chaperone, and highlight how controlled sequential exchange of these co-chaperones can drive Hsp90 around its conformational and catalytic cycle in a unidirectional manner (Figure 1.9) (150–152).

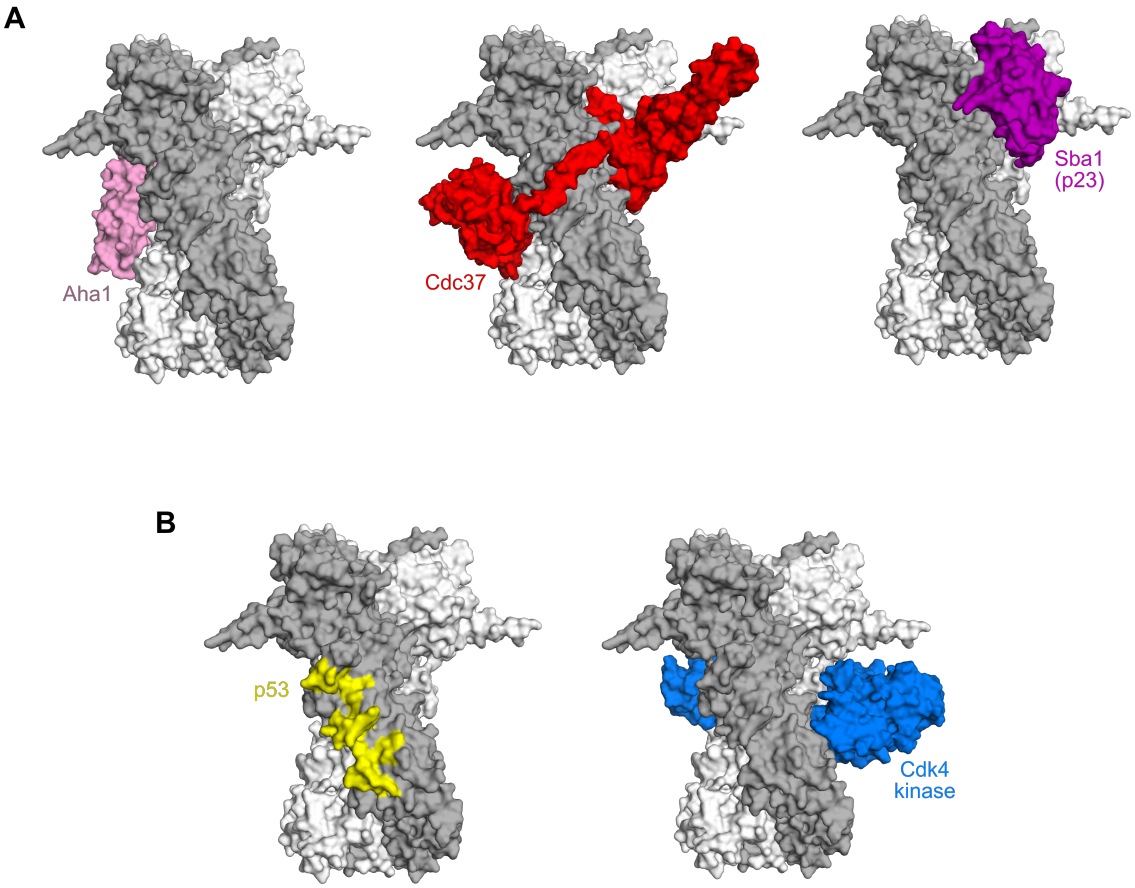


Figure 1.11. Co-chaperone and client binding sites on Hsp90.
(A) The binding sites of the co-chaperones Aha1, Cdc37 and Sba1 (yeast homologue of p23) mapped onto the closed structure of Hsp82 (yeast homologue of Hsp90) (PDB entry 2CG9), shown with surface representation. The structures of co-chaperones are from PDB entries 1USU (Aha1), 5FWL (Cdc37) and 2CG9 (Sba1). Note that NMR studies indicate that Aha1 also displays binding to the NTD of Hsp90 that is not shown here (147). (B) As in A, except showing the binding sites of the clients p53 and Cdk4 (cyclin dependent kinase 4) mapped onto Hsp82. The binding interface of p53 is as reported in Hagn *et al.* (153), and the Cdk4 structure is from PDB entry 5FWL. Note that p53 also shows binding to Hsp90 CTD that is not illustrated here (153).

Table 1.4. Co-chaperones of Hsp90.

Co-chaperone	Hsp90 binding site	Function
Aha1	NTD, MD	Stimulates ATPase activity, Encourages the closed 1 state
Cdc37	NTD, MD	Recruitment and maturation of kinases, Inhibits ATPase activity, Prevents NTD lid closure and dimerisation
CYP40 (Cpr6/7 in yeast)	MEEVD	PPIase, Implicated in the maturation of the glucocorticoid receptor
FKBP1/2	MEEVD	General co-chaperones
HOP	Mainly MEEVD, Also, MD and CTD	Client transfer from Hsp40-Hsp70, Inhibits ATPase activity, Stabilises the open state
p23 (Sba1 in yeast)	MTD, MD	Inhibits ATPase activity, Stabilises the closed 2 state
PP5	MEEVD	Dephosphorylation of Hsp90, Implicated in kinase maturation
Tah1	MEEVD	Component of the Rvb1-Rvb2-Tah1-Pih (R2TP) complex
Unc45	MEEVD	Implicated in myosin-dependent processes

Note: Aha1 (activator of Hsp90a ATPase homologue 1); Cpr6 (cyclosporin-sensitive proline rotamase 60; CTD (carboxy-terminal domain); FKBP51 (51 kDa FK506-binding protein); HOP (Hsp70/Hsp90-organising protein); MEEVD (Met-Glu-Glu-Val-Asp motif); MD (middle domain); ND (amino-terminal domain); PP5 (Ser/Thr phosphatase 5); PPIase (peptidyl-prolyl cis-trans isomerase); Tah1 (TPR-containing protein associated with Hsp90). Reproduced from Schopf *et al.*(134).

1.3.4. The diverse client pool of Hsp90

As previously mentioned, Hsp90 interacts with a wide range of protein clients; in fact, several hundred clients have been catalogued (http://www.picard.ch/Hsp90Int, (154)). While diverse, many of these are components of signal transduction pathways, with an estimated over 60% of the human kinome, 30% of human E3 ubiquitin ligases, and approximately 7% of all transcription factors being found to interact with Hsp90 (155). Other clients include membrane proteins such as GPCRs and ion channels, to cytoskeletal proteins. The involvement of Hsp90 with such a diverse set of client proteins means that it plays a central role in cell biology, being involved in processes including DNA repair, development, and cancer (134).

Unlike clients of molecular chaperones such as heat shock protein 70 (Hsp70) that act earlier in the proteostasis network, the determining features of Hsp90 clients are unclear. While there is evidence that Hsp90 interacts with its clients via unfolded features as found with other molecular chaperones (148,156), studies on clients such as kinases and the glucocorticoid receptor have indicated that clients do not inherently present these features for recognition by Hsp90, instead requiring their recruitment to Hsp90 via co-chaperones (148,157). Different studies have also provided conflicting evidence as to the ‘foldedness’ of clients bound to Hsp90, as in the case of the tumour suppressor p53

(153,158,159). These findings raise questions as to how Hsp90's diverse set of clients, each with different structural features, come to interact with Hsp90 and how the chaperone can fulfil its rather different roles using one universal ATPase cycle.

The dynamic nature of interactions between Hsp90 and clients has made structural studies of client-chaperone complexes difficult, but several breakthroughs have enabled the binding sites of a number of clients on Hsp90 to be defined. Prior to obtaining structural data of these complexes, mutagenesis studies indicated that the middle domain (MD) of Hsp90 was important for client interactions (160,161). Later, structures of bacterial and yeast Hsp90 with MD coverage suggested that hydrophobic patches of residues in the MD constituted the binding site of clients (145,162). These hypotheses have largely held true where client-chaperone complex structures have been determined, with the glucocorticoid receptor (GR), kinases, Tau and p53 all making contacts with the MD (148,153,156,163). However, interactions with other domains of Hsp90 have also been observed; for example, p53 also interacts with the CTD of Hsp90 (153), and Tau interacts with the NTD (156). While there is still a relative paucity of information in this area, parallels between the interactions of clients are emerging with clients using overlapping interaction surfaces (Figure 1.11B).

1.3.5. The chaperoning of kinases by Hsp90

Kinases represent one of the largest and best studied clients of Hsp90. As mentioned, over 60% of the human kinome has been proposed to interact with Hsp90 (155), yet the molecular determinants of kinases which define whether a kinase is a client or not are still unclear. Curiously, despite eukaryotic kinase domains being structurally very similar, not all kinases are considered clients, even within kinase families; for example, in the FGFR family, only FGFR3 (and to some extent FGFR4) interact significantly with Hsp90 (164). Moreover, point mutations in kinases that do not significantly affect kinase structure can transform kinases between non-client and client status (165). While a shared structural motif among client kinases has not been found, evidence suggests that properties of the kinase N-lobe alone influence whether or not a kinase is a client (164–169).

Following the observation that there is a clear, strong positive correlation between kinase thermal stability and client status (155,165), a model of 'clientness' has emerged where all kinases are thought to contain the features necessary for interaction with Hsp90, but that some kinases present these features more readily and more often than others (170). Molecular dynamic simulations of different variants of Src kinase that show a relationship between the conformational dynamics, the thermal stability, and the client status of kinases further supports this idea (171). These suggest that rather than there being a binary description of kinases as clients (i.e. clients or not), all kinases can be thought of

as lying somewhere on a continuum of Hsp90 dependence and independence. This model nicely satisfies the observation that different kinases have different 'client strengths', and how point mutations can alter this strength.

1.3.6. The recruitment of kinases to Hsp90 by the co-chaperone Cdc37

As introduced previously, the interaction of kinases with Hsp90 is dependent on their prior recruitment by the co-chaperone Cdc37 (172). Previously thought to only recognise client kinases, recently, Cdc37 has been proposed to also partially remodel kinases prior to their transfer to Hsp90, though this effect is poorly defined (173). The recruitment of client kinases by Cdc37 occurs in a two-state process beginning with the recognition of and binding to all kinases, followed by the formation of stable binary co-chaperone-kinase complexes using a second interaction interface upon client detection. These interactions are mediated by the amino-terminal (N-Cdc37) and carboxy-terminal (C-Cdc37) ends of Cdc37, respectively (Figure 1.12) (173). The middle domain of Cdc37 (M-Cdc37) is involved in binding to Hsp90 (149).

The role of N-Cdc37 in client recognition and sorting of kinases was first established from biochemical studies that showed that mutation of residues in the N-terminus of Cdc37 (namely V2, D3, Y4 and W7) significantly affected or abolished client interaction (174), later reinforced by the finding that Cdc37 lacking its first 30 residues is unable to bind to the kinase bRaf (175). From the findings that features of the kinase N-lobe dictate client strength, that Cdc37 outcompetes ATP binding to kinases, and that it itself is outcompeted by ATP-competitive inhibitors of kinases (175), it is likely that N-Cdc37 detects and sorts kinases through interactions at or near the ATP-binding cleft between the two kinase lobes. Following client detection, likely by probing the conformational stability of bound kinases, secondary interactions via a hydrophobic patch of residues in C-Cdc37 can occur (Figure 1.12) (173,176). This formation of a stable binary Cdc37-kinase complex appears reliant on the ability of Cdc37 to remodel the kinase, a characteristic that likely allows N-Cdc37 to distinguish between kinases on the client to non-client continuum. While the regions of the kinase remodelled by Cdc37 are not well defined, Cdc37 appears to partially locally unfold a small portion of the kinase at a region that can be also be marginally perturbed by Hsp90 alone (173).

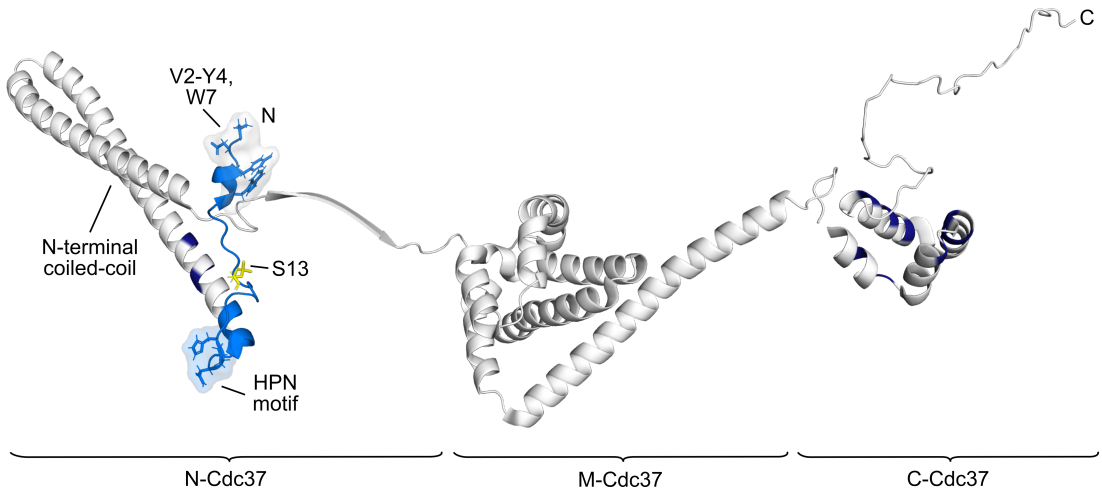


Figure 1.12. Client-binding regions of co-chaperone Cdc37.
A cartoon representation of full-length co-chaperone Cdc37 stitched together using structures of residues 1-240 of PDB entry 5FWL, residues 241-293 of PDB entry 1US7, and residues 194-378 of one structure of the NMR ensemble of PDB entry 2N5X. The amino-terminal (N-Cdc37), middle (M-Cdc37), and carboxy-terminal (C-Cdc37) regions are annotated. Residues 1-30 found to be important for kinase recognition (also termed the N-terminal coil) are coloured blue, with residues V2-Y4 and W7 shown in stick representation. The His-Pro-Asn (HPN) motif found to mimic αC-β4 of Cdk4 in the cryo-EM structure of kinase-bound Hsp90 (148) is shown in stick and surface representation. Further residues found to be implicated in the binding of sb-Raf^{V600E} kinase, notably in C-Cdc37 (173), are highlighted in dark blue. Residue S13, found to have a role in Hsp90 ternary complex stability is shown in yellow sticks. This figure is inspired by and based upon that from Verba and Agard (170).

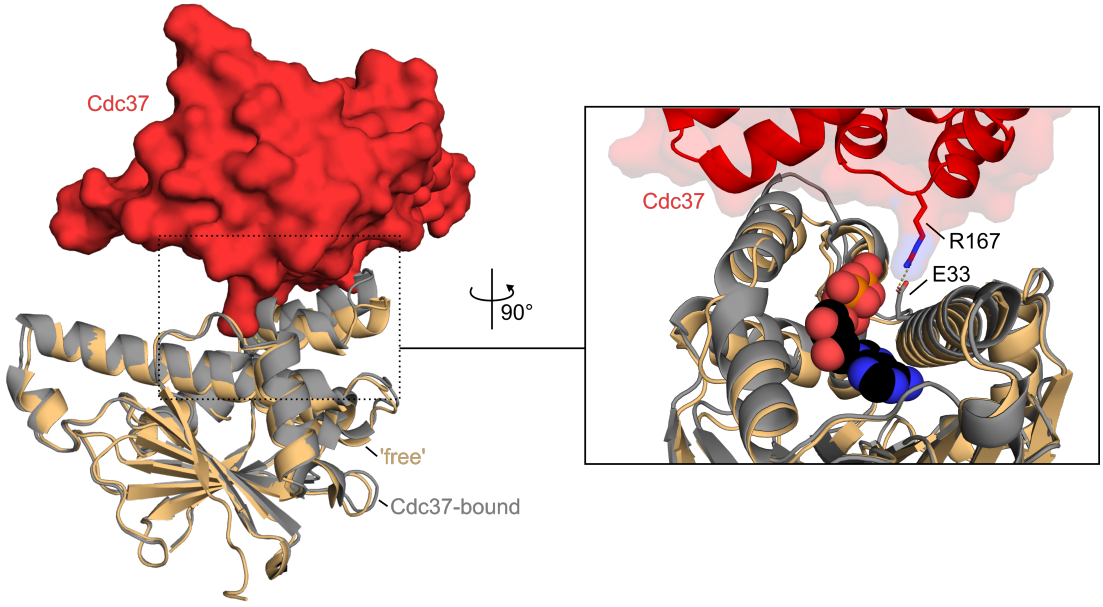


Figure 1.13. The binding of Cdc37 to Hsp90 blocks conformational rearrangements of the Hsp90 amino-terminal domain 'lid'.
Structure of human Cdc37 (red surface representation) bound to the NTD of yeast Hsp90 (grey cartoon), showing the locking of the 'lid' motif in its open conformation (PDB entry 1US7). Nucleotide-free Hsp90 NTD is shown in orange cartoon for comparison (PDB entry 1AH6). In the expanded panel, the sequestering of Hsp90 residue E33 by R167 of Cdc37 is shown, with the binding location of ATP (should it be possible) in spheres, as observed in the ATP-bound structure (PDB entry 2CG9).

Consistent with its role in the recruitment of kinases, Cdc37 inhibits the ATPase activity of Hsp90, presumably allowing kinase transfer to take place (177). Co-crystallisation of the middle and C-terminal domains of human Cdc37 (residues 138-378) with yeast Hsp90 revealed that Cdc37 binds to the NTD of Hsp90 in its open conformation, preventing NTD 'lid' closure, sequestering the catalytic residue Glu33 of Hsp90, and blocking dimerisation of its NTDs to fulfil this catalytic inhibition (149,178) (Figure 1.13). Cdc37 has also been found to bind to the MD of Hsp90, first reported for nematode Cdc37 (179), and later observed in a high-resolution structure of a ternary complex of Hsp90β bound to Cdc37 and Cdk4 (148). These data thus suggest that Cdc37 can bind to Hsp90 at different sites, possibly reporting on different states during kinase transfer; however, this has yet to be explicitly demonstrated.

1.3.7. Hsp90 locally unfolds kinases

In the high-resolution structure of a Cdk4-bound ternary complex of Hsp90, determined by cryo-EM, an unexpected dramatic local unfolding of Cdk4 kinase was observed (148). Capturing Hsp90β in a nucleotide-bound state, in this structure, the N and C-lobes of Cdk4 are separated, lying on opposing sides of the Hsp90 dimer at its MD, connected by an unfolded polypeptide corresponding to strands β4 and β5 of Cdk4's N-lobe that passes through the centre of the Hsp90 dimer (Figure 1.14). Cdc37 is found to bind to both lobes of the kinase, wrapping around the periphery of the complex. Critically, in the complex, where N-Cdc37 binds to the kinase C-lobe via the His-Pro-Asn (HPN) motif at the N-terminal end of its coiled coil, Cdc37 appears to mimic the interactions that would usually be made by the kinase N-lobe in solution, namely the αC-β4 loop (Figure 1.14C). It has been proposed that these interactions are key in allowing the separation of the two kinase lobes (148). Meanwhile, the kinase N-lobe interacts with M/C-Cdc37. The complex structure also revealed new modes of Cdc37 binding to Hsp90; while bound at the MD of Hsp90, consistent with that found with nematode Cdc37, in this instance, Cdc37 contributes a β-strand to the β-sheet at the N-terminal side of the MD, and also makes a previously unseen interaction with the NTD of Hsp90 via N-Cdc37, similar to interactions made by p23 (148).

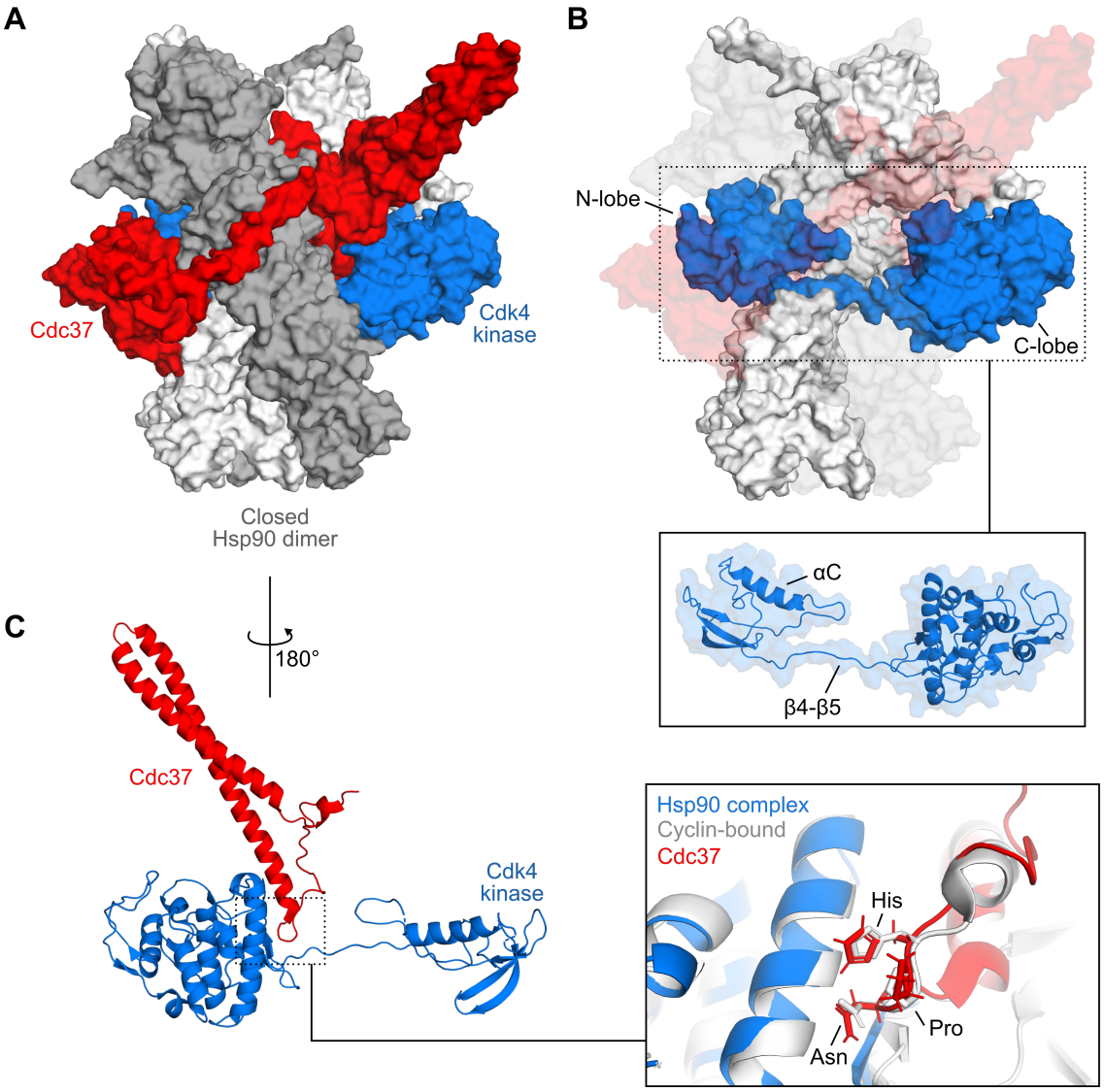


Figure 1.14. Hsp90-bound Cdk4 kinase is locally unfolded.
(A) Cryo-EM structure of a ternary complex of Cdk4 and Cdc37 bound to a closed dimer of Hsp90β (PDB entry 5FWL) (148). Proteins are shown with surface representation. (B) As in A, except Cdc37 and one copy of Hsp90 are transparent, allowing visualisation of the locally-unfolded kinase threading through the centre of the Hsp90 dimer. In the panel beneath, the kinase is shown with cartoon representation illustrating unfolding of strands β4 and β5 and separation of the two kinase lobes. (C) Cartoon representations of Cdk4 kinase and the amino-terminal domain (N-Cdc37) of Cdc37 showing interactions between the kinase C-lobe and the N-terminal coiled-coil of Cdc37. In the expanded panel, a structural alignment of Cdk4 bound to Hsp90 (blue) and D-type cyclin (PDB entry 2W9Z) (white) shows structural mimicry of the His-Pro-Asn (HPN) motif of the N-lobe of Cdk4 by the HPN motif of Cdc37.

The unfolded state of Cdk4 kinase observed in this ternary complex is undoubtedly inactive as separation of the two lobes breaks both the ATP binding cleft and hydrophobic spines of the kinase. While the unfolding and sequestering of partially unstable kinases in ternary complexes with Hsp90 may be useful to prevent their aggregation and degradation, and to promote their correct folding (multiple reports show that inhibition of Hsp90 typically leads to ubiquitination and degradation of its client kinases, suggesting that Hsp90 may 'rescue' many kinases from this fate (164,180–182)), it is currently disagreed why this would be useful for kinase function.

On one hand, it has been argued the role of Hsp90 in regulation of kinases may have evolved through the capacity of Hsp90 to 'buffer' the acquisition of gain-of-function mutations that, though often destabilising, confer activity advantages to kinases by allowing them to be more conformationally labile (183). However, it is unclear why this method of regulation will have evolved for some but not all kinases, and to have also evolved inconsistently among kinase families (170). Alternatively, a second hypothesis speculates that this interaction with kinases plays a distinct functional role in kinase regulation and that the unfolding induced by Hsp90 primes the kinase to be active upon its release (170). This proposal is supported by the observation that molecular dynamic simulations of EGFR show that the conformational transition between inactive and active states proceeds via a degree of local unfolding at the kinase hinge (184). The similar unfolding event induced by Hsp90 as observed in the cryo-EM structure of the Cdk4-bound complex may therefore enable a kinase to rapidly sample its active states following its dissociation from Hsp90. There is growing evidence that Hsp90 appears to have an active role in kinase activation: Raf-1, for example, is dependent on Cdc37 and Hsp90 for its activity (185), and v-src has been demonstrated to be activated by Hsp90 and Cdc37 *in vitro* (165). Interestingly, the Hsp90-Cdc37 system has also been implicated in regulating the formation of complexes of CDKs with their activator cyclins, possibly suggesting that Hsp90 may (at least in some instances) hold kinases in a primed but inhibited state for rapid activation under the correct cellular conditions (186).

1.3.8. A speculative Hsp90 cycle for the chaperoning of kinases

Building upon findings from previous studies, a speculative model for the chaperoning of kinases by Hsp90 has been proposed, where following sorting, recruitment and partial remodelling of kinases by Cdc37, kinases are transferred to and further remodelled by Hsp90. This begins with binding of Cdc37-kinase binary complexes at the NTD of Hsp90. Through undetermined means, structural rearrangements possibly involving NTD movement, could then lead to kinase lobe separation and relocation of Cdc37 to the Hsp90 MD. Subsequent processes such as the recruitment of other co-chaperones like Aha1 and the displacement of Cdc37, performed through a dynamic series of post-translational modifications (187–190), would allow the cycle to progress, ultimately leading to ATP hydrolysis and kinase release.

Obtaining structures of kinase-bound ternary complexes will be essential to confirm and illustrate that this is the case. Interestingly, a negative-stain EM map of Cdk4-bound ternary complex, reported a decade prior to the cryo-EM structure described above, proposes different modes of kinase and Cdc37 binding to Hsp90 (Figure 1.15) (148,191). Though possible that this may represent an alternative state of the kinase-bound complex that would be useful for building a model of the Hsp90 cycle with relation to kinases, the resolution of the complex is realistically insufficient to confidently place proteins in the map; higher-resolution structures will be necessary to confirm and embellish the structural and mechanistic details of the cycle.

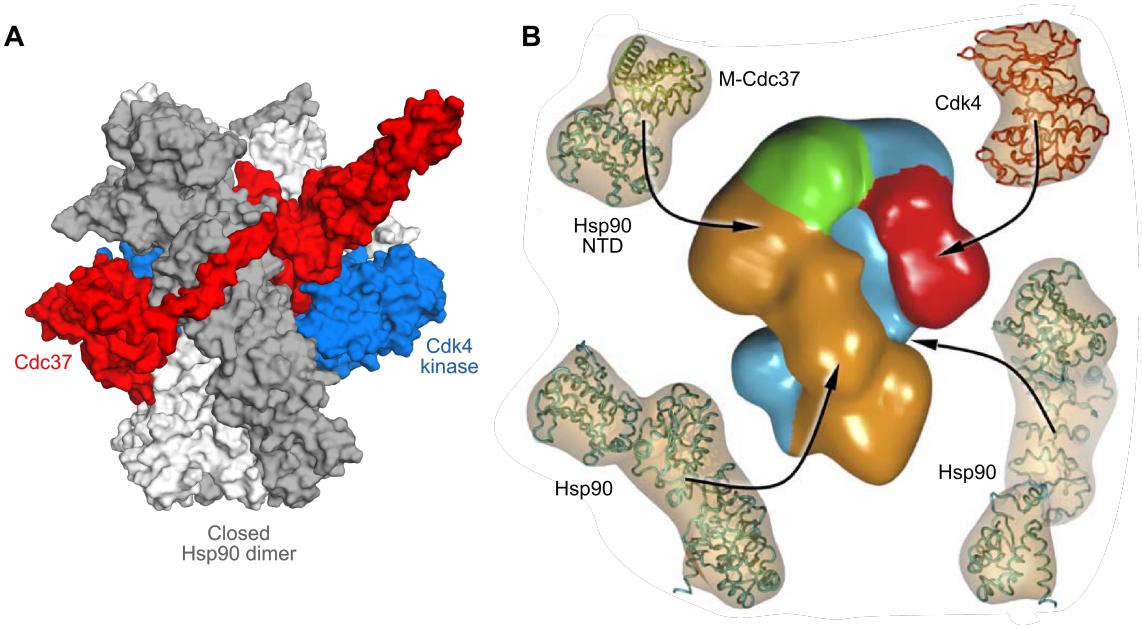


Figure 1.15. Two alternative Cdk4-bound ternary complexes.
(A) Cryo-EM structure of Cdk4-bound ternary complex (PDB entry 5FWL) (148). (B) Negative-stain EM model of an alternative Cdk4-bound ternary complex produced a decade earlier, adapted from Vaughan *et al.* (191).

1.3.9. Hsp90-kinase interactions as a therapeutic target

Given its role in the maintenance and regulation of kinases and its proposed ability to 'buffer' gain-of-function mutations from aggregation, Hsp90 has emerged as a desirable therapeutic target, being of particular interest for the treatment of cancer. In fact, it has been suggested that cancer cells become 'addicted' to Hsp90, a consequence of an increased protein load on the proteostasis network that arises from proteins gaining mutations that may lead to their destabilisation, as is observed with kinases (192,193). With respect to kinases, it is interesting to note that both Hsp90 and Cdc37 are overexpressed in cancer cells (194).

Efforts to target Hsp90 have largely focused on the development of inhibitors acting at its NTD to inhibit ATPase activity, binding at the nucleotide binding cleft. Historically, inhibitors have been based upon the natural product inhibitors of Hsp90, geldanamycin and radicicol (193); however, toxicities associated with these have driven interest in the development of novel synthetic Hsp90 inhibitors with different scaffolds such as the purine-based inhibitor PU-H71 (195). A number of clinical trials assessing inhibitors of Hsp90 are currently in progress (196).

Given the central role of Hsp90 in a host of cellular processes, off-target effects through inhibition of 'healthy' Hsp90 interactions are a concern for Hsp90-targeting therapies. While studies show that cancer cells are more sensitive to Hsp90 inhibitors than healthy cells (195), targeting specific Hsp90-client interactions may still be desirable. Curiously, it has already been noted that the inhibitors of Hsp90 can selectively interact with certain pools of client-bound Hsp90 complexes; for example, PU-H71 uniquely interacts with kinase-bound Hsp90 complexes (197,198), indicating that selective targeting of Hsp90-mediated kinase chaperoning may be possible. Targeting kinase-co-chaperone interactions may provide another kinase-selective approach, especially given that Cdc37 is able to stabilise some client kinases independently of Hsp90 (199). Alternatively, blocking kinase transfer to Hsp90 may present a further inhibition mode. Opportunities here are highlighted by the inhibitor celastrol that binds to the NTD of Hsp90 in a non-competitive manner to ATP, also blocking Cdc37-Hsp90 interactions (200), and ATP-competitive kinase inhibitors which stabilise kinases such that they no longer interact with Cdc37 (175). While these are encouraging, to best exploit kinase-Hsp90 interactions, a better structural understanding of the mechanisms of recruitment and chaperoning of kinases is required; these will help to both better understand established inhibition modes such that they can be improved upon and potentially also identify novel ones.

1.3.10. The role of Hsp90 in the function of RTKs and FGFRs

As for all membrane proteins, the biogenesis and folding of receptor tyrosine kinases takes place via the secretory pathway (201). In this pathway, nascent membrane proteins are inserted co-translationally into the membrane at the endoplasmic reticulum (ER) by the SEC translocon, then proteins fold and mature with the help of numerous chaperones (both ER-resident and cytosolic, including Hsp90) while progressing through the ER, to the Golgi apparatus and towards the plasma membrane (201). Proteins pass through several checkpoints during this process, ensuring that only mature, correctly folded proteins reach the cell surface. For many membrane proteins, and as is the case for RTKs, progress through this secretory pathway can be monitored by the glycosylation state of the receptor, as mature proteins are typically more heavily-glycosylated (201). Membrane proteins resident in the ER and Golgi also possess unique glycosylation signatures, as ER-resident glycosylated proteins are Endo-H sensitive whereas Golgi-located proteins are not (202). During expression of RTKs in mammalian systems, often both immature (ER-located) and mature (plasma membrane-located) can be observed.

As discussed previously, the dependency of kinases on Hsp90 for function differs even within kinase families, as is also observed among RTKs (141). This is best exemplified by the dependency of ErbB family members, which not only include both clients and non-clients, but also shows dependency differences among the clients themselves (203). For example, while ErbB-1 is dependent on Hsp90 during its maturation only (and is independent upon reaching the plasma membrane) (204,205), ErbB-2 depends on Hsp90 throughout its lifetime (205). This is demonstrated by a rapid degradation of both immature, ER-resident and mature forms of ErbB-2 upon treating cells with Hsp90 inhibitors (205).

In the FGFR family, only FGFR3 is considered to be a strong client of Hsp90, and it, like ErbB-2, is rapidly degraded upon treating cells with either 17-AAG or radicicol, both following ectopic expression in HEK293S cells or when endogenously expressed in RTT112 bladder cancer cells (155). Pull-down experiments indicate that Hsp90 interacts with both immature and mature forms of the receptor (155), suggesting that like ErbB-2, FGFR3 remains dependent on Hsp90 throughout its lifetime. Hsp90 thus likely plays a critical role in FGFR3 function, supported by the reduced cell viability of RTT112 cells (which are dependent on FGFR3 signalling activity for survival) following their treatment with Hsp90 inhibitors (155). Hsp90 also appears to play a key role in the function of other FGFR family members, most notably when they are present as fusions, as both FGFR1 (206) and FGFR2 (207) oncogenic fusions are also susceptible to Hsp90 inhibition.

1.4. Methodology introduction

In this thesis, the structure of membrane proteins and of protein complexes will be investigated using a variety of tools including electron microscopy (EM) and nuclear magnetic resonance (NMR) spectroscopy. For aid in the understanding of these data, these techniques and relevant background information are briefly introduced here.

1.4.1. Structural studies of membrane proteins

Despite membrane proteins accounting for approximately a third of the human proteome (208,209) and 60% of known drug targets for FDA-approved drugs (210), less than 2% of structures deposited to the Protein Data Bank (www.wwpdb.org, PDB) (211) are of this essential class of proteins. This severe underrepresentation in the PDB can be attributed to difficulties in obtaining samples of membrane proteins for structural studies due to problems faced with protein overexpression, solubilisation, purification and sample stability, among others.

1.4.1.1. Recombinant expression of mammalian membrane proteins

Expression of membrane proteins at a high enough level for structural studies, especially those of higher eukaryotes such as mammals, presents the first hurdle in many structural biology projects (212). Historically, structures of membrane proteins have been dominated by those of prokaryotic origin, in many cases of homologues of the 'more interesting' mammalian protein targets which were difficult to obtain. More recently, technical advances have resulted in a shift towards studying mammalian membrane proteins themselves.

While success has been found with recombinant expression in *Escherichia coli* (*E. coli*), more and more often, higher complexity expression hosts such as yeast, insect cells and mammalian tissue culture, and even cell-free expression systems, have been used to obtain membrane proteins of interest (213). Unsurprisingly, expression hosts most similar to the native host of the protein produce higher yields of functional protein (214). This has been attributed to these systems using post-translational modifications, protein biosynthesis machinery, and membrane environments closer to those found in the native organism (212–214). However, higher complexity expression systems come with drawbacks; insect cell and mammalian expression systems are more demanding and expensive, and both options present multiple expression hosts to consider (e.g. Sf9, Sf21 or Hi5 for insect cells, and HEK293, CHO, BHK-21 or COS-7 for mammalian tissue culture) (213).

When choosing an expression host, a compromise is often made between the desire to express in a close-to-native system and doing so in a time and cost-effective manner. In this work, the insect cell-based baculovirus expression vector system (BEVS) was

chosen. While not able to fulfil the full repertoire of post-translational modifications of mammalian cells, insect cells offer high-mannose type N-linked glycosylation (215) and are generally considered a cheaper and an easier to manipulate alternative to their mammalian counterparts. They have been successfully used to express a wide range of mammalian membrane proteins including G-protein coupled receptors (GPCRs), ion channels, transporters, and full-length RTKs and their subdomains, including FGFR family members (20,47,223–225,126,216–222).

1.4.1.2. Purification of membrane proteins

Following successful recombinant expression, researchers next face the hurdle of being able to isolate their membrane protein of interest in a pure yet functionally active form. Purification of membrane proteins first requires their extraction from the membrane, a process which often leads to high protein losses through protein denaturation and aggregation (212).

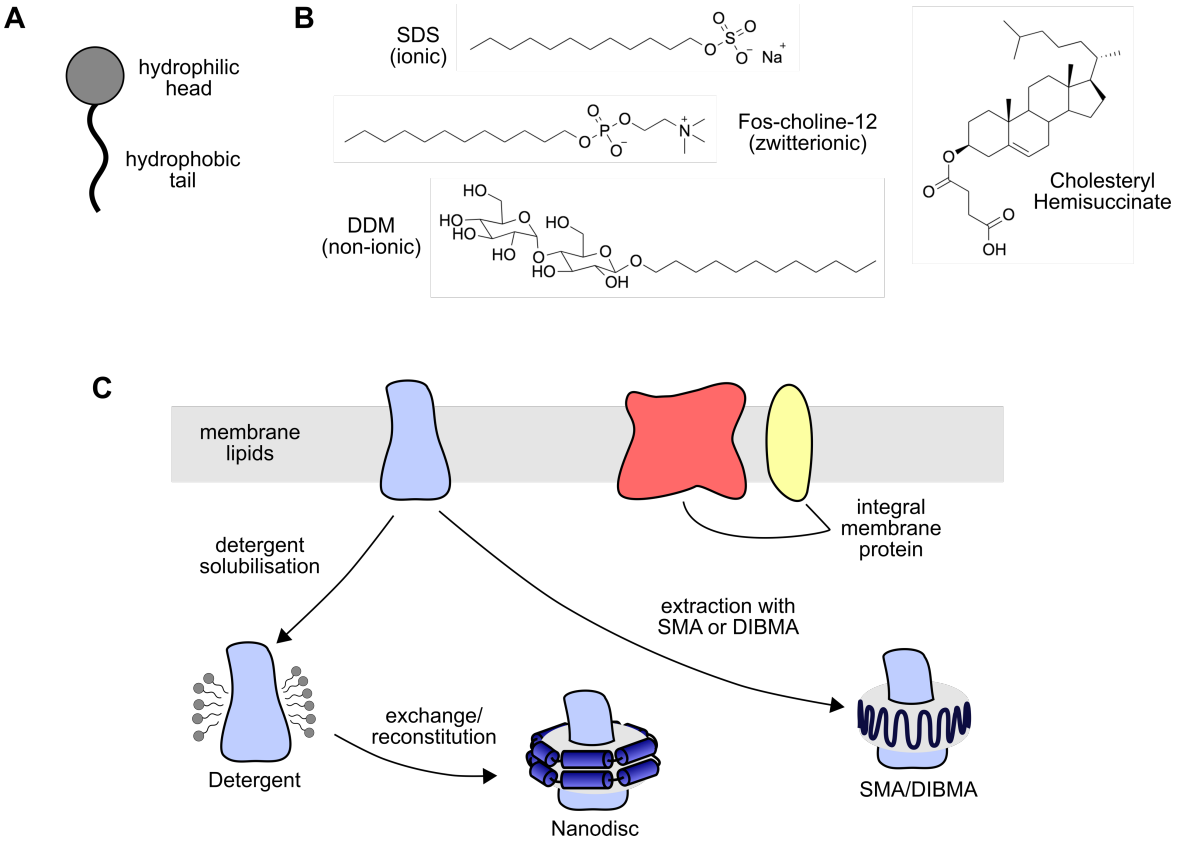


Figure 1.16. Solubilisation of membrane proteins.
(A) A schematic of a detergent molecule. (B) Chemical structures of example ionic, zwitterionic and non-ionic detergents, and the cholesterol mimic cholesteryl hemisuccinate. (C) Illustration of membrane protein solubilisation by detergents and their reconstitution into nanodiscs versus direct extraction from the membrane into disc-like structures by the co-polymers SMA and DIBMA.

Traditionally, detergents have been the primary choice for the solubilisation of membrane proteins. Although differing in chemical structures, all detergents are amphipathic, containing a hydrophilic head group and hydrophobic tail (Figure 1.16A). These properties enable detergents to disrupt membranes and extract membrane proteins by protecting the hydrophobic surfaces of their membrane-embedded regions from the aqueous environment, mimicking the lipid membrane (226). To achieve this effect while maintaining proteins in protein-detergent micelles, detergents are used above their critical micelle concentration (CMC), i.e. the concentration required for detergents to self-assemble into micelles. No two detergents are equal; in fact, several families of detergents have been developed, each with different chemical and physical properties including 'harshness', detergent CMC, molecular weight and micelle size (226,227). In order to extract proteins from the membrane, detergents are required to be relatively 'harsh', meaning that they are able to replace the lipids within which the proteins are embedded. However, if a detergent is too harsh, one risks disrupting the protein's structure causing its inactivation. Furthermore, lipids which may be essential for protein activity or stability can be lost. Thus, a balance needs to be found between detergent 'harshness' which improves yields from membrane protein extraction and the desire to obtain functional protein, often instead requiring milder options. Detergents can be classified into three categories: ionic, non-ionic and zwitterionic (Figure 1.16B) (226). Of these, ionic detergents are considered the harshest and are generally not used in structural studies as they are able to interfere with protein-protein interactions, often resulting in protein denaturation; sodium dodecyl sulfate (SDS), used to denature proteins for separation in gel electrophoresis, is an anionic detergent, for example (228). Instead, non-ionic detergents are preferred. Zwitterioinic detergents lie between ionic and non-ionic detergents on the 'harshness' scale and are of some utility, having been used fairly commonly in NMR studies (229).

n-Dodecyl-β-D-maltoside (DDM) is the most commonly used detergent in membrane protein structural biology and is the standard starting choice for solubilisation and crystallisation of membrane proteins. This is in part due to its relatively low CMC of 0.0087%, requiring less detergent during sample preparation than its counterparts (thus being more economical). Moreover, the CMC of DDM is also sufficiently high that it can be removed and exchanged with other detergents if desired (230). This can be attractive in cases where the large micelle size of DDM (~70 kDa) is found to be inhibitory to structure determination, or where a membrane protein requires reconstitution into an alternative membrane environment for functional studies. While DDM is often a good starting point for solubilisation and purification, there are no established rules dictating which detergent will be a good choice for a particular membrane protein of interest; instead, detergent screens are necessary to identify the most promising candidates in

each case. Moreover, it is often the case that the most ideal detergents for solubilisation, purification, and structure determination are not the same, necessitating detergent exchange between stages, though these exchange processes are often accompanied by protein loss.

Recently, though DDM remains the primary choice for membrane protein solubilisation and purification, alternative detergents such as lauryl maltose neopentyl glycol (LMNG), decyl maltose neopentyl glycol (DMNG), digitonin, and its substitute glycol-dicosgenin (GDN) have emerged as popular choices for structure determination (231). These latter two have only gained traction in structural studies by EM, highlighting the emergence of preferences in detergent choice depending on the method of structure determination. During the solubilisation of eukaryotic membrane proteins, the cholesterol derivative cholesteryl hemisuccinate (CHS) is frequently added. This inclusion is based upon the findings that the stability and activity of isolated eukaryotic membrane proteins is substantially improved in the presence of CHS (232–234), generally thought to derive from the recapitulation of a more membrane-like lipidic environment than is possible with detergents alone.

As the importance of the lipids on the structure and function of membrane proteins is ever-more realised (235), there have been increased efforts to extract and reconstitute isolated membrane proteins in more native-like environments and in alternatives to detergents. Examples of this can be found in the use of bicelles (a mixture of lipid and detergent containing a membrane-like lipid bilayer) (236) and the advent of nanodiscs (237). These lipid-nanoparticles are formed from an apolipoprotein component whose amphipathic helices encompass and encapsulate a lipid bilayer into discoid-like particles (Figure 1.16C) (237). Nanodiscs have also encouraged the development of alternative surfactants for protein solubilisation such as amphipols (238), peptidiscs (239), Saposin-lipoprotein (Salipro) nanoparticles (240), and the co-polymer styrene-maleic acid (SMA) and its derivatives such as diisobutylene-maleic acid (DIBMA) (241–244). The ability of the co-polymers SMA and DIBMA, and more recently Salipro, to extract membrane proteins directly from the lipid membrane, has enabled proteins to be studied in an environment that retains lipids from the membrane that the protein was embedded in (Figure 1.16C).

1.4.1.3. Membrane protein structural determination

After purification of the membrane protein of interest in a functional form, structural determination presents a further hurdle to be tacked. Historically, X-ray crystallography has been the choice for membrane protein structure determination. However, structure solution by this method is by no means easy nor routine, as, like with all crystallography projects, obtaining crystals remains a significant bottleneck. In fact, crystallisation of membrane proteins is generally more difficult than of their soluble counterparts, in part due to the stability problems mentioned previously, and in many cases due to a reduced ability to form crystal contacts required for crystal formation. Where success is achieved, crystals obtained are also often smaller than those from soluble proteins; membrane protein structure determination has thus benefited substantially from developments in synchrotron radiation facilities, robotics and micro-focus beamlines for collection of X-ray diffraction data (212). Many successful membrane protein crystallisation projects have required extensive efforts to improve protein stability through detergent screening, using ligand additives, using fusion partners and the introduction of stabilising mutations, and through the aim to increase the number of hydrophilic surfaces for crystal contacts by various means including using antibodies/nanobodies, fusion proteins, and through complex formation. Though also shown elsewhere, this is best exemplified by the various efforts used to solve structures of GPCRs in the past two decades (245–247).

In recent years, cryo-EM has established itself as major technique to challenge and complement X-ray crystallography in protein structure determination. Previously limited to low resolution information, the ‘resolution revolution’ driven by technical developments have led to an ever-increasing role of cryo-EM in membrane protein structural biology (248,249). In fact, in 2018, marginally more *unique* membrane protein structures were solved by cryo-EM than by crystallography for the first time (54% versus 45%) (121,231). Structure determination by cryo-EM is particularly attractive in cases where the amounts of purified proteins available for structural studies are limited and insufficient for crystallography, as can be the case for many membrane proteins. Moreover, it allows visualisation of protein structure in a more ‘solution-like’ state and is not hampered by the requirement to obtain protein crystals, a feat that can be challenging if the protein of interest harbours flexible regions or few opportunities for the formation of crystal contacts. Despite these benefits, cryo-EM is subject to its own limitations, as will be discussed in the following subchapter.

1.4.2. Structure determination by single-particle electron microscopy

While EM was often referred to as ‘blob-ology’ in the past, cryo-EM datasets now regularly produce structures at resolutions between 3-5 Å. This leap in resolution can be largely attributed to developments in detector instrumentation with the introduction of direct electron detectors, complemented by development of stable high-energy electron sources and microscope optics, improved automation and computing power, and sophisticated image processing programmes (248,250). While these improvements have also benefited cryo-electron tomography, only single-particle cryo-EM will be discussed.

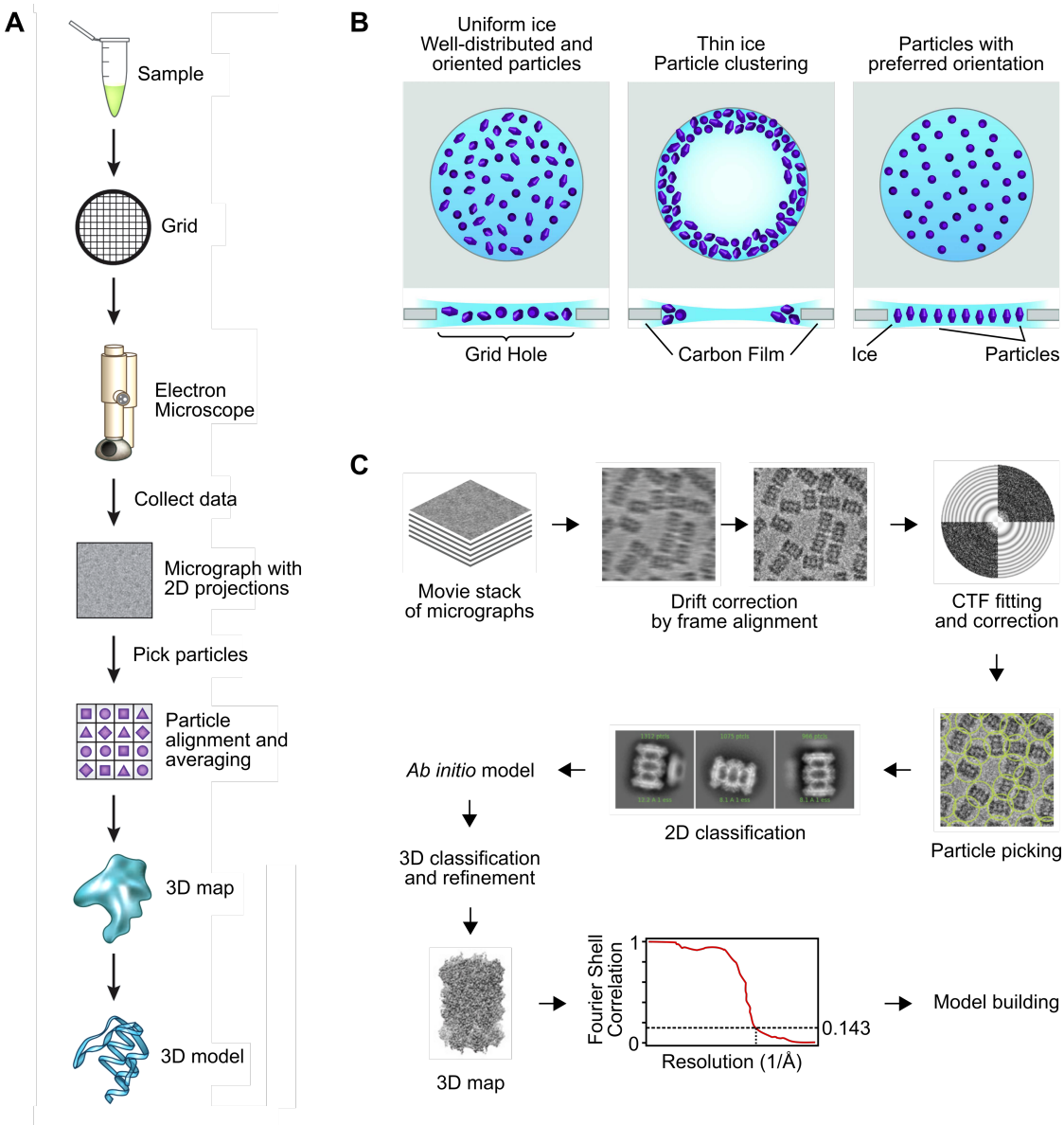


Figure 1.17. Structure determination by electron microscopy.
(A) An illustrative overview of the workflow of protein structure determination by electron microscopy. Adapted from Doerr (251). (B) Illustration of some of the possible outcomes from sample vitrification in cryo-EM grid preparation and problems that can be faced. Adapted from Drulyte *et al*. (252). (C) The workflow of cryo-EM image processing and structure determination in more detail. Adapted from Tegunov & Cramer (253).

1.4.2.1. Anatomy of an EM data collection

Structure determination by single-particle EM begins by measuring how electrons interact with a protein sample placed in a transmission electron microscope. Occurring under vacuum, electrons originating from an electron source at the top of the microscope are focused onto the sample of interest, then electron scattering by the sample measured as an image (called a micrograph) on a detector beneath the sample at the microscope base. A micrograph is effectively a 2-dimensional (2D) projection image of the 3-dimensional (3D) sample. Using image processing and reconstruction software, these 2D projections can be reconstructed into a 3D volume of the sample (Figure 1.17A).

1.4.2.2. EM sample preparation

To image a sample by electron microscopy, it needs to be prepared in a way that allows it to survive unperturbed under vacuum. This can be achieved in a number of ways including chemical fixation, negative-staining with high molecular weight stains, and, as in the case of cryo-EM, sample vitrification (254,255). To be imaged, samples are prepared on EM grids. These are typically composed of a carbon film (either continuous or holey) on a copper (or gold) grid support, though alternatives are available (254–256). As these grids are hydrophobic, in order to apply a sample for grid preparation, the surface of these grids are first made to be hydrophilic, either through a process known as glow-discharging or through the use of a plasma cleaner (257). To prepare a sample for negative-stain EM, protein samples are first adsorbed onto the continuous carbon film of the glow-discharged grid, then following blotting of excess liquid and washing if desired, samples are stained with a heavy metal salt solution and dried. Multiple heavy metal stains are available, each with advantages and drawbacks, but uranyl formate and uranyl acetate stains are often used as standard (256). In addition to making the sample suitable for imaging, heavy metal staining also introduces high contrast, permitting easy and rapid visualisation of the sample (though indirectly, as it is the stain rather than the sample that is observed). This enables negative-stain EM to play a critical role in structure determination as a means to assess sample integrity and heterogeneity in ways which biochemical techniques such as SDS-PAGE and size exclusion chromatography (SEC) cannot (254). Unfortunately, contrast enhancement by negative-staining comes at the expense of resolution, effectively limiting 3D reconstructions of samples to the 20 Å range (256); high-resolution studies instead require the vitrification of samples.

These days, cryo-EM grid preparation occurs in a semi-automated manner with specialised grid-plunging devices such as Thermo Fisher Scientific's Vitrobot Mark IV and Leica's EM GP2 being commercially available. Grid preparation is similar to that in negative-stain except holey rather than continuous carbon films are used and grids are

plunge-frozen in a cryogen such as liquid ethane to obtain the biological sample suspended in a thin layer of vitreous ice (generally between ~10 and 90 nm thick) (Figure 1.17B) (252,255). As ice thickness correlates inversely with image contrast and can limit the resolutions that are obtainable from reconstructions, ice layers should be as thin as possible (258); an ideal ice layer is just thick enough that it can support a single layer of particles (258). After grid preparation, samples are kept at liquid nitrogen temperature for storage and data collection. Ideally, biological samples are frozen as well-spaced particles in multiple orientations in a uniform-thickness ice layer between the carbon film supports (Figure 1.17B). In practice, this can be difficult to achieve reproducibly or at all, with issues such as the ice being too thick, too thin, or inconsistent in thickness, particle clustering and preferred particle orientation, sample denaturation and others being observed (252). Often these indicate that better sample preparation prior to grid making is required, though grid properties can also be adapted to improve the prepared grids (252). Recently, the air-water interface has gained a lot of attention as an effect which, exacerbated by grid blotting prior to vitrification, can cause preferred orientations and denaturation of samples by presenting an exceptionally hydrophobic surface for proteins to interact with during grid preparation. This effect can mean that samples that may otherwise look promising by negative-stain EM are unsuitable for cryo-EM, preventing high-resolution structure determination (259,260). Several approaches have been suggested and shown to be useful in minimising the effect of the air-water interface, including: chemical crosslinking to improve sample stability (261–263), attempts to block the air-water interface using surfactants such as detergents (264,265), using specialised grids backed with a graphene layer to adsorb biological samples onto (by analogy with what occurs in negative staining) (259), and alternative rapid-freezing techniques, namely sample spraying and the ‘writing’ of samples onto grids (266–269). However, while each of these have been shown to be useful in particular cases, none are a ‘one glove fits all’ solution.

In this thesis, chemical crosslinking was used during sample preparation for EM. This approach has found a particular use in the study of macromolecular complexes which are unstable and ‘fall apart’ during EM grid preparation, and can be achieved in a number of ways including ‘in-solution’ crosslinking following complex purification and ‘on-column’ crosslinking during purification by SEC. An alternative method which has gained traction and produced numerous low-to-intermediate-resolution structures, particularly of large macromolecular complexes, is a process called ‘Gradient Fixation’ (GraFix) (261,270–274). This sample preparation method has also been used on difficult-to-handle samples such as full-length RTKs like the KIT receptor and PDGFRβ, in these cases being highlighted as a necessary step to obtain stable and monodisperse samples for negative-stain EM (125,126). In this method, complexes are simultaneously purified and mildly

chemically-crosslinked by ultracentrifugation in a density gradient of either glycerol or sucrose containing a crosslinker such as glutaraldehyde. Following ultracentrifugation, a sample which contains intra-complex crosslinks is obtained following fractionation of the gradient. This sample can be used directly for negative-stain EM or for cryo-EM following buffer exchange. Despite its uses, chemical crosslinking may introduce artefacts (such as through capture of rare and otherwise conformationally-unlikely states, or by stabilisation of false protein-protein contacts as a consequence of high protein concentration (275)) and thus should be used with caution.

1.4.2.3. Data collection and image processing

After grid optimisation, an EM dataset in the form of many micrographs is collected; the processing of these micrograph images is necessary to obtain a 3D reconstruction of the biological sample (Figure 1.17C). Following the advent of direct electron detectors, cryo-EM micrographs are typically collected in so-called 'movie mode' where the total dose applied to the sample (typically around 50 $e^-/Å^2$) is split into many fractions and a stack of micrographs image frames is acquired (254). Crucially, this enables drift correction of the sample (whether induced by the beam or other factors), overcoming the blurring of images that limits the resolution attainable from these reconstructions due to the loss of high-resolution details. Following drift correction, micrographs are contrast transfer function (CTF)-corrected, a process that corrects for image distortions introduced into micrographs due to the use of defocus in data collection. This is necessary because biological molecules scatter electrons very weakly, thus micrographs collected at true focus have very little contrast between the sample and the empty ice. While the use of greater defocus during sample imaging introduces contrast and allows the sample to be observed more easily, micrographs recorded with greater defocus contain lower spatial frequency information (i.e. lower resolution information) than those recorded closer to true focus. Partly due to this reason, micrographs are typically recorded over a range of defocus values such that the dataset used for 3D reconstruction contains both high- and low-resolution information. Alternatively, contrast can be introduced at true focus by using newly-developed phase plates, a strategy that is particularly useful for samples <200 kDa in mass (276).

Reconstruction begins through particle picking of micrographs. Ideally, these particles correspond to a homogeneous population of the sample in multiple orientations in the vitreous ice, representing different views of the sample. Particle picking can be completed both manually and automatically using a number of software packages (254,277–281). In practice, the approach used is chosen on a case by case basis and often a semi-automated method is appropriate. In this scenario, particles are first manually picked, then these particles used to generate templates for auto-picking

algorithms. While laborious, manual particle picking can be useful as a means to assess the quality of the data and is helpful when fully automated auto-picking programmes fail. However, it is important to note that this semi-automated procedure is also prone to user- and template-introduced biases. Particularly, users may miss uncommon views of the sample during manual picking, and through the use of templates, this bias can be propagated through the auto-picking procedure. A further concern of template-based auto-picking is the risk of picking noise and observing the so-called 'Einstein from noise' effect (282). To try and avoid this, templates used for auto-picking are typically low-pass filtered to remove high-frequency information. In recent years, the reliability of fully automated auto-picking algorithms has taken great strides and to such an extent that datasets can now in many cases be routinely fully auto-picked and processed 'on the fly' (280,283–285) with little user intervention.

After particles have been picked, they are next classified into 2D classes of similarly oriented particles, typically by maximum likelihood algorithms (278,286–288), and averaged together. The obtained 2D class averages are good gauges of data quality and can indicate the likelihood of high-resolution 3D reconstruction, as secondary structure elements are visible in the classes in these cases. Following 2D classification (often in multiple rounds to remove 'poor' particles that contribute to poorly-defined classes), 'good' particle stacks are used for *ab initio* determination of an initial 3D model that is subsequently used for 3D classification of these particles. This is followed by structure refinement to the final map (Figure 1.17C). Two-dimensional projections are reconstructed into a three-dimensional map using the 'projection-slice theorem' (289); effectively, as the Fourier transform of a 2D projection corresponds to a slice of the Fourier transform of an object in 3D space, the back-projection of many 2D projections with different but known views of an object generates a 3D volume of that object. Multiple competing software packages now exist for EM image processing, including but not limited to: RELION (278), CryoSPARC (290), WARP (253), and others.

Resolution estimates in EM are determined by the Fourier Shell Correlation (FSC) curve, a calculation of correlation between two maps each corresponding to half of the data used for reconstruction. As high-resolution cryo-EM is still in its infancy compared to X-ray crystallography, there is still some dispute regarding how best to estimate resolution from FSC curves and how to measure the quality of fit between the refined reconstruction and the acquired data (291); however, the so-called 'gold-standard' approach has emerged as the current common choice, where a threshold of 0.143 is used with a FSC curve of two independently-refined half-datasets (292). Regardless of these estimates, resolution is not uniform across the reconstructed 3D map, with the cores of structures typically showing higher resolution than the structure as a whole.

1.4.2.4. Challenges in EM

While modern electron microscopes are capable of collecting data to obtain reconstructions better than 2 Å in resolution, reconstructions are often much poorer (254). In fact, although the current resolution record from single-particle EM stands at 1.54 Å as achieved for apoferritin in 2019 (EM Databank accession code: EMDB-9865), the average resolution of a structure deposited to the PDB was only 6.6 Å in 2018 (293). This discrepancy can in part be attributed to features of each particular sample of interest that can limit the resolutions obtainable by EM (e.g. sample mass, symmetry and heterogeneity). While proteins smaller than 100 kDa have been solved to subnanometer resolution (294–296), protein molecular weight and particle size still pose challenges for structural determination in EM, and only 11% of entries to the Electron Microscopy Data Bank (EMDB) are <250 kDa in mass (297). Sample heterogeneity can also significantly hamper the resolution of EM reconstructions; while EM is unique in that it is able to handle some sample heterogeneity through computational means, enabling multiple conformational states to be extracted from a single dataset to describe the structural dynamics of a system (298), sometimes, heterogeneity can be too complex. In these cases, reconstructions are limited to lower resolutions as insufficient data relating to each state of the mixture is obtainable. Symmetry (or lack thereof) of a sample also has a considerable impact on the resolution that can be achieved. Samples possessing symmetry are more likely to return higher resolution reconstructions than those that do not, as imposing this symmetry relationship during processing effectively increases the dataset size and thus artificially improves signal to noise. However, as imposing symmetry can also very easily introduce artefacts, doing so is generally inadvisable, especially if there is any reason to believe that the sample contains asymmetric features.

While we have recently seen a boom in structure determination of membrane proteins by cryo-EM, doing so is by no means trivial. In addition to those discussed, membrane protein samples pose additional challenges in EM not met elsewhere. Notably, the material that the membrane protein sample is solubilised in (often detergent, but increasingly nanodisc or nanodisc-like materials) can contribute significantly to the size and shape of particles in electron micrographs. This is a particular problem for smaller membrane proteins which are largely embedded in the membrane environment and have few soluble domains; consequently, particle averaging during data processing can be dominated by the detergent micelle, limiting the ability to resolve structural features of the protein. In a similar way to those used in crystallography, in these instances, researchers have turned to a toolbox of options to increase the mass of the protein with an emphasis on enhancing features outside of the membrane(-like) environment, such

as through complex formation and the use of antigen-binding fragments or nanobodies, among others (299,300).

1.4.3. Structural studies by NMR spectroscopy

In this thesis, NMR spectroscopy was used to study protein interactions, ligand-protein interactions and the effect of point mutations on protein structure. To help understanding of this data, the principles of NMR (301–304) and the techniques used in this thesis will be briefly described here.

1.4.3.1. Basic principles of NMR

Many nuclei have nuclear spin, occurring due to unpaired protons and neutrons, and can be thought of as tiny magnets spinning on a rotation axis (Figure 1.18A). In absence of a magnetic field, these 'tiny magnets' (or magnetic dipole moments) are in random orientations. However, when placed in an external magnetic field (B_0), in the instance of spin-½ nuclei (e.g. ^{1}H, ^{13}C, ^{15}N and ^{31}P), they align in one of two preferred orientations: aligned with or aligned against this field (Figure 1.18A). These two orientations can be thought of as two energy states, with the difference in the population of magnetic moments across these states described by the Boltzmann distribution. In NMR experiments, nuclei are excited from the lower energy state to the upper energy state through irradiating the sample with radiofrequency waves. Following irradiation and excitation, nuclei return to thermal equilibrium through a process called relaxation. In essence, the change in the population difference between these two states is detected by the NMR spectrometer, as is discussed later.

The energy required to induce 'flipping' events corresponds to the energy difference between the upper and lower energy states, and can be described by:

$$\Delta \mathbf{E} = \boldsymbol{\gamma} \hbar \mathbf{B_0} \qquad (1.1)$$

The energy required is thus proportional to both the strength of the external static magnetic field (B_0) and the gyromagnetic ratio (γ) of the nucleus, a constant that reports on the strength of the magnetic moment. Different nuclei have different gyromagnetic ratios, a factor which affects the sensitivity of NMR measurements and is also exploited in multidimensional NMR. The sensitivity of NMR spectroscopy is inherently low, a consequence of the very small population difference between the upper and lower energy states, and the 'weakness' of the magnetic moment. While the population difference of magnetic moments can be maximised using nuclei with large gyromagnetic ratios (i.e. protons, ^{1}H) and magnets with high field strengths, this difference is still extremely small; the percentage population difference of protons present in a 1000 MHz NMR spectrometer at room temperature is only 0.008% (305). Consequently, NMR

samples are generally required to be of a relatively high concentration, and experiments comprise multiple repeated scans to improve signal to noise.

1.4.3.2. Relaxation in NMR

To explain signal detection and relaxation processes in NMR, it can be useful to describe magnetisation moments as vectors precessing about the axis of the applied magnetic field (the z axis), at a frequency proportional to γB_0. In this simplified model, due to the population difference between the two orientations of magnetic moments, the nuclei of the sample with equivalent spins can be considered as a bulk net magnetisation that is aligned with the external magnetic field (Figure 1.18B). It is this net magnetisation that is manipulated and observed during NMR experiments.

NMR signals are detected as a free-induction decay (FID) that corresponds to precession of nuclei at the Larmor frequency. This FID is observed in the x-y (transverse) plane and reports on an oscillating voltage that is induced due to precession of net magnetisation and is detected by a receiver coil in the NMR spectrometer. An NMR spectrum is obtained through Fourier transform of this FID. In a typical NMR experiment, so that it can be observed, net magnetisation is transferred to the transverse plane by applying a radiofrequency pulse resonant with the frequency of precession for a defined amount of time. Typically, radiofrequency pulses are short in length (usually microseconds) but high in power, allowing excitation of all nuclei of interest equally. The simplest way to excite nuclei such that net magnetisation is moved to the transverse plane is by using a 90° pulse (Figure 1.18B). As introduced previously, following their excitation, nuclei return to thermal equilibrium through a process called relaxation. In this instance, relaxation can be described as the processes by which net magnetisation is lost in the transverse plane (T_2 relaxation), and is returned to the z axis, restoring equilibrium magnetisation (T_1 relaxation) (Figure 1.18B). This is achieved by exchange of energy at the frequency of precession (the Larmor frequency).

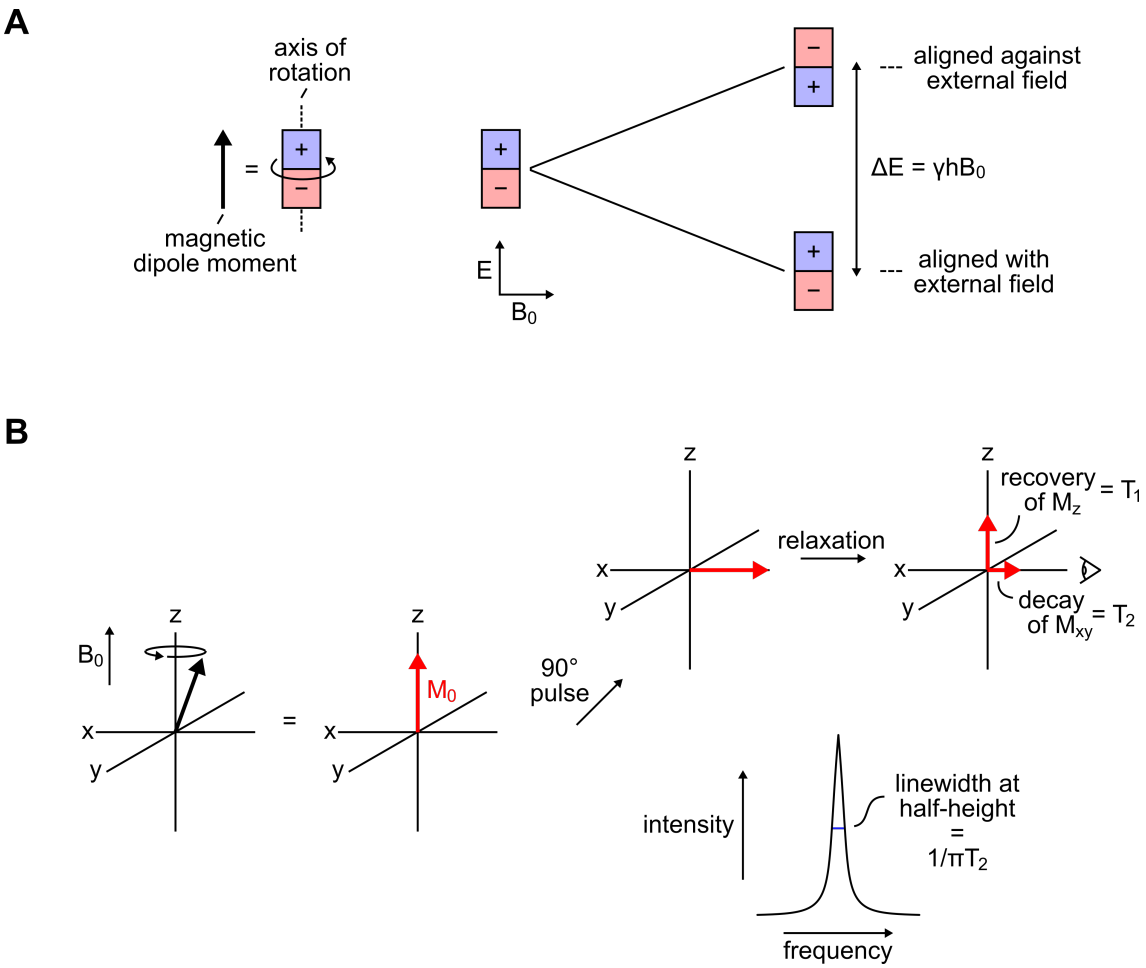


Figure 1.18. Principles of NMR spectroscopy.
(A) Schematic illustration of the behaviour of magnetic dipole moments in the presence of an external magnetic field (B_0), where they can align with or against the magnetic field. The difference in energy between these two states (ΔE) is dependent on the strength of the external field and the gyromagnetic ratio of the nucleus (γ). (B) The vector model of net magnetisation (M_0) illustrating its transfer to the transverse plane by sample irradiation with radiofrequency pulses, and T_1 and T_2 relaxation processes. The relationship between T_2 relaxation time and peak linewidth in an NMR spectrum is illustrated beneath. M_z and M_{xy} refer to net magnetisation in the z (longitudal) and x-y (transverse) planes.

As restoring equilibrium magnetisation in the z axis concurrently reduces magnetisation in the transverse plane, T_2 rate of relaxation is always smaller than or equal to T_1; transverse T_2 relaxation can be smaller than longitudal T_1 relaxation as it also reports on the loss of coherence of magnetic moments in the transverse plane, a consequence of nuclei experiencing imperfect magnetic fields which are different for individual spins of the sample ensemble. These imperfections can be a consequence of B_0, or due to the local magnetic fields of neighbouring nuclei. For small molecules in solution, rapid tumbling effectively averages these contributions to relaxation (meaning T_2 is very large and relaxation is dominated by T_1 processes); however, larger molecules (such as proteins) that tumble more slowly in solution experience imperfect averaging, resulting

in more rapid transverse relaxation and a smaller T_2 value. The transverse relaxation of a system is important as the linewidth of peaks in an NMR spectrum is inversely related to T_2 (Figure 1.18B). This places a size limitation on solution NMR studies, as large molecules with slow tumbling have broad peaks. With very large molecules, peaks can be broadened beyond detection.

1.4.3.3. Protein NMR

NMR permits structural studies of proteins in solution, lending itself as an excellent tool to study protein dynamics in ways that other structural techniques such as X-ray crystallography and electron microscopy cannot. While NMR can be used to solve protein structure, it is typically reserved for smaller protein systems (306) is not used nor discussed here. As proteins are abundant in protons, one-dimensional ^{1}H NMR spectra can conveniently inform on protein structure. However, for the same reason, these spectra are very complex with many overlapping peaks, each corresponding to different ^{1}H nuclei (Figure 1.19A). While generally not used in protein NMR, 1D spectra of proteins are useful to help describe two properties of NMR peaks: intensity and chemical shift. The intensity (or more accurately the area) of a peak corresponds to the number of nuclei contributing to this peak. Chemical shift refers to the position of a peak on the spectrum with respect to a reference, and arises due to the shielding of nuclei from the magnetic field by electrons. When protons are neighboured by electron withdrawing groups such as oxygen, nitrogen and aromatic rings, they experience less electron shielding and have a more positive chemical shift, expressed in parts per million (ppm) of the static external magnetic field (Figure 1.19A). The position of a peak on an NMR spectrum thus reports on the local chemical environment of that particular nucleus; in protein NMR, peak position therefore (at least partly) reports on protein structure.

To simplify NMR spectra of proteins so that they are more interpretable and informative, two-dimensional heteronuclear single quantum coherence (HSQC) spectra describing correlation between protons and a heteronuclear atom (typically ^{15}N or ^{13}C) are acquired. In these spectra, every peak corresponds to a unique proton directly bonded to a heteronucleus; in the case of ^{1}H-^{15}N HSQC experiments, peaks generally correspond to the backbone amide group of peptide bonds. With the exception of proline residues, a peak for every residue should be observable, providing a ‘fingerprint’ of the protein. Residues with side chains containing amine groups (e.g. glutamine and asparagine) have additional peaks. As before, the intensity (or rather area) of each peak indicates the population of nuclei pertaining to this peak, and chemical shift reports on the local chemical environment of the amide bond nuclei that the peak corresponds to. This is useful in interpreting protein structure, most simply illustrated by the differences in peak dispersion in ^{1}H-^{15}N HSQC spectra of a protein in its folded and unfolded state (Figure

1.19B). If the identity of each peak is known (i.e. they are assigned to residues), finer structural information can also be obtained.

By using a technique called chemical shift perturbation (CSP) mapping, NMR is particularly useful in providing structural insights about binding interactions between proteins and other molecules (e.g. small molecules, proteins etc.). In these studies, ^{1}H-^{15}N HSQC spectra can be acquired in the absence and presence of potential binding partners, then any changes in peak movement and intensity (a consequence of the changing chemical environments of the nuclei they report on), can be observed and quantified. If peaks are assigned, those which are perturbed can be mapped onto protein structures (if known) to define binding interfaces, or to identify allosteric structural effects induced elsewhere in the protein as a consequence of partner binding. This approach is used frequently in the screening of small molecule binding to target proteins (307). The way in which a peak is perturbed is also informative, reporting on the rate of exchange between the bound and unbound forms of the protein. Exchange can be described as occurring in fast, intermediate or slow regimes (a rate of exchange measured with respect to the difference in chemical shift (in Hz) between bound and unbound states), and can be identified by the characteristics of perturbed peaks, both in their changes in position and their intensity during a titration with a binder (Figure 1.19C). Typically, binding events with strong interactions (i.e. with binding affinities in the low micromolar range or lower) occur in the slow exchange regime, while weak binding events occur in the fast exchange regime (307). These exchange phenomena can also be extended to other exchange processes, such as those between a folded and unfolded state, or between two conformational states sampled by a protein.

1.4.3.4. Sample preparation for protein NMR

As ^{15}N has a very low natural abundance (~0.3%), to obtain ^{1}H-^{15}N HSQC spectra with good sensitivity, protein samples need to be uniformly labelled with ^{15}N for NMR studies. Though also NMR-active, as ^{14}N nuclei are quadrupolar (i.e. they have a spin number greater than ½), they are generally not used in protein NMR as they produce very broad signals. To achieve uniform ^{15}N labelling, proteins of interest are typically recombinantly expressed in *E. coli* cultures that are grown in minimal media supplemented with ^{15}N-labelled ammonium chloride as the sole nitrogen source. Using the flexibility of minimal media and knowledge of *E. coli* amino acid metabolic pathways, expression media can also be adapted to allow for labelling strategies generating triple (^{1}H/^{13}C/^{15}N)-labelled proteins for triple-resonance NMR experiments (used in residue assignment of backbone amides) or to introduce ^{1}H/^{13}C labelling of amino acid side chain methyl groups for use in methyl-NMR. This latter labelling strategy proves to be particularly useful for large protein systems which are intractable to NMR with traditional ^{15}N-labelling due to

relaxation-induced broadening effects. Though developments such as sample deuteration (308) and transverse relaxation optimised spectroscopy (TROSY) detection methods (309) have allowed larger molecular weight systems up to 70-100 kDa to be studied (306), methyl-NMR can be used to study assemblies of up to 1 MDa in mass (310).

Expression media can also be adapted to selectively label only select amino acids (e.g. Leu residues (or others) only) (311). Alternatively, proteins can also be 'selectively unlabelled' in which they are uniformly-labelled except for select amino acids that contain NMR-'inactive' nuclei (e.g. ^{14}N-labelled Leu residues in a ^{15}N-labelled protein) (311). These two strategies are effectively opposites of one another, but both can be used to minimise spectral crowding. Crucially, through amino acid type identification, both can also help to resolve ambiguities that arise in backbone connectivity during triple resonance assignment efforts as a consequence of groups of amino acids sharing similar Cα and Cβ chemical shifts. For selective unlabelling, this is achieved through comparison of spectra acquired with uniformly-labelled and selectively-unlabelled samples, and the identification of 'disappearing' peaks in spectra of the selectively-unlabelled sample (311). This disappearance indicates that these correspond to amino acids which were unlabelled, the identity of which is known from the adaptations made to the media during sample preparation. Both selective labelling and unlabelling are achieved through supplementing minimal media with labelled or unlabelled amino acids or their precursors, as desired (311–313). Though selective labelling is typically 'cleaner' and generates more rapidly interpretable spectra, as ^{13}C/^{15}N-enriched amino acids and their precursors are relatively expensive, selective unlabelling can be preferable to minimise experimental costs. Both strategies suffer from metabolic scrambling of NMR-inactive nuclei, but using knowledge of the metabolic pathways in *E. coli* and the extent of cross-metabolism of amino acids, one can account for these effects during data analysis (311). Through the same understanding of *E. coli* metabolic scrambling, combinatorial selective unlabelling strategies can be also be employed (313). In these instances, certain combinations of amino acids are simultaneously selectively-unlabelled to maximize the amount of information that can be gained with as few protein samples as possible. Different methods for choosing amino acid combinations have been proposed (311). One of these proposes selectively unlabelling four pairs of amino acids: Gln/Ile ('-QI'), Asn/Arg ('-NR'), Lys/Leu ('-KL'), and Phe/Val ('-FV') pairs (313). Using these combinations and the fact that Gly, Ala, Ser and Thr residues have unique ^{13}Cα/^{13}Cβ chemical shifts, the authors of this method claim that amino acid type identification can be obtained for ~80% of residues in a protein (313).

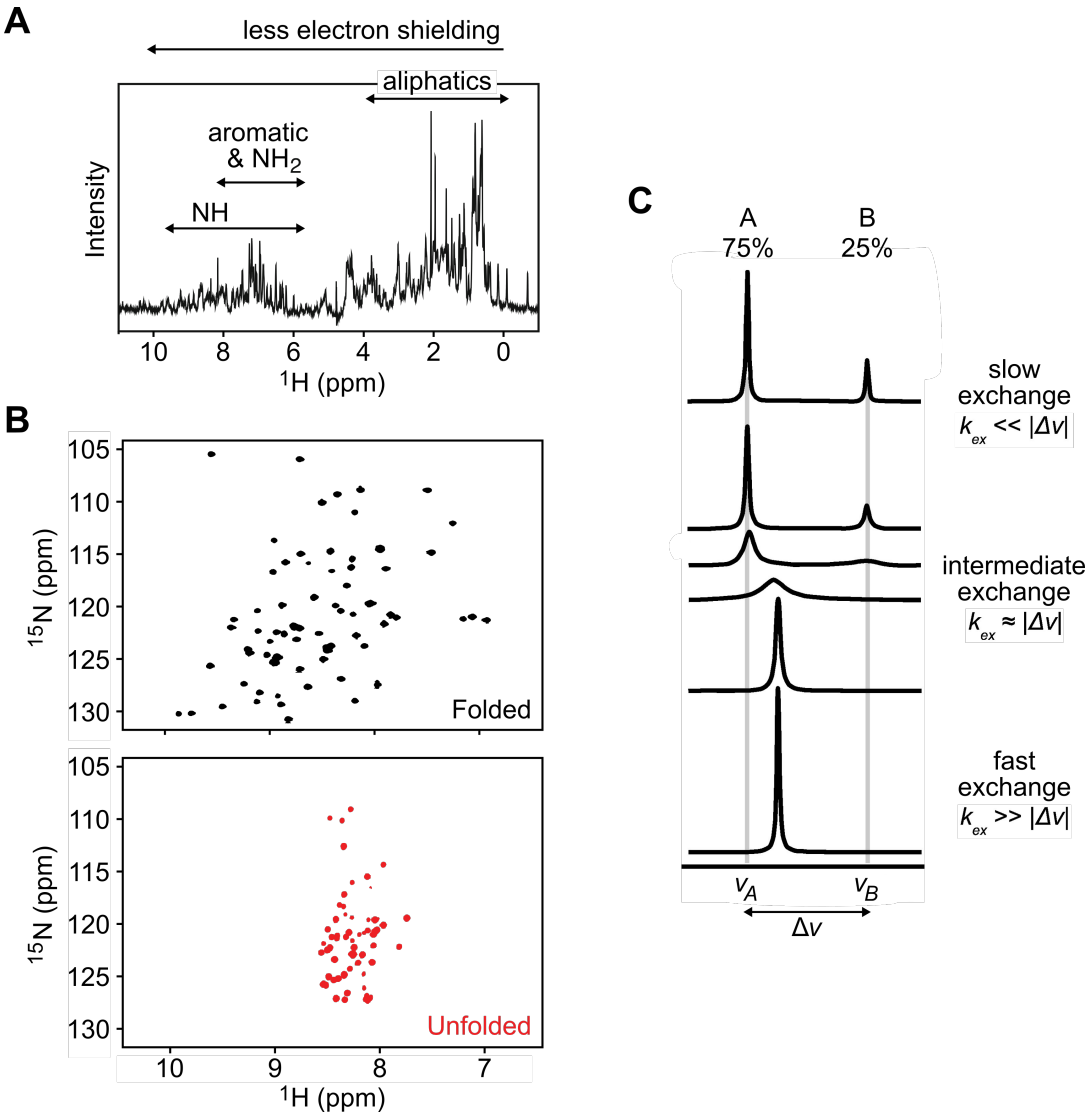


Figure 1.19. Protein analyses by NMR spectroscopy.
(A) 1D ^{1}H spectrum of a 15 kDa protein with the regions of the spectrum corresponding to different types of protons highlighted, illustrating the effect of electron shielding of nuclei on their observed chemical shift. Adapted from Bersch *et al.* (314). (B) ^{1}H-^{15}N HSQC spectra of a 10 kDa protein in folded (black) and unfolded (red) states, illustrating the effect of protein structure on the 'HSQC fingerprint' of the protein. Adapted from Alderson (315). (C) Illustration of the different exchange regimes observable when exchange between two states A and B occurs during data acquisition. Example spectra are shown for a case where 75% of the sample is in state A and 25% is in state B. The regime in which exchange occurs in depends on how the rate of exchange (k_{ex}) compares to the difference in chemical shift frequency ($\Delta\nu$) of the two states. Adapted from Kleckner *et al.* (316).

1.5. Aims

The aim of this project was to investigate how the FGFR family of receptor tyrosine kinases is regulated on a structural and mechanistic level. To this end, two unique but key aspects of FGFR activity regulation were considered: (a) how FGFR activity is stimulated in response to exogenous ligand binding in the context of full-length receptors, and (b) how kinases are regulated by the cellular chaperone Hsp90. FGFR3 was used as a representative of FGFRs and kinases in each case.

(a) How are full-length FGFRs (and RTKs) activated by exogenous ligands?

As described previously, there is currently a lack of understanding as to how intact full-length RTKs are activated and function from a structural view. As RTKs, this also holds true for the FGFR family. To address this missing understanding, one aim of this project was to resolve a high-resolution structure of full-length FGFR3. While focusing on this individual member of the FGFR family, multiple forms of the receptor would be considered to maximise the likelihood of obtaining a full-length structure, and if possible, permit comparisons between 'wild-type' and pathogenic variants to help understand how receptor activation can be achieved in different settings. Given the paucity of structural information in the RTK field with respect to full-length receptors, any intact receptor structure, even at intermediate resolutions (e.g. 6-10 Å), would greatly improve our understanding of their biology.

(b) How are kinases regulated by the cellular chaperone Hsp90?

While studies thus far have revealed some details as to how kinases are recruited to and are chaperoned by Hsp90, many aspects of kinase structure during this chaperone process are poorly or entirely undefined. Notably, the structural basis of kinase client strength is not fully understood, and while appreciated that the co-chaperone Cdc37 locally remodels kinases during their recruitment to Hsp90, precisely where kinases are remodelled is unclear. Moreover, while a structural snapshot of a kinase bound to Hsp90 has been captured, this corresponds to only one of likely several Hsp90-bound states of kinases. How kinases progress through the conformational cycle of Hsp90 is unclear. To address these gaps in our understanding, using FGFR3 kinase domain as a representative kinase, this project first aimed to define how kinases and the co-chaperone Cdc37 interact with one another, detailing precisely which areas of the kinase are remodelled prior to delivery to Hsp90. Building on this, a further aim was to investigate the processes of kinase transfer to Hsp90, capturing kinase-bound ternary complexes for structural evaluation.

As a preface to each chapter, the aims relative to that chapter will be reiterated and the specific objectives set to realise these aims will be introduced.

Chapter 2. Materials and Methods

2.1. Chemicals

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich, Honeywell Fluka™, Melford Laboratories, VWR International, Alfa Aesar, or Fisher Scientific.

2.2. General molecular biology

2.2.1. Buffers

All buffers were prepared using Milli-Q® water then filtered using 0.22 µm or 0.45 µm membrane filters (GE Healthcare Life Sciences) or syringe filters (Acrodisc®, Pall Corporation or Minisart™, Sartorius). Buffers prepared for use with ÄKTA or HPLC systems were additionally degassed prior to use.

2.2.2. Primers

All primers used herein were standard, desalted primers, synthesised by Eurofins Genomics or Sigma Aldrich. A list of primers used in this project are given in Appendix 1.

2.2.3. Media

All media used for the culturing of *Escherichia coli (E. coli)* (summarised in Table 2.1) were prepared in Milli-Q® water and sterilised by autoclaving. Where necessary, some components (e.g. glucose, antiobiotic stocks, isopropyl- β-D-thiogalactoside (IPTG) etc.) were sterilised by 0.22 µm filtration for later addition.

2.2.4. Agarose gel electrophoresis

To visualise DNA samples, 0.8-1% *w/v* agarose gels containing 0.5 x SYBR™ Safe DNA Gel Stain (Invitrogen) were prepared with 1 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH ~8.5). Samples were prepared in 6 x Purple Gel Loading Dye (New England BioLabs) then loaded alongside either 100 bp, 1 kb or 2-log DNA Ladders (New England BioLabs) or GeneRuler DNA Ladder Mix (Thermo Fisher Scientific). Gels were run in 1 x TAE buffer at a constant 100 V then visualised using a G:BOX (Syngene).

Table 2.1. Media for *E. coli* cultures.

Medium	Components
Luria Broth (LB), 1 L	12.5 g granulated LB mix (Melford Laboratories) 1 L Milli-Q® water *autoclaved*
Terrific Broth (TB), 1 L	46.7 g granulated TB mix (Melford Laboratories) 4 mL glycerol 996 mL Milli-Q® water *autoclaved*
2YT Broth (2YT), 1 L	16 g tryptone 10 g yeast extract 5 g NaCl 1 L Milli-Q® water *autoclaved*
LB-agar, 200 mL	5 g granulated LB mix (Melford Laboratories) 3 g agarose 200 mL Milli-Q® water *autoclaved*
TB-agar, 200 mL	9.34 g granulated TB mix (Melford Laboratories) 3 g agarose 200 mL Milli-Q® water *autoclaved*

2.2.5. SDS-PAGE

Samples were prepared in 5 x SDS sample buffer (250 mM Tris-HCl pH 6.8, 10% *w/v* SDS, 50% *v/v* glycerol, 0.25% *w/v* bromophenol blue, 500 mM DTT). As appropriate, samples were boiled at 95 °C for 10 minutes then spun by centrifugation at 10,000 g for 1 minute, or simply incubated at room temperature for 10 minutes, before loading into either Mini-Protean® TGX (4-20%) or Criterion™ TGX (4-20%) Precast gels (Bio-Rad) alongside either Precision Plus Protein™ Dual Color Standards (Bio-Rad) or Color Prestained Protein Standard (Broad Range, New England BioLabs). Samples containing membrane proteins were not boiled but were incubated at room temperature. Proteins were separated by running at a constant voltage of either 120 or 150 V in SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% (*w/v*) SDS, pH ~8.3), then visualised using in-gel green fluorescent protein (GFP) fluorescence, by western blotting or Coomassie staining (Quick Coomassie Stain, Generon), each visualised with a G:BOX (Syngene) using the appropriate settings. Where required for protein identification by mass spectrometry, SDS-PAGE gels were stained with Bio-Safe™ Coomassie Stain (Bio-Rad).

2.2.6. Native-PAGE

Samples were prepared in 4 x Native sample buffer (62.5 mM Tris-HCl pH 6.8, 40% *v/v* glycerol, 0.01% *w/v* bromophenol blue) then loaded into Mini-Protean® TGX Precast gels (4-20%, Bio-Rad). Proteins were separated by running at a constant voltage of

either 120 or 150 V in Native-PAGE running buffer (25 mM Tris, 192 mM glycine, pH ~8.3) at room temperature, then visualised using in-gel GFP fluorescence, by western blotting or Coomassie staining (Quick Coomassie Stain, Generon), each visualised with a G:BOX (Syngene) using the appropriate settings.

2.2.7. Western blotting

Following the separation of proteins by SDS-PAGE or Native-PAGE, proteins were transferred onto nitrocellulose or PVDF membranes using Trans-Blot® Turbo™ Transfer Packs (Bio-Rad) and a Trans-Blot® Turbo™ Transfer System (Bio-Rad). After transfer, membranes were blocked in Tris-buffered saline (20 mM Tris pH 7.6, 150 mM NaCl) plus 0.1% *v/v* Tween®-20 (i.e. TBST) containing 2-3% *w/v* bovine serum albumin (BSA) or dried skimmed milk powder (Marvel), either at room temperature for 1 hour, or at 4 °C overnight, on a roller mixer. As required, blocked membranes were incubated with the desired HRP-conjugated antibody or primary antibody on a roller mixer for 1 hour at room temperature. After incubation with primary antibodies, membranes were washed three times with TBST, blocked in TBST containing 2-3% *w/v* BSA or milk for 1 hour at room temperature, and then incubated with a secondary antibody, as appropriate, for a further hour at room temperature. After antibody incubations, membranes were washed three times with TBST then developed with Clarity™ Western ECL Substrate (Bio-Rad) and imaged with a G:BOX (Syngene). Where necessary, to re-probe with additional antibodies, membranes were stripped with Abcam mild stripping buffer (200 mM glycine pH 8.0, 5 mM SDS, 1% *v/v* Tween®-20) according to their guidelines, or to visualize the proteins transferred, membranes were stained with Ponceau Red stain (0.1% *w/v* Ponceau S, 5% *v/v* acetic acid). Antibodies used in this study are listed in Table 2.2.

2.2.8. *E. coli* transformations

Unless otherwise stated, all transformations into various *E. coli* strains were completed using a standard ‘heat-shock’ protocol. Following plasmid/DNA addition to an aliquot of the appropriate *E. coli* strain of competent cells (generally prepared in-house), cells were incubated on ice for 5-15 minutes then ‘heat-shocked’ for 30-40 seconds at 42 °C. After further incubation on ice for at least 15 minutes, cells were supplemented with 500 or 950 µL sterile LB media and incubated at 37 °C, shaking at 200 rpm, for 1 hour, then spread on LB-agar plates containing the appropriate antibiotics and incubated at 37 °C overnight. OmniMAX™ competent cells were used throughout cloning and various *E. coli* expression strains were used during protein expression as detailed in the relevant sections.

Table 2.2. Antibodies used in western blotting.

HRP-conjugated antibodies			
Antigen	**Type**	**Dilution**	**Identifier/Source**
His-tag	Mouse, mAb HRP	1:5,000-10,000	MAB050H, R&D Systems RRID: AB_357354
Primary antibodies			
Antigen	**Type**	**Dilution**	**Identifier/Source**
Phospho-FGFR1* (Tyr653/654)	Rabbit, pAb	1:2,500	#3471, Cell Signaling Technology RRID: AB_331072
Phospho-CDC37 (Ser13) (D8P8F)	Rabbit, mAb	1:5,000	#13248, Cell Signaling Technology RRID: AB_2783724
Mouse FGFR3**	Rabbit, pAb	1:1,000	OAAB11172, Aviva Systems Biology RRID: AB_2630645
Phospho-Tyr	Mouse, mAb	1:1,000	Sc-7020, Santa Cruz Biotechnology RRID: AB_628123
Human GAPDH	Rabbit, pAb	1:1,000	ab37168, Abcam, RRID: AB_732652
Secondary antibodies			
Antigen	**Type**	**Dilution**	**Identifier/Source**
Rabbit IgG	Goat, pAb HRP	1:5,000	Jackson ImmunoResearch Europe RRID: AB_2307391
Mouse IgG (H+L)	Goat, pAb HRP	1:10,000	#31430, Invitrogen RRID: AB_228307

Note: mAb (monoclonal antibody); pAb (polycolonal antibody); HRP (horseradish peroxidase); RRID (antibody registry ID); the anti-PhosphoFGFR1 (Tyr653/4) antibody was raised against a peptide corresponding to residues surrounding Tyr653/654 of human FGFR1 sharing 100% sequence identity with FGFR3 (*); the anti-mouse FGFR3 antibody was raised against a KLH-conjugated synthetic peptide of mouse FGFR3 residues 468-495, sharing 100% sequence identity to human FGFR3 (**).

2.2.9. Polymerase chain reaction

All polymerase chain reactions (PCRs) were completed similarly using Q5® High-Fidelity 2X Master Mix (New England BioLabs) and a T100™ Thermal Cycler (Bio-Rad). Generally, reactions containing 0.4 ng/µL vector, 0.4 µM forward and reverse primers and 1 x Q5® High-Fidelity Master Mix were thermocycled as recommended by the manufacturer. Annealing temperatures were first calculated using the NEB Tm Calculator (tm.calculator.neb.com) then optimised thereafter as necessary using temperature gradients.

For colony PCR, a master mix containing 1 x Q5® High-Fidelity Master Mix and 0.4 µM forward and reverse primers was prepared and 5 µL aliquots made in PCR tube strips

(Bio-Rad). Following cloning (see 2.2.10 and 2.2.11), colonies were picked with a sterile P10 pipette tip and touched onto a 'master LB-agar plate' containing the appropriate antibiotics, aspirated 5-10 times in a 5 µL PCR master mix aliquot, then the tip stored in LB media containing the appropriate antibiotics for overnight culturing, if positive. The reactions were thermocycled as above, except including a 7-minute initial denaturation step at 95 °C.

2.2.10. Ligation-independent cloning

Generally, vectors and inserts were linearised by PCR, then products separated by agarose gel electrophoresis (see 2.2.4). Primers for PCR were designed using the Clontech In-Fusion Cloning Primer Design Tool v1.0 (https://www.takarabio.com/learning-centers/cloning/in-fusion-cloning-tools) and edited where required to obtain more suitable annealing temperatures and to avoid repetitive sequences (such as when cloning around a His tag). Alternatively, the vector was linearised by restriction digestion using the appropriate restriction enzymes in CutSmart® Buffer (New England BioLabs), as recommended by the manufacturer, before separation of products as above. Using a Safe Imager™ 2.0 Blue-Light Transilluminator (Invitrogen), bands at the expected molecular weight were cut from the gel and DNA extracted using a Nucleospin® Gel and PCR Clean Up Kit (Macherey-Nagel), then yields determined using the 'dsDNA' application on a DS-11 Spectrophotometer (DeNovix). In assembly reactions, 20-50 ng of linear vector and various amounts of insert (initially at a 2:1 molar ratio of insert:vector and optimised thereafter) were mixed with either 1 x In-Fusion® cloning mix (Clontech Laboratories) or 1 x NEBuilder® HiFi DNA Assembly cloning mix (New England BioLabs) and incubated at 50 °C for 20 minutes. OmniMAX™ competent *E. coli* cells were transformed with 1 µL of the reaction product and colonies selected for on LB-agar plates containing the appropriate antibiotics. Colony PCR was used to identify colonies containing vectors with the correct sized insert and these used to inoculate 5-10 mL LB overnight cultures (incubated at 37 °C, 200 rpm) supplemented with the appropriate antibiotics. From these, plasmids were isolated using a Nucleospin® Plasmid Miniprep kit (Macherey-Nagel). Successful assembly was verified by sequencing of the entire open reading frame of each plasmid (Eurofins Genomics).

2.2.11. Site-directed mutagenesis

Point mutations and deletions were achieved using Q5® site-directed mutagenesis (New England BioLabs) as per the manufacturer's protocol with minor adaptions. Following PCR, 10 µL 'KLD reactions' were prepared containing 1 µL PCR product, 1 x T4 DNA Ligase buffer (New England BioLabs), and 1 µL of an in-house prepared 'KLD enzyme mix' comprising a 1:1:2 ratio of T4 Polynucleotide Kinase, T4 DNA Ligase and DpnI (New

England BioLabs). 'KLD reactions' were incubated at room temperature for 20 minutes or overnight, then 1 µL used to transform OmniMAX™ competent *E. coli* cells and colonies selected for on LB-agar plates containing the appropriate antibiotics. Colonies were used to inoculate 5-10 mL overnight cultures of LB media and plasmids isolated as in 2.2.10. Successful mutagenesis was verified by sequencing of the entire open reading frame of each plasmid (Eurofins Genomics).

2.2.12. Mass Spectrometry

All mass spectrometry data were acquired and processed by Dr Rachel George of the Faculty of Biological Sciences Mass Spectrometry Facility, University of Leeds.

Denatured intact molecular mass measurements were used to assess protein quality, confirm protein identification, and identify the presence of any (post-translational) modifications. Briefly, proteins were diluted to 5 µM in 0.1% *v/v* trifluoroacetic acid, then desalted using a MassPREP protein desalting column (Waters Corporation, UK) that was monitored by an ACQUITY UPLC M-Class system (Waters Corporation, UK). The desalted eluent from this column was directed into a Xevo QToF G2-XS mass spectrometer (Waters Corporation, UK) via a Z-spray electrospray source, and mass spectrometry (MS) operated in positive time-of-flight mode. Data were processed using the MassLynx v4.1 suite of the software supplied with the mass spectrometer.

In some cases, proteins and post-translational modifications were further identified by protein digestion with trypsin and LC-MS detection of the resulting peptides. Generally, samples were provided to the MS facility for analysis via this method as an SDS-PAGE gel stained with Bio-Safe™ Coomassie Stain (Bio-Rad), or less commonly as a purified in-solution protein sample. Briefly, following the trypsin digestion, peptides were isolated, reconstituted in 0.1% *v/v* trifluoroacetic acid, then loaded onto a Symmetry C18 trap column (Waters Corporation, UK) monitored by an ACQUITY UPLC M-Class system (Water Corporation, UK). The eluent from this column was directed into a Xevo QToF G2-XS mass spectrometer (Waters Corporation, UK) via a Z-spray nanoflow electrospray source, and MS operated in positive time-of-flight mode. Data were processed as for intact mass measurements, except that peptide MS/MS data were also processed with PEAKS Studio (Bioinformatics Solutions Incorporated), and searched against a UniProt database for *Homo sapiens* and the specific sequence of the construct of interest (as in Appendix 2). Searches were allowed for peptide modifications including acetylation, carbamiodomethylation, deamidation, oxidation and phosphorylation events, using an MS mass tolerance of 20 ppm and a fragment ion mass tolerance of 0.05 Da. The false discovery rate was set to 1%.

2.3. Cloning of constructs

The full sequences of all constructs are provided in FASTA format and primers used in cloning in Appendices 1 and 2.

2.3.1. Cloning of full-length and fusion constructs of FGFR3

2.3.1.1. Cloning for insect cell expression

All full-length constructs of FGFR3 used herein are based on a variant of wild-type FGFR3 IIIb in which all tyrosine phosphorylation sites of the kinase domain are mutated to phenylalanine residues bar the kinase-activity essential Y647 of the kinase activation loop. This FGFR3-IIIb-Y579F/Y650F/Y726F/Y762F/Y772F variant is henceforth called FGFR3-5F1Y and is considered to be analogous to the 'wild-type' receptor. Constructs were also prepared with substitutions of C228R, S249C or Y375C mutations in the extracellular domain, each with and without the 'constitutively-activating' kinase domain mutation K652E. These eight constructs were assembled from synthetic double-stranded DNA GeneArt® Strings™ (Thermo Fisher Scientific) which had been codon-optimised for expression in insect cells. For assembly, FGFR3-5F1Y was split into two fragments, FGFR3-Frag1 and FGFR3-Frag2, and six GeneArt® Strings™ ordered corresponding to 'wild-type' FGFR3-Frag1, or FGFR3-Frag1 containing C228R, S249C or Y365C mutations, and 'wild-type' FGFR3-Frag2, or FGFR3-Frag2 containing K652E. These would be able to be recombined in a combinatorial manner to generate the various full-length constructs of interest (Figure 2.1A). The full-length FGFR3-5F1Y-Y375C/K652E construct was generated by overlap-extension PCR of the two relevant GeneArt® Strings™ DNA fragments, then assembled into a pFastBac vector encoding a C-terminal HRV 3C-cleavable monomeric GFP-His8 (CVmGH) tag using ligation-independent cloning (see 2.2.10) (Figure 2.1B and C). As two of the ordered GeneArt® Strings™ were not successfully synthesised, the remaining seven constructs and a 'kinase-dead' variant (FGFR3-5F1Y-K510A) (317) were generated by site-directed mutagenesis (see 2.2.11) (Figure 2.1C). Monomeric GFP was generated through introduction of the mutation A206K into the GFP molecule of the C-terminal tag (318).

To generate C-terminal fusion constructs of FGFR3-5F1Y, plasmids encoding FGFR3-RT4, FGFR3-RTT112, FGFR3-SW780 and FGFR3-deldel, reported previously (89), were kindly provided by Prof Margaret Knowles (University of Leeds). The fusion-encoding regions of these plasmids were PCR-amplified then cloned into the pFastBac vector encoding FGFR3-5F1Y using ligation-independent cloning, replacing exon 19 of the receptor. Each of these fusion constructs thus comprised residues 1-760 of FGFR3-5F1Y followed by the relevant fusion protein. A truncated construct, FGFR3-5F1Y-del19, was generated by deletion of exon 19 from FGFR3-5F1Y.

2.3.1.2. Cloning for mammalian tissue culture expression

For expression of FGFR3-5F1Y in mammalian tissue culture, FGFR3-5F1Y was cloned into pEG_BacMam (kindly provided by Dr Eric Gouaux (319)), preceded by the native Kozak sequence of FGFR3 (GCCCCCGCC) and followed either by a HRV 3C-cleavable GFP-His8 tag or HRV 3C-cleavable FLAG tag using ligation-independent cloning.

2.3.2. Cloning of FRS2α PTB domain

To generate variously tagged constructs of FRS2α phosphotyrosine binding (PTB) domain, a synthetic gene corresponding to the entire coding region of Human FGFR signalling adaptor SNT-1/FRS2α (GenBank ID: AF036717.1) was obtained from GeneArt® Gene Synthesis (ThermoFisher Scientific). Using ligation-independent cloning, FRS2α PTB domain (residues 8-122) was cloned into a pRSET vector containing an N-terminal His6-HRV 3C-cleavable tag (named hereafter as FRS2α$^{N8\text{-}122}$). Alternatively, FRS2α PTB domain with slightly adapted construct boundaries (residues 9-122) with either an N-terminal His6-HRV 3C-GGS tag (FRS2α$^{N9\text{-}122}$) or a C-terminal GGS-HRV 3C-His6 tag (FRS2α$^{C9\text{-}122}$) was prepared in a pET28 vector.

2.3.3. Materials and cloning for FGFR3 kinase domain, Cdc37 and Hsp90

A pJ821 vector encoding N-terminally His6-SUMO-tagged human FGFR3 kinase domain (residues 455-768, IIIc) harbouring C482A, C582S and I538F mutations (known herein as FGFR3-KD^{I538F}), a pET-28 vector encoding human Cdc37 with a HRV 3C-cleavable C-terminal His6-tag, a pRSET vector encoding human Hsp90α with a HRV 3C-cleavable N-terminal His6-tag, a pCDF-Duet dual-expression vector encoding lambda phosphatase (λPP) and human Cdc37 in open reading frames 1 and 2 for co-expression with FGFR3-KD^{I538F}, and a pTWO vector encoding casein kinase II for Ser13 phosphorylation of Cdc37, were kindly gifted by Dr Tom Bunney (University College London). C482A and C582S substitutions present in the FGFR3-KD^{I538F} construct improve the stability of kinase domain by removal of free cysteines prone to inter-protein crosslinking without affecting kinase structure, as reported previously (49).

To co-express FGFR3-KD^{I538F} with λPP alone, human Cdc37 was deleted from the second open reading frame of the pCDF-Duet vector using site-directed mutagenesis. An erroneous V376G mutation present in human Cdc37 in pCDF-Duet and pET28 vectors was also corrected using site-directed mutagenesis prior to use.

Variants of Hsp90α containing E47A or D93N substitutions were generated by site-directed mutagenesis.

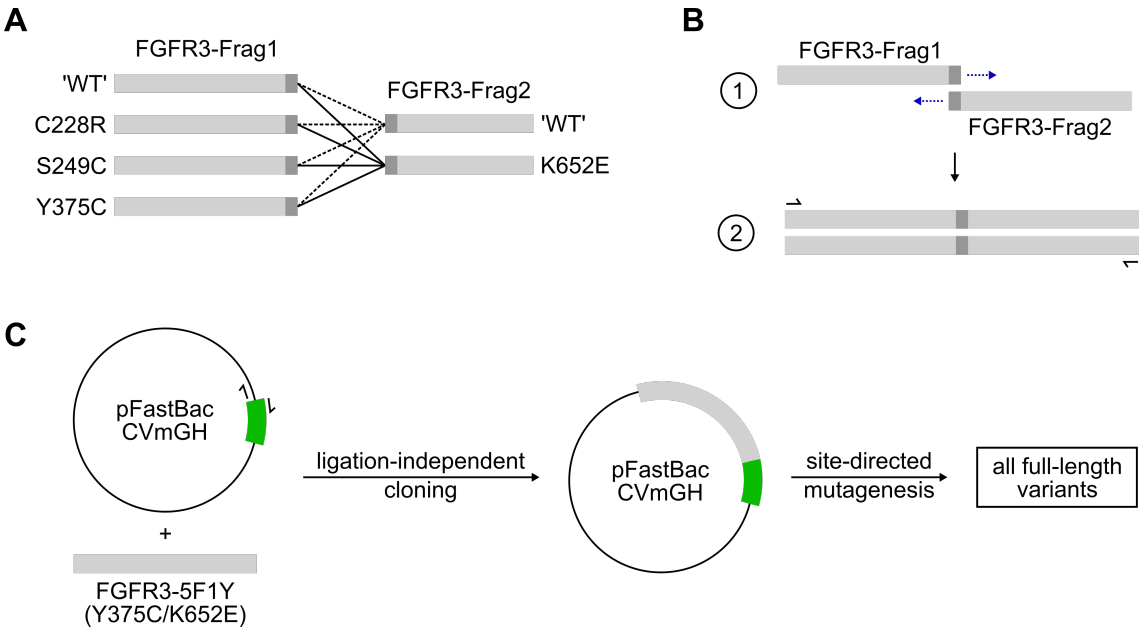


Figure 2.1. Cloning strategy for FGFR3-5F1Y constructs.
(A) Schematic illustrating how DNA GeneArt® Strings™ of FGFR3 split into two fragments, FGFR3-Frag1 and FGFR3-Frag2, could be recombined in a combinatorial manner to generate eight full-length FGFR3-5F1Y constructs containing the highlighted substitutions. The dark grey panel on each fragment corresponds to a 20 bp overlap between FGFR3-Frag1 and FGFR3-Frag2. (B) An illustration of the generation of a full-length construct of FGFR3-5F1Y using PCR overlap extension of fragments using their 20 bp overlap (stage 1) followed by PCR amplification of the product (stage 2). (C) FGFR3-5F1Y constructs in pFastBac vectors were generated by ligation-independent cloning of FGFR3-5F1Y-Y375C-K652E into a pFastBac plasmid containing a C-terminal HRV 3C-cleavable monomeric GFP-His8 tag (CVmGH), then the remaining seven full-length constructs generated by site-directed mutagenesis.

2.4. Insect cell expression of full-length and fusion constructs of FGFR3

2.4.1. Maintenance of insect cells

Sf9 insect cells were maintained in either Insect-XPRESS™ Protein-free Insect Cell Medium with L-glutamine (Lonza) or Sf9-900 II SFM (Serum-free Media) (Gibco™) at cell densities of between 0.5 and 4 million cells/mL with >95% cell viability in Erlenmeyer cell culture flasks at 27 °C, shaking at 120 rpm. Cell densities were counted either manually using a haemocytometer or using a Countess™ II Automated Cell Counter (Invitrogen). Cell viability was assessed by staining with trypan blue stain (Invitrogen).

2.4.2. Bacmid preparation

To prepare bacmids encoding FGFR3-5F1Y constructs for transfection of Sf9 cells, the relevant pFastBac plasmids were transformed into either DH10Bac™ or DH10EmBacY competent cells using an adapted 'heat-shock' protocol. Following incubation at 42 °C for 30 seconds and recovery on ice for 15 minutes, cells were supplemented with 500 µL of sterile LB media and incubated at 37 °C overnight, shaking at 200 rpm. Transformed cells with recombinant bacmids were selected for by blue-white screening on LB-agar plates containing 50 µg/mL kanamycin, 10 µg/mL tetracycline, 7 µg/mL gentamycin, 100 µg/mL X-Gal and 0.16 mM IPTG (DH10Bac™ plates), incubated at 37 °C for at least 24 hours. After confirmation of a blue-white phenotype by re-streaking on and incubation of fresh DH10Bac™ plates, 2 mL cultures of LB media containing 50 µg/mL kanamycin, 10 µg/mL tetracycline, and 7 µg/mL gentamycin were prepared from white (i.e. recombinant bacmid-containing) colonies and incubated at 37 °C overnight, shaking at 200 rpm. From these, bacmids were isolated using an adapted protocol from the Bac-to-Bac® Expression System Manual (Invitrogen). Briefly, overnight cultures were harvested and lysed according to a Nucleospin® Plasmid Miniprep Kit (Macherey-Nagel), then the lysate clarified by centrifugation at 15,000 g for 10 minutes. The supernatant was transferred to a clean tube, clarified once more, then 700 µL isopropanol was carefully layered onto the supernatant (an approx. 3:2 ratio of supernatant:isopropanol), followed by gentle mixing through inversion. After centrifugation at 15,000 g for 10-15 minutes, the supernatant was carefully discarded by aspiration leaving a transparent pellet which was washed in 200 µL of 70% *v/v* ethanol, added dropwise to the side of the tube opposite to the pellet. After a final 15,000 g centrifugation for 5 minutes, the bacmid DNA pellet was stored under 50 µL of fresh 70% *v/v* ethanol, added dropwise as above, then used immediately for transfection.

2.4.3. Transfection

Transfection of Sf9 cells was completed in 12-well plates (Sarstedt). For each construct, two wells were seeded with 2 mL of Sf9 cells at a density of 0.325 million cells/mL and

>98% viability, and cells allowed to adhere to the well surface. Meanwhile, bacmids (see 2.4.2) were prepared for transfection by aspiration of the 70% *v/v* ethanol used for temporary bacmid storage and, following air drying, were resuspended in 30 µL of sterile Milli-Q® water and incubated at room temperature for 10 minutes. For each well, a transfection mixture containing 13 µL of bacmid, 195 µL of Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza) and 6.5 µL of X-TremeGENE HP DNA Transfection Reagent (Roche) was prepared, then 100 µL added dropwise to the well. Control wells containing Sf9 cells or media only were also prepared. Plates were incubated at 27 °C for 3 days then the media containing V0 virus collected. Viruses were stored in LightSafe tubes at 4 °C.

To assess transfection and expression efficiency, following V0 isolation, 2 mL of fresh media was added to each well and the plates incubated for a further 3 days at 27 °C. Cells were visualised using an EVOS FL Auto 2 Imaging System (Thermo Fisher Scientific) in bright-field and GFP fluorescence modes (λ_{ex}: 470 nm, λ_{em}: 525 nm), then, following media removal, cells were scraped and resuspended in 1 x PBS (10 mM Na_2HPO_4, 2.7 mM KCl, 1.8 mM KH_2PO_4, 137 mM NaCl, pH 7.4) then pelleted by centrifugation at 300 g for 5 minutes. Pellets were resuspended in 150 µL of chilled NP-40 lysis buffer (50 mM Tris-HCl pH 8.0, 120 mM NaCl, 0.5% *v/v* Nonidet® P 40 substitute) containing 1 mM PMSF and 1x cOmplete EDTA-free protease inhibitors (Roche) and transferred to a black, clear-bottomed 96 well microplate and GFP fluorescence (λ_{ex}: 380 nm, λ_{em}: 520 nm) or yellow fluorescent protein (YFP) fluorescence (λ_{ex}: 485 nm, λ_{em}: 520 nm) measured on a FluoSTAR Optima microplate reader (BMG LabTech). Cell lysates were further analysed by SDS-PAGE followed by visualisation of in-gel GFP fluorescence and anti-His western blotting (see 2.2.5 and 2.2.7).

2.4.4. Virus amplification and small-scale expression trials

To amplify baculoviruses for expression, V0 viruses were used to infect Sf9 cells at a cell density of 1 million cells/mL in Erlenmeyer flasks, then cells maintained at 27 °C and 120 rpm with cell density and viability monitored daily. Typically, 17-25 mL of Sf9 cells were infected with 2-3 mL of V0 virus. Infected cells were maintained at below 1.5 million cells/mL until DPA+24h (24 hours after the day following proliferation arrest) when V1 viruses were harvested by the centrifugation of cultures at 300 g for 5 minutes and the supernatant transferred to LightSafe tubes for storage at 4 °C. If desired, to determine the day of best expression, cell pellets were resuspended in fresh medium then cultures maintained until DPA+72h or until cell viability was below 65%, whichever occurred first. To assess the day of optimum recombinant protein expression, approximately 1 million cells were resuspended in 50 µL of NP-40 lysis buffer supplemented with 1 mM PMSF and 1 x cOmplete EDTA-free protease inhibitors

(Roche) and incubated on ice for 1 hour. Total lysate samples were clarified by centrifugation at 10,000 g for 5-10 minutes then the supernatant (i.e. the clarified lysate) used to prepare SDS-PAGE samples and analysed as before.

2.4.5. Preparation of baculovirus-infected insect cells

For long term storage of viruses and for protein expression, baculovirus-infected insect cells (BIICs) (320) were prepared. Multiples of 100 mL of Sf9 cells at 1 million cells/mL, pre-adapted in 500 mL Erlenmeyer flasks, were infected with various volumes and dilutions (prepared in insect cell medium) of V1 viruses. Cultures were incubated at 27 °C, shaking at 120 rpm, and the cell density monitored. To prepare BIICs, a low titer of virus is desired that allows the cells to double once then experience proliferation arrest (PA); this corresponds to a multiplicity of infection (MOI) of less than 1. Upon identifying a suitable virus titer, cultures were incubated until the day following proliferation arrest (DPA), then harvested by centrifugation at 200 g for 10 minutes. The supernatant was discarded, then the cell pellet resuspended to a cell density of 10 million cells/mL in insect cell medium containing 10 g/L BSA and 10% *v/v* DMSO, and 1 mL aliquots prepared in cryovials. These cryovials containing BIICs were transferred to a Mr. Frosty™ Freezing Container (Thermo Scientific) then placed in a -80 °C freezer for 24-48 hours prior to long-term storage in liquid nitrogen.

2.4.6. Large-scale expression

FGFR3-5F1Y was expressed using BIICs as a virus source. Per BIIC, 4 x 300 mL of Sf9 cells at 1 million cells/mL in 1 L Erlenmeyer flasks were prepared. To infect cells, a BIIC was thawed quickly through warming by hand and diluted into 50 mL of pre-warmed insect cell medium, then 12.5 mL added to each flask. Infected cultures were incubated at 27 °C, shaking at 120 rpm, and the cell density monitored. Typically, following infection, the cells double only once using this method. Upon reaching the day of best expression (DPA+48h in the case of FGFR3-5F1Y), cells were harvested by centrifugation at 700 g for 30 minutes at 4 °C. Cell pellets were washed in ice-cold 1 x PBS (pH 7.4), flash frozen in aliquots corresponding to 600 mL of Sf9 cells, then stored at -80 °C.

2.5. *E. coli* expression of proteins

2.5.1. Expression of FRS2α PTB domain

2.5.1.1. Expression optimisation

$FRS2\alpha^{N8\text{-}122}$ was initially expressed using protocols present in the literature as guidelines (59,321). Following poor expression and purification yields, small-scale expression trials for $FRS2\alpha^{N8\text{-}122}$, $FRS2\alpha^{N9\text{-}122}$ and $FRS2\alpha^{C9\text{-}122}$ were completed to optimise expression. In

these trials, constructs were transformed into BL21 Gold pRARE2 and BL21 Star pRARE2 *E. coli* competent cells and colonies selected for on LB-agar plates supplemented with the appropriate antibiotics, overnight at 37 °C. Starter cultures were prepared by inoculation of 10 mL of LB media supplemented as appropriate, then incubated at 37 °C overnight, shaking at 200 rpm. The next morning, the starter cultures were used to inoculate 60 mL supplemented LB media in 250 mL baffled flasks at a starting OD_{600} of 0.05, then cultures grown at 37 °C, shaking at 200 rpm. Upon reaching an OD_{600} of ~0.5-0.7, expression was induced with 0.3 mM IPTG and cultures incubated overnight at 18 °C, shaking at 200 rpm. Media was supplemented with 25 µg/mL chloramphenicol and either 100 µg/mL carbenicillin (for pRSET FRS2α^{N8-122}) or 50 µg/mL kanamycin (for pET28 FRS2α^{N9-122} and FRS2α^{C9-122}).

2.5.1.2. Large-scale expression

FRS2α^{C9-122} was expressed in BL21 Gold pRARE2 as detailed above. Briefly, multiples of 600 mL of LB supplemented with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol were inoculated using an overnight starter culture, then the cultures incubated at 37 °C, shaking at 100 rpm, until reaching an OD_{600} of ~0.7. Expression was induced with 0.3 mM IPTG then cultures incubated at 18 °C and 100 rpm for 20 hours. Cells were harvested by centrifugation at 6,000 g for 20 mins, and then pellets stored at -20 °C until required.

2.5.2. Expression of Ulp1

pET-28b(+) encoding Ulp1 (pFGET19_Ulp1, gifted by Hideo Iwaï, Addgene #64697) was transformed into T7 Express Competent (High Efficiency) cells (New England BioLabs) and colonies selected for on LB-agar plates containing 50 µg/mL kanamycin overnight at 37 °C. Starter cultures were prepared by inoculation of LB media (50 mL in 250 mL baffled flasks) with a single colony followed by incubation at 37 °C and 220 rpm overnight. Starter cultures were used to inoculate LB media (1 L in 2L baffled flasks) at a starting OD_{600} of 0.05, then cultures were incubated at 37 °C and 150 rpm until reaching an OD_{600} of 0.5 when they were induced with 1 mM IPTG for 21 hours. After induction, cultures were harvested by centrifugation at 6,000 g for 20 minutes at 4 °C.

2.5.3. Expression of FGFR3-KD^{I538F}

pJ821 encoding FGFR3-KD^{I538F} and pCDF-Duet encoding λPP alone were transformed into OverExpress™ C41 (DE3) cells (Lucigen) and colonies selected for on TB-agar plates containing 10 mM glucose, 50 µg/mL kanamycin and 50 µg/mL streptomycin at 37 °C overnight. Multiples of 600 mL of TB (supplemented as before) were inoculated with several colonies, then cultures shaken at 100 rpm and 37 °C until reaching an OD_{600} of 0.8-1.0. Cultures were cooled to 15 °C for 2 hours then induced with 0.1 mM IPTG

and 1 mM L-rhamnose for 16 hours. After induction, cells were harvested by centrifugation at 6,000 g for 20 mins and pellets stored at -20 °C until required.

2.5.4. Expression of Cdc37

pET28 encoding Cdc37 and pTWO encoding casein kinase II were transformed into C41 (DE3) pRARE2 cells and colonies selected for on LB-agar plates containing 50 µg/mL kanamycin, 100 µg/mL carbenicillin and 25 µg/mL chloramphenicol at 37 °C overnight. Starter cultures were prepared by inoculation of 60 mL of 2YT media supplemented with the same antibiotics and the culture grown overnight, shaking at 220 rpm and 37 °C. Overnight cultures were used to inoculate multiples of 500 mL of 2YT media supplemented with antibiotics as before, then these shaken at 100 rpm and 37 °C until reaching an OD_{600} of 1.0. After cooling to 25 °C for 2 hours, expression was induced with 1 mM IPTG for 4 hours. Cultures were harvested by centrifugation at 6,000 g for 20 mins and pellets stored at -20 °C until required.

2.5.5. Expression of Hsp90

pRSET encoding Hsp90α (wild-type, E47A or D93N variants) was transformed into C41 (DE3) pRARE2 cells or BL21 DE3 cells and colonies selected for on LB-agar plates containing 100 µg/mL carbenicillin (and 25 µg/mL chloramphenicol for C41 DE3 pRARE2 cells) at 37 °C overnight. Multiples of 500 mL of 2YT media containing 100 µg/mL carbenicillin were inoculated with several colonies each, then shaken at 180 rpm and 37 °C until reaching an OD_{600} of ~0.4. Cultures were cooled to 25 °C for 2 hours then induced with 1 mM IPTG, shaking at 180 rpm overnight. Cultures were harvested by centrifugation at 6,000 g for 20 mins and pellets stored at -20 °C until required.

2.6. Purification of proteins

All steps were performed at 4 °C unless stated otherwise.

2.6.1. Purification of full-length FGFR3

2.6.1.1. Membrane preparation

Generally, pellets corresponding to 600 mL of Sf9 culture were thawed on ice, then resuspended in 30 mL of FGFR3 homogenisation buffer (20 mM HEPES-NaOH pH 8.0, 200 mM NaCl, 1 mM TCEP, 10% *v/v* glycerol) supplemented with 1 mM PMSF, 1 mM EDTA, and either 1 x Pierce Protease Inhibitors (EDTA-free, ThermoFisher Scientific) or 1 x cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche). Cells were lysed through 30 passes in a dounce homogeniser with a tight plunger then cell debris and non-lysed cells pelleted by centrifugation at 3,000 g for 20 mins. The supernatant was recovered, then the pellet resuspended and any remaining cells lysed by dounce homogenisation once more as before. Following a second round of centrifugation at

3,000 g for 20 mins, the recovered supernatants were combined and membranes pelleted by centrifugation at 42,000 rpm for 1 hour in a Type 45 Ti rotor (Beckman Coulter). The membrane pellet was washed through multiple cycles of homogenisation and centrifugation at 42,000 rpm for 40 mins in a Type 45 Ti rotor. Membranes were first washed with FGFR3 homogenisation buffer containing 1 mM PMSF and 1 mM EDTA, then with FGFR3 homogenisation buffer containing 1 mM PMSF only. Following washing, the membrane pellet was resuspended in 5 mL of FGFR3 homogenisation buffer containing 1 mM PMSF, and the total protein concentration estimated using a Pierce™ BCA Protein assay kit (ThermoFisher Scientific). Typically, the concentration obtained was ~6-7 mg/mL. Membranes were flash frozen in liquid nitrogen as droplets, then stored at -80 °C.

2.6.1.2. Solubilisation trials

Solubilisation trials were completed both using whole-cell pellets and prepared membranes.

Firstly, solubilisation trials were completed using whole cells using a range of detergents (Fos-choline-12, n-Decyl-β-D-Maltoside (DM), n-Dodecyl-β-D-Maltoside (DDM), decyl maltose neopentyl glycol (DMNG), lauryl maltose neopentyl glycol (LMNG), Cyclofos-6, octaethylene glycol monodecyl ether (C12E8), n-Dodecyl-N,N-Dimethylamine-N-Oxide (LDAO) and Anzergent 3-8) in the presence and absence of the cholesterol mimic cholesteryl hemisuccinate (CHS). The co-polymers SMA and DIBMA were also tested. Solubilisation mixtures were prepared through dilution of Sf9 cells to an OD_{600} of 2.0 in FGFR3 homogenisation buffer lacking TCEP but supplemented with 1 x cOmplete Protease Inhibitor Cocktail (EDTA-free, Roche) and the solubilising agent of interest. Detergents were used at a 1% *w/v* final concentration, and co-polymers SMA and DIBMA at a 2% *w/v* final concentration. Where used, CHS was present at a 0.2% *w/v* concentration. A control sample containing no solubilising agent was also prepared. Samples were incubated at 4 °C (for detergents and control) or at room temperature (for SMA and DIBMA) on a roller mixer, and solubilisation allowed to proceed overnight. After incubating for the desired time, the insoluble fraction of samples was pelleted by centrifugation at 55,000 rpm for 1 hour at 4 °C using a TLA-100 rotor (Beckman Coulter). To quantify solubilisation efficiency, total and soluble fractions were analysed by SDS-PAGE followed by measurement of in-gel GFP fluorescence (for SMA and DIBMA) and anti-His tag western blotting (for all others), and quantification by densitometry using Fiji/ImageJ. Solubilisation efficiencies are expressed as a percentage of the total protein for each condition. In this case, reported solubilisation efficiencies correspond to a single measurement.

Building upon these whole-cell experiments, solubilisation trials using a narrower range of detergents, SMA and DIBMA were completed using prepared membranes. Solubilisation mixtures were prepared containing membranes (~0.15 mg/mL final concentration) in FGFR3 homogenisation buffer and the solubilising agent of interest. Detergents (Fos-choline-12, DM, DDM, DMNG, LMNG and Cyclofos-6) were used at a 1% *w/v* final concentration plus 0.2% *w/v* CHS, and SMA and DIBMA co-polymers at a 2% *w/v* final concentration (without CHS). A control sample containing no solubilising agent was also prepared in parallel. Samples were treated and analysed as in whole-cell solubilisation trials. Solubilisation efficiencies are reported as the mean value plus standard deviation; experiments were completed between one and four times, each as reported.

2.6.1.3. Purification trials

All trial purifications of FGFR3-5F1Y were completed similarly with minor adaptations. Generally, membranes were diluted in FGFR3 homogenisation buffer to a total protein concentration of 2 mg/mL, then solubilised for 2 hours (or overnight in one instance) with the desired solubilising agent. The receptor was solubilised with either LMNG or DDM (at either 2% *w/v* with 0.4% *w/v* CHS, or 1% *w/v* with 0.2% *w/v* CHS) at 4 °C or the co-polymer DIBMA (at 2% *w/v*) at room temperature. Following solubilisation, insoluble material was pelleted by centrifugation at 50,000 rpm for 1 hour in a TLA110 rotor at 4 °C (Beckman Coulter), then the supernatant bound to Nickel Sepharose Fast Flow 6 resin pre-equilibrated with FGFR3 homogenisation buffer for 1-2 hours at 4 °C. After incubation, the resin was transferred to an Econo-Pac® Chromatography Column (Bio-Rad), then washed with FGFR3 His Wash Buffer 1 (20 mM HEPES-NaOH pH 8.0, 200 mM NaCl, 10% *v/v* glycerol, 1 mM TCEP, 10 mM imidazole; supplemented with 0.02% LMNG and 0.004% CHS, 0.05% DDM and 0.001% CHS, or alone in the case of DIBMA) and FGFR3 His Wash Buffer 2 (20 mM HEPES-NaOH pH 8.0, 200 mM NaCl, 5% *v/v* glycerol, 1 mM TCEP, 20 mM imidazole; plus 0.01% LMNG and 0.002% CHS, 0.05% DDM and 0.001% CHS, or alone in the case of DIBMA). At this stage, protein elution and tag cleavage occurred in one of two ways. For on-column tag cleavage, the resin was exchanged into HRV Cutting Buffer (20 mM HEPES-NaOH pH 8.0, 200 mM NaCl, 5% *v/v* glycerol, 1 mM TCEP; supplemented as required like with FGFR3 His Wash Buffer 2) and incubated in the presence of HRV 3C protease at 4 °C overnight, then the flow-through collected. Following washing of the resin with FGFR3 homogenisation buffer, any bound proteins were eluted with FGFR3 His Elution Buffer (20 mM HEPES-NaOH pH 8.0, 200 mM NaCl, 5% *v/v* glycerol, 1 mM TCEP, 250 mM imidazole; supplemented as before as required). Alternatively, for off-column tag cleavage, bound proteins were eluted using FGFR3 His Elution Buffer, then elution fractions containing

target protein dialysed into HRV Cutting Buffer in the presence of HRV 3C protease, overnight at 4 °C. Dialysed samples were further purified by reverse-immobilised metal affinity chromatography (IMAC) through binding to Nickel Sepharose Fast Flow 6 resin pre-equilibrated with HRV Cutting Buffer and the resin washed with HRV Cutting Buffer to collect any unbound protein. After tag cleavage, untagged FGFR3-5F1Y was expected to be in this unbound fraction. Finally, bound proteins were eluted with FGFR3 His Elution Buffer as before.

Alternatively, purification of FGFR3-5F1Y was attempted using GFP-Trap® resin (Chromotek). In this case, purification of FGFR3-5F1Y alone and in complex with ligand FGF1 was tried. To purify a ligand-bound complex, membranes were diluted to 2 mg/mL in FGFR3 homogenisation buffer and incubated with human FGF1 (ab155608, Abcam) and heparin dp8 oligosaccharide (H008, Iduron) at equal amounts (i.e. 60 µg of each component) for 1 hour at 4 °C. This approximates to 7.5 x molar excess of FGF1 and 50 x molar excess of heparin oligosaccharide dp8. The amount of FGFR3-5F1Y in isolated membranes was estimated by GFP quantification. Following incubation with ligands, membranes were solubilised using 2% *w/v* DIBMA at room temperature for 2 hours, then insoluble material pelleted by centrifugation at 48,000 rpm in a TLA-110 rotor (Beckman Coulter) at 4 °C for 1 hour. After centrifugation, the supernatant fractions were incubated with GFP-Trap® resin for 2 hours in a 1.5 mL microcentrifuge tube. Resin was pelleted by centrifugation at 5,000 g, then washed with FGFR3 homogenisation buffer three times with centrifugation between each wash to remove unbound proteins. Proteins were eluted from the GFP-Trap® resin by on-column tag cleavage through incubation with HRV 3C protease in FGFR3 homogenisation buffer overnight. In this instance, unbound proteins following the initial sample incubation with GFP-Trap® resin were further purified by IMAC using Nickel Sepharose Fast Flow 6 resin as above, except with an overnight incubation with the resin and no tag cleavage.

Where required, samples were concentrated using Vivaspin ultrafiltration devices (50K MWCO, PES) at 4 °C. Sample concentrations were estimated using absorbance at 280 nm and calculated molecular weight and extinction coefficients.

2.6.2. Purification of FRS2α PTB domain

Pellets were thawed on ice then resuspended in a 1:1 cell pellet:buffer ratio of *E. coli* lysis buffer (25 mM Tris-HCl pH 8.0, 250 mM NaCl, 1 mM $MgCl_2$, 100 uM $CaCl_2$) supplemented with 1 mM TCEP, 1 x Proteoloc™ Protease Inhibitor Cocktail (EDTA-free, Expedeon) and 1000 U/g cell pellet OmniCleave™ Endonuclease (Epicentre). Cells were lysed by incubation on ice in the presence of 500 µg/mL lysozyme and 2% *v/v* Triton X-100 for 1 hour, followed by sonication on ice (25 x 10 sec pulse, 30 sec rest, 70% amplitude, QSonica). Cell debris was pelleted by centrifugation at 40,000 g for

20 mins, then the supernatant bound to Nickel Sepharose Fast Flow 6 resin pre-equilibrated with *E. coli* lysis buffer for 1 hour in the presence of 10 mM imidazole. After washing the resin with 3 x 10 column volumes (CV) of *E. coli* lysis buffer, 2 x 10 CV of FRS2α wash buffer 1 (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM TCEP, 20 mM imidazole) and 2 x 10 CV of FRS2α wash buffer 2 (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM TCEP, 40 mM imidazole), bound protein was eluted with 10 x 1.2 CV FRS2α elution buffer 1 (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM TCEP, 250 mM imidazole) and 8 x CV FRS2α elution buffer 2 (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM TCEP, 500 mM imidazole). Elution fractions were combined then dialysed overnight into FRS2α dialysis buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP) in the presence of HRV 3C protease to cleave the His6 tag, using Snakeskin Dialysis Membrane (10K MWCO). The dialysed and tag-cleaved sample was purified further by reverse-IMAC through incubation with Nickel Sepharose Fast Flow 6 resin pre-equilibrated with FRS2α dialysis buffer for 15 minutes, then the flow through collected. Any remaining unbound protein was collected through washing the resin with 2 x 20 CV FRS2α dialysis buffer. Bound proteins were eluted with 4 x 2 CV of FRS2α elution buffer 2. The flow through and wash fractions containing FRS2α PTB domain were concentrated using Vivaspin 20 ultrafiltration units (10K, PES), then injected onto a HiLoad™ Superdex 75 16/600 column (GE Healthcare) equilibrated with FRS2α gel filtration buffer (25 mM Tris-HCl, 200 mM NaCl, 1 mM TCEP, 0.5 mM EDTA, pH 8.0). Ion exchange chromatography using a HiTrap Q HP column (GE Healthcare) prior to size exclusion chromatography did not yield any purity enhancement. Fractions containing monomeric FRS2α$^{C9\text{-}122}$ were concentrated using Vivaspin ultrafiltration units (10K, PES) to ~2.5 mg/mL, then flash frozen in liquid nitrogen and stored at -80 °C.

2.6.3. Purification of Ulp1

Bacterial pellets were resuspended in Ulp1 Lysis Buffer (50 mM sodium phosphate, 300 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, pH 8.0, supplemented with 6 µg/mL bovine pancreatic DNAse I and 1x cOmplete protease inhibitors EDTA-free (Roche)) at a 1:1 ratio of wet pellet mass:volume, then lysed by sonication on ice (25 x 10 second pulses with 30 second rests, 70% amplitude, QSonica). After clarification by centrifugation at 40,000 g for 30 minutes at 4 °C, the supernatant was incubated with Nickel Sepharose Fast Flow 6 resin (GE Healthcare) pre-equilibrated with Ulp1 Wash Buffer A (50 mM sodium phosphate, 300 mM NaCl, pH 8.0). After washing with Ulp1 Wash Buffer A and Ulp1 Wash Buffer B (50 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 8.0), proteins were eluted with Ulp1 Elution Buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted proteins were dialysed into 1 x PBS using Snakeskin Dialysis Tubing (10K MWCO, Thermo Fisher Scientific), overnight at 4 °C,

then concentrated with Vivaspin ultrafiltration units (10K MWCO, PES, Sartorius). Protein concentration was estimated by absorbance at 280 nm and an estimated extinction coefficient. Purified Ulp1 was supplemented with 50% *v/v* glycerol and 25 mM DTT, flash frozen with liquid nitrogen and stored at - 80 °C at 4.2 mg/mL

2.6.4. Purification of FGFR3-KD^{I538F}, Cdc37 and Hsp90

Pellets were thawed on ice then resuspended in 20-50 mL of chilled *E. coli* Lysis Buffer supplemented with 100 µg/mL lysozyme and 1 x cOmplete protease inhibitors (EDTA-free, Roche) and agitated on a roller mixer at 4 °C. After 1 hour, the suspension was supplemented with a pinch of bovine pancreatic DNAse I or 1000 U/g cell pellet OmniCleave™ Endonuclease (Epicentre) and 2% *v/v* final Triton X-100, then agitation continued at 4 °C for a further 1 hour. Lysed cells were clarified by centrifugation at 40,000 g for 45 minutes at 4 °C, then the supernatant fraction incubated for 1 hour at 4 °C on a roller mixer with Nickel Sepharose Fast Flow 6 resin (GE Healthcare) pre-equilibrated with *E. coli* Lysis Buffer. In the case of Hsp90, higher target protein recovery was achieved by incubation with the resin overnight. After incubation, the supernatant-resin mixture was transferred to an Econo-Pac® Chromatography Column (BioRad), the flow through collected and reapplied to the resin once more. For the initial purifications, the resin was washed with 10 CV of *E. coli* Lysis Buffer, 15 x 1 CV of His Buffer A (25 mM Tris-HCl, 500 mM NaCl, 1 mM TCEP, 40 mM imidazole, pH 8.0) collected separately, then proteins eluted with 5 x 1 CV of His Buffer B (25 mM Tris-HCl, 500 mM NaCl, 1 mM TCEP, 250 mM imidazole, pH 8.0), also collected separately. The length and collection of washes was adapted as appropriate in subsequent purifications (e.g. shorter or longer washes, washes not collected as separate column volume fractions). The protein composition of wash and elution fractions was assessed by SDS-PAGE, then fractions containing semi-purified target protein were pooled together and dialysed against Dialysis Buffer (25 mM Tris-HCl, 1 mM TCEP, pH 8.0) using Snakeskin Dialysis Tubing (7K or 10K MWCO, Thermo Fisher Scientific) overnight at 4 °C with gentle stirring. The next morning, Dialysis Buffer was exchanged with fresh buffer and dialysis continued for a further hour. In the case of FGFR3-KD^{I538F}, up to 300 µL of 4.2 mg/mL Ulp1 (produced in house, see 2.5) was added to pooled fractions to cleave the N-terminal His6-SUMO tag during dialysis. On occasion, the N-terminal His6 tag of Hsp90 was cleaved similarly by incubation with HRV 3C protease during dialysis.

For His6-SUMO tag and Ulp1 separation from FGFR3-KD^{I538F}, and His6 tag and HRV 3C protease removal from Hsp90, reverse-IMAC was performed where the dialysed sample was applied to Nickel Sepharose Fast Flow 6 resin (GE Healthcare) pre-equilibrated with Dialysis Buffer, the flow through collected and the resin washed with Dialysis Buffer to recover all non-bound protein. Target proteins were further purified by the loading of

pooled reverse-IMAC flow through and wash fractions (in the case of FGFR3-KD^{I538F} and occasionally Hsp90) or dialysed samples (in the case of Cdc37 and Hsp90) onto a HiTrap Q HP column (5 mL, GE Healthcare) pre-equilibrated with Q Buffer A (25 mM Tris-HCl, 20 mM NaCl, 1 mM TCEP, pH 8.0). The column was washed with Q Buffer A until a baseline was reached then proteins eluted using a 0-50% gradient to Q Buffer B (25 mM Tris-HCl, 1 M NaCl, 1 mM TCEP, pH 8.0) over 20 CV, collected in 10 mL fractions. Alternatively, in the case of Hsp90, the column was washed with 30% Q Buffer B, then proteins eluted using a 30-50% gradient to Q Buffer B over 10 column volumes. After analysis by SDS-PAGE, fractions containing the target protein were pooled and concentrated using Vivaspin ultrafiltration units (10K or 30K MWCO as appropriate, PES, Sartorious). During concentration, proteins were exchanged into Storage Buffer (25 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, pH 8.0). Concentrations were estimated using absorbance at 280 nm and calculated molecular weight and extinction coefficients, then purified proteins (of ~1-2 mg/mL for FGFR3-KD^{I538F}, and 7-20 mg/mL for Cdc37 and Hsp90) were flash frozen in liquid nitrogen and stored at -80 °C.

2.7. Probing autophosphorylation of FGFR3-KD^{I538F}

FGFR3-KD^{I538F} (0.6 mg/mL) was incubated in Kinase Assay Buffer (25 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, 10 mM $MgCl_2$, 2 mM $NaVO_3$, pH 8.0) on ice for 1 hour then phosphorylation induced by addition of ATP (A2383, Sigma Aldrich) at a final concentration of 2.3 mM. The reaction was allowed to proceed for 1 hour at 4 °C on ice with samples taken at various time points. Reactions were quenched by dilution of the samples into 5 x SDS-PAGE sample buffer. Autophosphorylation was probed by SDS-PAGE followed by western blotting with an anti-Phospho-FGF Receptor (Tyr653/564) polyclonal antibody (#3471; Cell Signaling Technology) which detects phosphorylation of Tyr residues of the YYKK motif of FGFR A-loops (Tyr647/648 in FGFR3 IIIc). After development and imaging, membranes were stained with Ponceau Red stain to verify equal protein loading.

2.8. Verification of Cdc37 phosphorylation at Ser13

Cdc37 (0.48 mg/mL) was incubated in 1 x PMP buffer (50 mM HEPES-NaOH, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35, pH 7.5) (New England BioLabs) supplemented with 1 mM $MnCl_2$ at room temperature for 30 minutes then dephosphorylation initiated by addition of 200 units of λPP (New England BioLabs). The reaction mixture was incubated at 30 °C for 2 hours with samples taken at various time points. Reactions were quenched by addition of 5 x SDS-PAGE sample buffer. Phosphorylation of Ser13 was probed for by western blotting with an anti-Phospho-CDC37 (Ser13) (D8P8F) monoclonal antibody (#13248; Cell Signaling Technology) which detects phosphorylation of Cdc37 at this site.

After development and imaging, membranes were stained with Ponceau Red stain to verify equal protein loading.

2.9. Isothermal titration calorimetry

To probe the binding of FRS2α to FGFR3 juxtamembrane domain, the interaction between a peptide corresponding to this region and FRS2α$^{C9\text{-}122}$ PTB domain was investigated by isothermal titration calorimetry (ITC). For this analysis, the protein was dialysed into ITC buffer (50 mM sodium phosphate, 200 mM NaCl, 1 mM EDTA, 1 mM TCEP, pH 7.0) overnight at 4 °C and the peptide of FGFR3 juxtamembrane domain (residues 409-428, N-terminally acetylated and C-terminally amidated, >99% purity), synthesised by Dr Zsofia Hegedus (Department of Chemistry, University of Leeds), was prepared at 1 mg/mL (~430 μM) using the same dialysis buffer. Measurements were performed on a MicroCal iTC200 system (GE Healthcare Life Sciences) at 25 °C. A total of 20 injections (1 x 0.5 μL first injection then 19 x 2 μL injections) of titrant into the sample cell were completed, with a 180 second delay between injections to allow re-equilibration of the sample cell to baseline. The sample cell contained FRS2α$^{C9\text{-}122}$ at a nominal concentration of 10-20 μM and the syringe contained the FGFR3 juxtamembrane domain peptide titrant at ~430 μM. Control injections of buffer into buffer and peptide into buffer were completed also. During analysis, the control injection of peptide titrant into buffer, measuring the heat of dilution of the peptide, was subtracted from the raw data. The normalised data were fitted to a 'one site' model until convergence allowing the values N (stoichiometry), K_a (association constant) and ΔH (change in enthalpy) to vary, using Origin software (OriginLab).

2.10. Activity testing of full-length FGFR3

2.10.1. Stimulation assay in cells

Sf9 cells expressing 'wild-type', 'kinase-dead', and 'constitutively-active' FGFR3-5F1Y were assayed for the ability of the receptor to autophosphorylate at Y647, alone and following stimulation with ligands FGF1 and heparin. Cells were infected with baculoviruses prepared from DH10Bac™ cells and incubated at 27 °C and 120 rpm in Erlenmeyer flasks until reaching DPA+24. For each construct, cells were harvested by centrifugation at 2,000 g for 3 minutes, then cell pellets resuspended in either Insect-XPRESS™ Protein-free Insect Cell Medium with L-glutamine (Lonza) alone, or media containing 20 μg/mL heparin (H3149, Sigma Aldrich), 200 ng/mL human FGF1 (ab155608, Abcam), or 20 μg/mL heparin and 200 ng/mL human FGF1, all pre-warmed to 27 °C. Following resuspension, 500 μL aliquots of cells containing approximately 0.8 million cells were prepared in sterile 2 mL microcentrifuge tubes, then incubated at 27 °C and 700 rpm in a Thermomixer® C (Eppendorf). At time points of 0, 10 and

30 minutes after resuspension, cells were harvested by centrifugation at 2,000 g for 90 seconds, then cell pellets resuspended in cold lysis buffer (50 mM HEPES-NaOH, 200 mM NaCl, 1 mM TCEP, 1% *v/v* Nonidet P-40 Substitute, 1 mM $NaVO_3$, 1 mM NaF, 1 mM EDTA, 10% *v/v* glycerol, pH 7.5) and incubated on ice for 20-30 minutes. Following lysis, cell debris was pelleted by centrifugation at 14,000 g for 10 minutes at 4 °C, then the protein content of the supernatant (the clarified cell lysate) estimated using a Pierce™ BCA Protein assay kit (ThermoFisher Scientific). SDS-PAGE samples were prepared from the clarified cell lysate in 5 x SDS sample buffer, then 20 µg of total protein loaded per lane. Following separation of proteins by SDS-PAGE and transfer to PVDF membranes, samples were western blotted with anti-Phospho-FGFR1 (Tyr653/654), anti-mouse FGFR3 and anti-human GAPDH antibodies (Table 2.2). Membranes were stripped between probing with each antibody as detailed in 2.2.7. Uninfected Sf9 cells were used as a negative control and treated as for FGFR3-5F1Y expressing cultures. Anti-human GAPDH blotting and Ponceau Red staining were used to verify equal loading of total protein. The assay was completed as a technical triplicate.

2.10.2. *In vitro* autophosphorylation assay

Following membrane preparation and receptor purification, the kinase activity of FGFR3-5F1Y was assessed by probing its ability to phosphorylate itself as in 2.10.1. Generally, samples were incubated in the presence of $MgCl_2$ and $NaVO_3$, then ATP added to initiate kinase activity. At various time points before and following ATP addition, the reaction was quenched through addition of 5 x SDS-PAGE sample buffer to fractions of the reaction.

Initially, the autophosphorylation kinase activity of LMNG+CHS-purified FGFR3-5F1Y was assayed. The elution sample of this purification was incubated with 4 mM $NaVO_3$ and 20 mM $MgCl_2$ in FGFR3 His Elution Buffer for 10 minutes on ice, then kinase activity initiated by addition of 4 mM ATP. Samples were taken for SDS-PAGE analysis prior to ATP addition and following incubation on ice in the presence of ATP for 3 hours and 30 minutes. Autophosphorylation was probed through western blotting with an anti-phosphoFGFR1 (Y653/4) antibody (Table 2.2), and samples were submitted for protein and post-translational modification identification by mass spectrometry. As the YYKK motif of the activation loop of FGFR3-5F1Y contains a Y>F substitution (i.e. is rather YFKK), to confirm that the anti-phosphoFGFR1 (Y653/4) antibody is able to detect phosphorylation of this YFKK motif, blotting was also completed against a variant of FGFR2 kinase domain which also contains this YFKK motif but is known to be phosphorylated (Supplementary Figure 2). This protein is referred to as FGFR2-KD* and was kindly provided by Sasha Evans.

Next, activity of DDM+CHS-purified receptor was assayed. In this instance, the assay proceeded as with LMNG+CHS except 10 mM $NaVO_3$, 10 mM $MgCl_2$ and 4 mM ATP were used, and the reaction quenched and samples taken at 0, 20, 40, and 60 minutes. On this occasion, samples were western blotted with an anti-mouse FGFR3 antibody and an anti-phosphoTyr antibody (Table 2.2), then stained with Ponceau Red stain. Blotting with the anti-mouse FGFR3 antibody and Ponceau Red staining was used to verify equal loading between the lanes.

To investigate the response of DDM+CHS-purified receptor to ligand stimulation, the assay was repeated with untagged FGFR3-5F1Y (previously flash-frozen in liquid nitrogen for storage at -80 °C). In this instance, samples were first dephosphorylated with λPP (400 units, New England BioLabs) in 1 x PMP buffer (50 mM HEPES-NaOH, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35, pH 7.5) containing 1 mM $MnCl_2$ for 2 hours at room temperature. Following quenching of the phosphatase with 10 mM $NaVO_3$, kinase activity was initiated by the addition of 7.8 mM $MgCl_2$ and 3.1 mM ATP. Samples were taken prior to and after λPP addition, then at 0, 20, 40, and 60 minutes following ATP addition. For ligand stimulation, the receptor (an estimated 8 µg) was incubated with 4 µg of human FGF1 and 2 µg of heparin oligosaccharide dp8. This approximates to 2.8 x molar excess of FGF1 and 9.3 x molar excess of dp8. In parallel, the same assay with and without ligand stimulation was completed using prepared membranes containing tagged FGFR3-5F1Y; in this case, ligand concentrations were an estimated 15 x molar excess of FGF1 and 50 x molar excess of dp8. FGFR2-KD* (4.5 µg) was used as an assay positive control and treated similarly except without ligand stimulation. In all cases, samples were western blotted with anti-phosphoFGFR1 (Y653/4), anti-mouse FGFR3, and anti-phosphoTyr antibodies (Table 2.2) before staining with Ponceau Red stain. Blotting with the anti-mouse FGFR3 antibody and Ponceau Red staining was used to verify equal loading between the lanes.

Lastly, the activity of DIBMA-purified receptor was assayed. On this occasion, the elution fractions of ‘apo’ and ‘complexed’ receptors were first dephosphorylated with λPP (200 units, New England BioLabs) in 1 x PMP buffer containing 1 mM $MnCl_2$ at room temperature for 30 minutes. Following inhibition of phosphatases with 10 mM $NaVO_3$, kinase activity was initiated through addition of 7.8 mM $MgCl_2$ and 3.1 mM ATP. Samples were taken prior to and after λPP addition, then at 0, 15, 23, 30, and 40 minutes after ATP addition. FGFR2-KD* (4.5 µg) was treated similarly as a positive control. Samples were western blotted using anti-phosphoFGFR1 (Y653/4), anti-mouse FGFR3, and anti-phosphoTyr antibodies (Table 2.2), then blots stained with Ponceau Red stain. Blotting with the anti-mouse FGFR3 antibody and Ponceau Red staining was used to verify equal loading between the lanes.

2.11. NMR of FGFR3 kinase domain and Cdc37 binary complex interactions

2.11.1. Preparation of proteins for NMR

All proteins used in NMR experiments detailed in this thesis were prepared by Dr Tom Bunney (University College London) as detailed in Bunney *et al.* (322). In brief, non-labelled Cdc37 was expressed and purified as in 2.5.4 and 2.6.4, and uniformly ^{15}N-labelled or selectively unlabelled FGFR3 kinase domains expressed with minor adaptions to protocols as in 2.5.3 and purified as in 2.6.4. To obtain uniformly ^{15}N-labelled kinase domains, *E. coli* were cultured in minimal media supplemented with 1 g/L of ^{15}N-labelled ammonium sulphate. To obtain selectively-unlabelled kinase domains with a uniform ^{15}N-labelled background, *E. coli* were cultured similarly except the minimal media was supplemented both with 1 g/L of ^{15}N-labelled ammonium sulphate and 1 g/L of the relevant selected amino acids (that were not labelled, i.e. ^{14}N). Five versions of selectively-unlabelled wild-type FGFR3 kinase domain (FGFR3-KD^{WT}) were prepared: those with unlabelling of the amino acid pairs Gln/Ile, Asn/Arg, Lys/Leu or Phe/Val, and one sample with unlabelling at the amino acid Trp alone. FGFR3-KD^{WT} was expressed from a pOPINS vector, with *E. coli* cultured as for FGFR3 kinase domain containing the I538F substitution (FGFR3-KD^{I538F}) except that protein expression was induced with 0.1 mM IPTG only (i.e. no rhamnose). Following purification by IMAC and ion exchange chromatography as in 2.6.4, all proteins were further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Healthcare) pre-equilibrated with Storage Buffer (25 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, pH 8.0).

2.11.2. General NMR methods

Protein samples for analysis of binary complex interactions between FGFR3 kinase domain and Cdc37 were prepared in NMR Buffer (50 mM PIPES-NaOH, 50 mM NaCl, 5 mM TCEP, 1 mM EDTA, pH 7.0 containing 5% *v/v* D_2O). ^{1}H-^{15}N 2D TROSY HSQC spectra of uniformly ^{15}N-labelled wild-type FGFR3 kinase domain (FGFR3-KD^{WT}) and FGFR3 kinase domain containing the I538F substitution (FGFR3-KD^{I538F}), each individually titrated with non-labelled Cdc37, were acquired by Dr Gary Thompson (University of Leeds) and provided for analysis. All others were acquired during this project by myself. Samples were approximately 200-300 µL in volume and of 77-230 µM kinase domain concentration, prepared in 5 mm Shigemi tubes. NMR spectra were recorded on Bruker Avance III HD 750 and 950 MHz spectrometers equipped with 5 mm TCI z-axis gradient cryoprobes, at temperatures ranging between 10 °C and 25 °C. Standard TROSY-detected HSQC experiments with water flip-back and WATERGATE pulses (323) were recorded as detailed previously. All data were processed using NMRPipe and NMRDraw (324) and analysed with CPPNMR Analysis (325).

2.11.3. CSP analysis of I538F substitution in FGFR3 kinase domain

To investigate the effect of the I538F substitution on kinase structure, a CSP analysis was completed comparing ^{1}H-^{15}N TROSY HSQC spectra of *apo* FGFR3-KD^{WT} and FGFR3-KD^{I538F}. To enable this analysis, assignments of *apo* FGFR3-KD^{WT} were transferred where unambiguous to FGFR3-KD^{I538F}. As HSQC spectra of these *apo* kinase domains were recorded at different temperatures to achieve the best quality data (at 25 °C for FGFR3-KD^{WT} and 10 °C for FGFR3-KD^{I538F} due to its lower thermal stability), additional spectra were acquired for wild-type kinase domain at 20 °C and 15 °C, and of the I538F variant at 15 °C to enable assignment transfer. Peak assignments of *apo* FGFR3-KD^{WT} (75% of non-Pro residues) were kindly provided by Dr Domenico Sanfelice (University College London) and were determined using spectra collected at 25 °C. These peak assignments are now published (326) and available in the BMRB (entry #27082) . Following assignment transfer, CSP analysis was completed using spectra acquired for each kinase at 15 °C. CSPs were calculated for peaks that were mutually assigned in each spectrum, quantified as a combined chemical shift ($\Delta\delta_{av}$) (327) where:

$$\boldsymbol{\Delta\delta_{av}} = \sqrt{(\boldsymbol{\delta H_{WT/I538F}})^2 + \mathbf{0.14}(\boldsymbol{\delta N_{WT/I538F}})^2} \qquad (2.1)$$

In this equation, $\delta H_{WT/I538F}$ and $\delta N_{WT/I538F}$ denote the changes in chemical shift observed in the proton and nitrogen dimensions for each assigned residue between spectra of *apo* FGFR3-KD^{WT} and FGFR3-KD^{I538F}. Significant perturbations were identified as those with values greater than the mean $\Delta\delta_{av}$. Note that this analysis also appears in Bunney *et al.* (322) but is repeated here. Notably, these analyses differ in their reported magnitude of perturbations in the DFG latch of the kinase, as these large perturbations were more difficult to identify through assignment transfer from FGFR3-KD^{WT} to FGFR3-KD^{I538F}, and thus not included here. The CSP calculation in the published study was completed by Dr Domenico Sanfelice (University College London).

2.11.4. Probing FGFR3 kinase domain-Cdc37 binary complex formation

To probe the binding interaction between FGFR3 kinase domain and Cdc37, uniformly ^{15}N-labelled FGFR3-KD^{WT} was titrated with non-labelled Cdc37 at various molar ratios (1:0, 1:0.7, 1:1.9 and 1:3.0, FGFR-KD^{WT}:Cdc37) and ^{1}H-^{15}N TROSY-HSQC spectra acquired at 25 °C for each titration point. Titration of FGFR3-KD^{I538F} with Cdc37 was completed in the same manner except molar ratios of 1:0, 1:0.25 and 1:0.5 (FGFR3-KD^{I538F}:Cdc37) were used and spectra recorded at 10 °C (a lower temperature of acquisition was used here due to the poor sample stability of FGFR3-KD^{I538F}). Note that these data were recorded by and provided by Dr Gary Thompson (University of Leeds).

As temperature can influence HSQC peak positions, to allow comparison between FGFR3-KD^{WT} and FGFR3-KD^{I538F} spectra, additional spectra were acquired of a 1:0.7 ^{15}N-labelled FGFR3-KD^{WT}:non-labelled Cdc37 complex at 20 °C and 15 °C, and of the ^{15}N-labelled FGFR3-KD^{I538F}:non-labelled Cdc37 complex at 15 °C.

Apo FGFR3-KD^{WT} backbone amide assignments were used for CSP mapping of binary complex formation. CSPs were calculated using a conservative chemical shift mapping approach where backbone assignments were transferred when unambiguous between *apo* FGFR3-KD^{WT} and FGFR3-KD^{WT}-Cdc37 complex spectra, then CSPs calculated for those mutually assigned peaks, quantified as a combined chemical shift ($\Delta\delta_{av}$) (327) where:

$$\boldsymbol{\Delta\delta_{av} = \sqrt{(\delta H_{apo/complex})^2 + 0.14(\delta N_{apo/complex})^2}} \qquad (2.2)$$

In this equation, $\delta H_{apo/complex}$ and $\delta N_{apo/complex}$ denote the changes in chemical shift observed in the proton and nitrogen dimensions for each assigned residue between spectra of *apo* FGFR3-KD^{WT} and of Cdc37-bound FGFR3-KD^{WT}. As discussed in 4.2.2, no perturbations were deemed to be significant.

In the Cdc37 titration series for both WT and I538F variants of FGFR3 kinase domain, sharp, complex-dependent peaks appeared in the centre of the HSQC spectra. These peaks were assigned in a corroborative manner: using standard triple-resonance experiments to obtain resonance spin systems and backbone connectivity (Dr Domenico Sanfelice, University College London), through prediction of random coil chemical shits of FGFR3-KD^{WT} backbone amide resonances with comparison to those observed experimentally using an approach detailed previously (328) (Dr Gary Thompson, University of Leeds), and through the use of a selective unlabelling strategy by myself to inform on the amino acid identity of these complex-dependent sharp resonances, as detailed in this thesis.

2.11.5. Using selective unlabelling to aid resonance assignments

To aid in backbone amide assignment of the complex-dependent sharp peaks (and for *apo* FGFR3-KD^{WT} assignment), 1:1 FGFR3-KD^{WT}:Cdc37 complexes with selective unlabelling of the kinase domain for QI, NR, LK, and FV pairs of residues, and of W alone, were prepared (Dr Tom Bunney, University College London). ^{1}H-^{15}N TROSY HSQC spectra of these samples and a uniformly ^{15}N-labelled FGFR3-KD^{WT}:Cdc37 (1:0.75) sample were acquired at 25 °C. Each sample had a nominal 100 μM kinase concentration. To account for variability in peak intensity due to slight concentration differences, selectively-unlabelled spectra were normalised with respect to the reference (uniformly-labelled) spectrum using the four most intense glycine resonances (G481,

G553, G549 and G574) as reference peaks; note that glycine residues should be unaffected by metabolic scrambling with the selectively unlabelling pairs used (311,313).

Following spectra normalisation, to assess the efficiency of unlabelling, the intensity of relevant assigned peaks in spectra acquired of the reference and selectively-unlabelled samples (i.e. Lys and Leu peaks for the -KL sample) were compared by dividing the intensity of those in the unlabelled spectrum with those of the reference spectrum; this obtained a peak intensity ratio where a value of 1 equates to equivalent intensities in each spectrum, and a value of 0 equates to complete unlabelling at this position. To account for stochastic noise in the data which could lead to the identification of false positive unlabelling, confidence thresholds were calculated to compare these peak intensity ratios against. To do so, expected peak intensity ratios of 1 were simulated for peaks with signal-to-noise ratios of 30, 20, 10 and 8:1, then these used to assess how stochastic noise can influence the observed ratio (Figure 2.2). These simulations showed that peaks with lower signal-to-noise are (as expected) more greatly influenced by stochastic noise than those with higher signal-to-noise. Confidence thresholds were calculated from the simulation data as the mean peak intensity ratio minus 3 standard deviations; these place a stricter requirement on peaks with low signal-to-noise in the HSQC spectra, requiring greater reductions in peak intensity to be considered as genuine unlabelling. When analysing the efficiency of unlabelling, peaks were classified as having signal-to-noise ratios of up to 30, 20, 10 or 8:1, then their calculated peak intensity ratio from the reference and selectively-unlabelled spectra compared against the appropriate confidence threshold.

After assessing the efficiency of selective-unlabelling for each sample using assigned resonances, these analyses were repeated for the complex-dependent sharp peaks to identify their residue type.

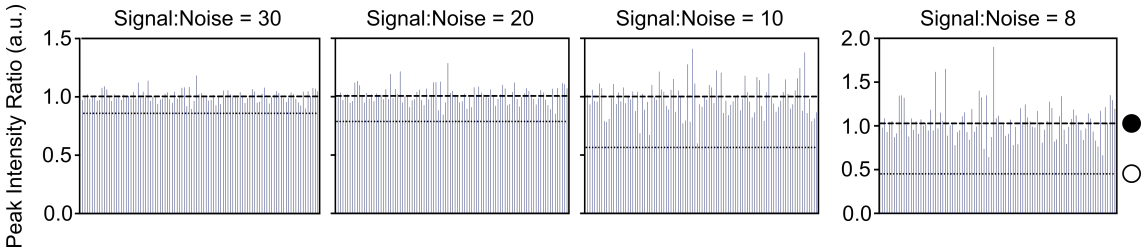


Figure 2.2. Calculation of confidence thresholds for selective unlabelling.
Simulations of the effect of different signal to noise ratios (30, 20, 10 and 8:1) on peak intensity ratios with an expected value of 1. The mean peak intensity (filled circle) and the calculated confidence threshold (open circle) corresponding to the mean minus three standard deviations are shown as dashed lines in each case.

2.11.6. Probing the effect of PD173074 on kinase-Cdc37 complex formation

To investigate the influence of kinase inhibitors on the formation and stability of the binary complex, PD173074 (Pfizer) was titrated into a pre-formed FGFR3-KD^{WT}:Cdc37 complex (1:0.7, 130 μM kinase) at various molar equivalents with respect to the kinase (0.3, 0.5, 1.0 and 2.0), and ^{1}H-^{15}N TROSY HSQC spectra were recorded at each titration point. PD173074 was added from a stock solution in DMSO; the final DMSO concentration was 1.26% *v/v*. Titration spectra were compared to ^{1}H-^{15}N TROSY HSQC spectra of *apo* FGFR3-KD^{WT}, a FGFR3-KD^{WT}:Cdc37 complex, and of PD173074-bound FGFR3-KD^{WT} (with 2 molar equivalents of PD173074). Additionally, spectra were acquired for FGFR3-KD^{WT}:Cdc37:PD173074 (1:1:2) complexes where PD173074 and Cdc37 were added in opposing orders with an incubation step between each addition (these samples were prepared by Dr Tom Bunney). Assignments of *apo* FGFR3-KD^{WT} and PD173074-bound FGFR3-KD^{WT} were used in analysis of the data (BMRB entries 27082 and 27083) (326). CSPs were calculated using a conservative chemical shift mapping approach as before, quantified as a combined chemical shift ($\Delta\delta_{av}$) (327) where:

$$\boldsymbol{\Delta\delta_{av}} = \sqrt{(\boldsymbol{\delta H_{apo/PD173074}})^2 + \mathbf{0.14}(\boldsymbol{\delta N_{apo/PD173074}})^2} \qquad (2.3)$$

In this equation, $\delta H_{apo/PD173074}$ and $\delta N_{apo/PD173074}$ denote the changes in chemical shift observed in the proton and nitrogen dimensions for each assigned residue between spectra of *apo* FGFR3-KD^{WT} and of PD173074-bound FGFR3-KD^{WT}. Significant perturbations were identified as those with values greater than the mean $\Delta\delta_{av}$.

2.12. Reconstitution of kinase-bound Hsp90 ternary complexes

2.12.1. Analysis of reconstituted complexes by size exclusion chromatography

30-50 μL samples were prepared by mixing of FGFR3-KD^{I538F}, Cdc37 and Hsp90 at various molar ratios in SEC Buffer (25 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, pH 8.0). Following incubation on ice for 10 minutes, samples were spun by centrifugation at 10,000 g for 10 minutes at 4 °C, then injected onto either a Superdex 200 Increase 5/150 GL column (GE Healthcare) monitored by a Prominence HPLC System (Shimadzu) at room temperature, or a Superdex 200 Increase 10/300 GL column (GE Healthcare) monitored by an ÄKTA pure chromatography system (GE Healthcare) at 4 °C, both which were pre-equilibrated in SEC Buffer. Injection volumes were of 10 or 20 μL for the Superdex 200 Increase 5/150 GL column and 45 μL for the Superdex 200 Increase 10/300 GL column. Each column was run at between 0.25 and 0.35 mL/min.

To saturate Hsp90 into ternary complexes, different ratios of components were tested; saturation was achieved with samples containing an equimolar concentration of each component, corresponding to a 2-fold excess of kinase and Cdc37 with respect to

dimeric Hsp90. To aid analyses, each individual component and an equimolar binary kinase-Cdc37 complex were separately injected. Sample concentrations varied between 20 and 60 µM during optimisation before choosing equimolar complex mixtures of either 20 or 25 µM.

Ternary complexes were initially reconstituted without any ligands present. To probe the effect of ATP or the Hsp90 inhibitor PU-H71 on complex reconstitution, equimolar samples containing 20 µM FGFR3-KD^{I538F}, Cdc37 and Hsp90 were prepared, then either ligand added to a final concentration of 100 µM (i.e. 5-fold excess, note that the ATP sample also contained 1 mM $MgCl_2$). After incubation on ice for 10 minutes, the samples were spun by centrifugation at 20,000 g for 5 minutes at 4 °C, then injected onto a Superdex 200 Increase 10/300 GL column as before.

2.12.2. Estimation of complex masses using SEC-MALS

In some cases, size-exclusion chromatography was coupled to a DAWN 8+ multi-angle light scattering (MALS) detector and an Optilab T-rEX differential refractive index (dRI) detector (Wyatt Technology) for MALS measurements. Molar masses were calculated using ASTRA 6.1 software (Wyatt Technology) using protein concentration estimations derived either from dRI (change in refractive index) signal or from calculated extinction coefficients, as required.

2.12.3. Isolation of ternary complexes by GraFix

A 10-30% *v/v* glycerol and 0-0.15% *v/v* glutaraldehyde linear gradient in GraFix Buffer (50 mM HEPES-NaOH, 100 mM NaCl, 1 mM TCEP, pH 7.5) was prepared by layering 2.1 mL of GraFix Buffer containing 10% *v/v* glycerol on top of 2.1 mL of GraFix Buffer containing 30% *v/v* glycerol and 0.15% *v/v* glutaraldehyde in SW 55 Ti centrifuge tubes (Beckman Coulter). After sealing with parafilm, the tubes were laid horizontally for 2 hours at room temperature then carefully returned to a vertical orientation and cooled to 4 °C for a further hour. Prior to sample application, a 200 µL cushion of 7% *v/v* glycerol in GraFix Buffer was applied on top of the linear gradient.

Reconstituted complexes were prepared as for size exclusion chromatography, then diluted in SEC buffer to obtain 200 µL samples containing a maximum of 200 pmol of the complex (as recommended by the GraFix manual (261)). These samples were pipetted on top of the cushion, then the tubes centrifuged at 40,000 rpm for 16 hours at 4 °C using a SW55 Ti rotor (Beckman Coulter). After centrifugation, the linear gradient was separated by removing 200 µL fractions taken from the top of the gradient. Excess glutaraldehyde was quenched with either 80 mM glycine pH 8.0 or 20 mM Tris pH 8.0. To aid analyses, complexes were also applied to control linear gradients prepared without glutaraldehyde and centrifuged in parallel. Fractions containing ternary

complexes were identified by SDS-PAGE analysis of crosslinked and control samples. GraFix was used to help optimise complex reconstitution to saturate Hsp90.

2.12.4. Concentration of ternary complexes for electron microscopy

SEC and GraFix fractions containing ternary complexes were of sufficient concentrations for direct use to prepare negative-stain EM grids. To prepare cryo-EM grids, these fractions required concentration. GraFix fractions also required buffer exchange to remove glycerol and quenched glutaraldehyde that would hamper contrast. Buffer exchange of GraFix samples into GraFix buffer lacking glycerol was tried in several ways: by dialysis, using Zeba™ Spin desalting columns (Thermo Fisher Scientific), on an ultrafiltration unit (see below), or a combination of these.

Concentration of GraFix and SEC fractions was also tried in several ways. Fractions were generally concentrated using ultrafiltration devices of different design, membrane type and molecular weight cut off; devices tested included Vivaspin ultrafiltration units (30K and 100K MWCO, PES membranes), Proteus X-spinner units (100K MWCO, PES membrane, Generon), and Amicon Ultra units (100K MWCO, regenerated cellulose membrane). Alternatively, fractions were rebound to Nickel Sepharose Fast Flow 6 resin (GE Healthcare) and eluted with His Buffer B in a smaller volume. During these processes, the absorbance at 280 nm was monitored to estimate protein concentration using a calculated molar extinction coefficient for the complex, and samples were analysed by SDS-PAGE and Native-PAGE.

2.13. Negative-stain electron microscopy

2.13.1. Grid preparation

2.13.1.1. Full-length FGFR3

3 µL aliquots of DDM- or DIBMA-purified FGFR3-5F1Y, *apo* or 'ligand-bound', were applied to glow-discharged carbon-coated copper grids and stained with 2% *w/v* uranyl acetate solution. Samples were allowed to adsorb onto the carbon for 30 seconds before blotting and twice staining with 3 µL 2% *w/v* uranyl acetate solution for 45 to 60 seconds with the side blotting of excess solution between each application. After a final blotting, grids were air dried under a lamp. Due to the small-scale nature of preparations, elution or unbound fractions from purifications were used directly to prepare grids at several dilutions in FGFR3 homogenisation buffer.

2.13.1.2. Kinase-bound Hsp90 ternary complexes

3 µL aliquots of ternary complex purified by GraFix were applied to glow-discharged carbon-coated copper grids and stained with 2% *w/v* uranyl acetate solution as above.

Grids were prepared of ternary complexes reconstituted in the absence of additional ligands (*apo* ternary complex) and of complexes prepared with PU-H71.

2.13.2. Data collection and image processing

Grids were screened and micrographs recorded for image processing on a Tecnai T12 (FEI) electron microscope operating at 120 KeV equipped with an Ultrascan 4000 CCD (Gatan) camera. Micrographs were recorded at a nominal 23,000x magnification yielding a pixel size of 4.8 Å/pixel. Alternatively, micrographs were recorded on a Technai F20 (FEI) electron microscope operating at 200 KeV equipped with CETA CMOS CCD (FEI) camera. In this case, micrographs were recorded at a nominal 50,000x magnification yielding a pixel size of 2 Å/pixel. In both cases, micrographs were recorded using a defocus range of -1 to -3 µm. All data were processed using RELION 2.1 or RELION 3.0 (278) and models were visualised using UCSF Chimera (326).

2.13.2.1. Full-length FGFR3

Micrographs of DIBMA-purified FGFR3-5F1Y in the presence and absence of the ligands FGF1 and heparin oligosaccharide dp8 (see 2.6.1.3) were recorded on the Technai T12 (FEI) electron microscope. In each case, approximately 1,000 particles were manually picked and subjected to 2D classification, then micrographs autopicked using these classes as templates. For the unliganded receptor, 10,673 particles were autopicked from 22 micrographs, then subjected to three rounds of 2D classification, discarding poor particles at each round that produced poorly defined classes. The final round of classification contained 3,433 particles, classified into 20 classes with a 200 Å mask. For the liganded receptor, 18,168 particles were autopicked from 18 micrographs, then also subjected to three rounds of 2D classification. The final round of classification contained 4,690 particles, classified into 30 classes with a 200 Å mask.

2.13.2.2. Kinase-bound Hsp90 ternary complexes

A total of 24 micrographs were recorded on the Technai F20 (FEI) electron microscope of the *apo* ternary complex. From these micrographs, 8,470 particles were manually picked then subjected to three rounds of reference-free 2D classification, discarding poor particles at each round. A final 2,208 particles were used to generate an initial *ab initio* 3D model, then these particles classified into two 3D classes using the initial model low pass filtered to 60 Å as a template. Each class (containing 1,290 and 918 particles respectively) was next individually refined. To improve their quality, each map was further focused by applying soft masks generated from the obtained models, then refinement continued. This obtained two final maps with estimated resolutions of 21 Å and 23 Å according to the gold-standard FSC 0.143 criterion. In all cases, no symmetry was

applied during processing. To interpret the obtained maps, rigid docking of a single protomer of Hsp90 (PDB entry 5FWL) was completed manually using Chimera (USCF).

For the PU-H71-bound ternary complex, 10 micrographs were recorded on the Technai T12 (FEI) electron microscope. From these micrographs, 6,915 particles were manually picked then subjected to three rounds of reference-free 2D classification. A final 3,306 particles were used to generate two initial *ab initio* 3D models, then classified into two 3D classes using the highest occupancy initial model low-pass filtered to 60 Å as a template. As the resulting 3D classes were very similar, all 3,306 particles were refined into a final 3D reconstruction with an estimated resolution of 31 Å according to the gold-standard FSC 0.143 criterion. No symmetry nor masking was applied during processing.

For comparison with the generated low resolution maps of *apo* and PU-H71-bound ternary complexes, the cryo-EM map of the Cdk4:Cdc37:Hsp90 ternary complex (PDB entry 5FWL) was low-pass filtered to 25 Å using RELION Image Handler.

2.14. Cryo-electron microscopy

2.14.1. Grid preparation and data collection

Cryo-EM grids were prepared using a Vitribot™ Mark IV (FEI Company) at 4 °C and 100% humidity. Generally, 3 µL aliquots of ternary complex were applied to glow-discharged grids, blotted for 6 seconds with a blot force of 6, then grids plunge frozen in liquid ethane cooled with liquid nitrogen. While Quantifoil® R1.2/1.3 300 mesh copper grids were generally used, a variety of grid types were tested during grid optimisation including: Quantifoil® R2/2 200 mesh and Quantifoil® R2/4 400 mesh copper grids, and Lacey carbon-backed 400 mesh copper grids. In the latter case, 3 µL samples were applied to the glow-discharged grid five times with 30 second incubations and side blotting between applications, then 3 µL of GraFix buffer applied before blotting and plunge freezing with the Vitribot™ as above. During grid optimisation, addition of the surfactants CHAPSO (at 8 mM, i.e. at its CMC) and DDM (at 0.085 mM, i.e. below its CMC) to samples was also tested.

Cryo-EM grids were screened on a Titan Krios G3i (FEI) electron microscope operating at 300 KeV, equipped with either a Falcon E3C (FEI) or an energy-filtered K2 XP summit (Gatan) direct electron detector. Datasets were acquired on the former using EPU software (FEI) for automated data collection with the detector operating in integrating mode. Cryo-EM datasets were collected for the GraFix-ed ternary complex (undetermined concentration, see 5.2.5), or the ternary complex (with no crosslinking) at 0.5 mg/mL in the presence of 0.085 mM DDM. For the GraFix-ed complex, micrographs were recorded at a nominal x75,000 magnification yielding a pixel size of 1.065 Å and at a defocus range of -1.5 to -3.5 µm. Each micrograph was dose-fractionated into 59

fractions with a dose per fraction of 1.2 $e^-/Å^2$, and a total dose of 71.9 $e^-/Å^2$ over a total exposure time of 1.5 seconds. Micrographs were recorded in the same manner for the non-crosslinked complex, except that each micrograph was dose-fractioned into 59 fractions with a dose per fraction of 1.34 e^-/A^2, and a total dose of 78.94 $e^-/Å^2$ over a total exposure time of 1.5 seconds.

2.14.2. Data processing

As with negative-stain EM, all data were processed using RELION 2.1 or RELION 3.0 (278) and models visualised using UCSF Chimera (326).

A total of 1,011 micrograph stacks were recorded of the GraFix-purified *apo* complex. Following motion correction and CTF estimation on-the-fly using RELION 2.1's implementation of MotionCorr and CtfFind, 5,918 particles were manually picked and subject to 2D classification. These 2D classes were used as templates to autopick 136,917 particles from the 1,011 motion-corrected micrographs. Following two rounds of reference-free 2D classification, discarding poor particles at each round that produced poorly-defined classes, there appeared to be two distinct groups of classes: one group with classes containing Hsp90 dimer-like 'V' features (31,288 particles), and a second group with classes containing more globular features (74,589 particles). These two groups were separated and particles from the latter (believed to be the ternary complex) were subjected to two further rounds of 2D classification. The final classification contained 66,355 particles, classified into 60 classes with a 135 Å mask, imposing a 15 Å maximum resolution. These images were not processed further.

A total of 1,263 micrograph stacks were recorded of the non-crosslinked complex prepared in the presence of 0.085 mM DDM. Micrographs were motion corrected and their CTF estimated on-the-fly using RELION 3.0's implementation of MotionCorr and Gctf, then 4,324 particles manually picked to generate 2D classes to use as templates in autopicking. This yielded a total of 204,792 particles that were binned two-fold, yielding a pixel size of 2.137 Å/pixel. Following two rounds of reference-free 2D classification, discarding poor particles after each round as above, a final 21,513 particles were re-extracted at their original pixel size (1.065 Å/pixel), then classified into 25 final 2D classes with a 150 Å mask. After generating an initial *ab initio* 3D model, these 21,513 particles were 3D classified into two classes while imposing a 20 Å maximum resolution; the two obtained classes were similar and had a fairly equal distribution of particles (54% and 46%, respectively). For comparison, one of these obtained 3D classes was manually fitted into the low-resolution 21 Å map of the *apo* complex from negative-stain EM analysis using Chimera (USCF).

2.15. Native mass spectrometry

All native mass spectrometry data were acquired and processed by Dr Rachel George (Faculty of Biological Sciences Mass Spectrometry Facility, University of Leeds).

60-100 µL samples of reconstituted ternary complexes at either 15 or 20 µM concentration were prepared by equimolar mixing of Hsp90, Cdc37 and FGFR3-KD^{I538F} in SEC buffer. After incubation on ice for 10 minutes, samples were centrifuged at 10,000 g for 10 minutes. A 25 µM sample of Hsp90 alone was prepared similarly. These samples were provided to the FBS Mass Spectrometry Facility for analysis. Briefly, samples were buffer exchanged into 0.2 M ammonium acetate using Zeba Spin 7K MWCO desalting columns (Thermo Fisher Scientific), then samples analysed by nano-electrospray ionisation mass spectrometry using gold- and palladium-coated nanospray tips prepared in-house on a Q-Exactive Plus quadrupole orbitrap mass spectrometer (Thermo Scientific) modified for transmission of high mass over charge (*m/z)* ions. The instrument was operated by Tune software version 2.10 provided with the instrument. Mass calibration up to *m/z* of ~12,000 was performed using 2 mg/mL CsI clusters in an isopropyl alcohol:water mixture. Data was processed using Xcalibur Qual Browser v4.0.27.19 and UniDec v2.7.1.

To explore the effect of the Hsp90 inhibitor PU-H71 on complex composition, the inhibitor was added at 5-molar excess before and during exchange of the complex sample into ammonium acetate, then analysed as above.

Chapter 3. Structural studies of full-length FGFR3 and variants

3.1. Aims

To address the paucity of structural information regarding RTK activation in the context of full-length receptors, I aimed to solve a high-resolution structure of full-length FGFR3. To accomplish this, a variety of full-length FGFR3 constructs including 'wild-type' and oncogenic variants of FGFR3 would be cloned for insect cell expression using the BEVS approach. After establishing and optimising expression, to obtain protein samples suitable for structural characterisation, the receptor would need to be purified; these purification protocols would also require optimisation. Building on prior attempts in the RTK field, efforts would be focused on using electron microscopy for receptor structure determination. Both *apo* and ligand-complexed forms of the receptor would be considered in aim to provide insights regarding activation mechanisms. To ensure the physiological relevance of any gained structural insights, it would also be important to demonstrate that expressed and purified constructs of full-length FGFR3 are indeed active. Anticipating that flexibilities in the receptor may be unfavourable for structure determination, strategies to minimise this would also be explored.

3.2. Results

3.2.1. Cloning of full-length FGFR3 and variants

Having hypothesised that conformational heterogeneity of the receptor could be problematic for high-resolution structural studies, measures were introduced during construct design to try and counteract this. To maintain functional relevance of the generated constructs, these measures were drawn from naturally-occurring pathological variants of FGFR3 where possible.

The phosphorylation status of the receptor kinase domain, ranging from a non-phosphorylated to a multi-phosphorylated state reflecting the activity status of the receptor, acts as one source of receptor heterogeneity. To minimise sample variation, a construct of FGFR3 IIIb was designed wherein only one of the phosphorylatable tyrosine residues of the receptor kinase domain was retained. This residue, Y649 of the YYKK motif of the activation loop, is analogous to Y653 of FGFR1, a residue which is essential and sufficient for FGFR1 kinase activity *in vitro* (51). Presuming conservation between FGFR1 and FGFR3, the remaining residues (Y579, Y650, Y726, Y762 and Y772) of FGFR3 are dispensable for kinase activity and thus were substituted with phenylalanine residues to prevent their phosphorylation (Figure 3.1A). This construct will be referred to herein as FGFR3-5F1Y and serves as the base of all other full-length FGFR3 constructs described. Although both Y649 and Y650 are required for full kinase activity, Y650 is not essential and was thus not retained (51).

Active-inactive transitions pose a further source of sample heterogeneity, whether from a monomer to dimer transition, or from conformational rearrangements of receptor dimers. Constructs were therefore designed to encourage receptors to exist in active, dimeric states, using pathological mutations and gene fusions as their basis (Figure 3.1 and Table 3.1). Fusion constructs would also provide an additional benefit for structural studies by single-particle EM by increasing the mass of the receptor. A truncated construct, 'del19', which also leads to receptor activation (89) was additionally generated; although this is smaller than the full-length receptor, understanding the structural mechanisms behind receptor activation through truncation would also be informative. Lastly, constructs containing the substitutions of either K652E (49) or K510A (317), which alter the activity status of the kinase domain, were generated (Table 3.1).

Table 3.1. Constructs of full-length FGFR3-5F1Y.

Construct	Description
'Wild-type'	FGFR3 IIIb containing Y579F, Y650F, Y726F, Y762F and Y772F substitutions*
+K652E	Substitution in the kinase domain resulting in constitutive kinase activity
+K510A	Substitution in the kinase domain resulting in loss of kinase activity (i.e. kinase-dead)
+C228R	Substitution in the extracellular domain
+S249C	Substitution in the extracellular domain
+Y375C	Substitution at the interface between the extracellular domain and the transmembrane domain
+RT4	C-terminal gene fusion of TACC3 (144-838)
+RTT112	C-terminal gene fusion of TACC3 (648-838)
+deldel	C-terminal gene fusion of TACC3 (144-663)
+SW780	C-terminal gene fusion of BAIAP2L1 (18-511)
+del19	Truncation of receptor through deletion of exon 19

Note: all constructs contain this 5F1Y backbone (*); BAIAP2L1 (brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1); TACC3 (transforming acidic coiled-coil containing protein 3); the residues of the fusion protein present in each construct are given in brackets. FASTA sequences of constructs can be found in Appendix 2.

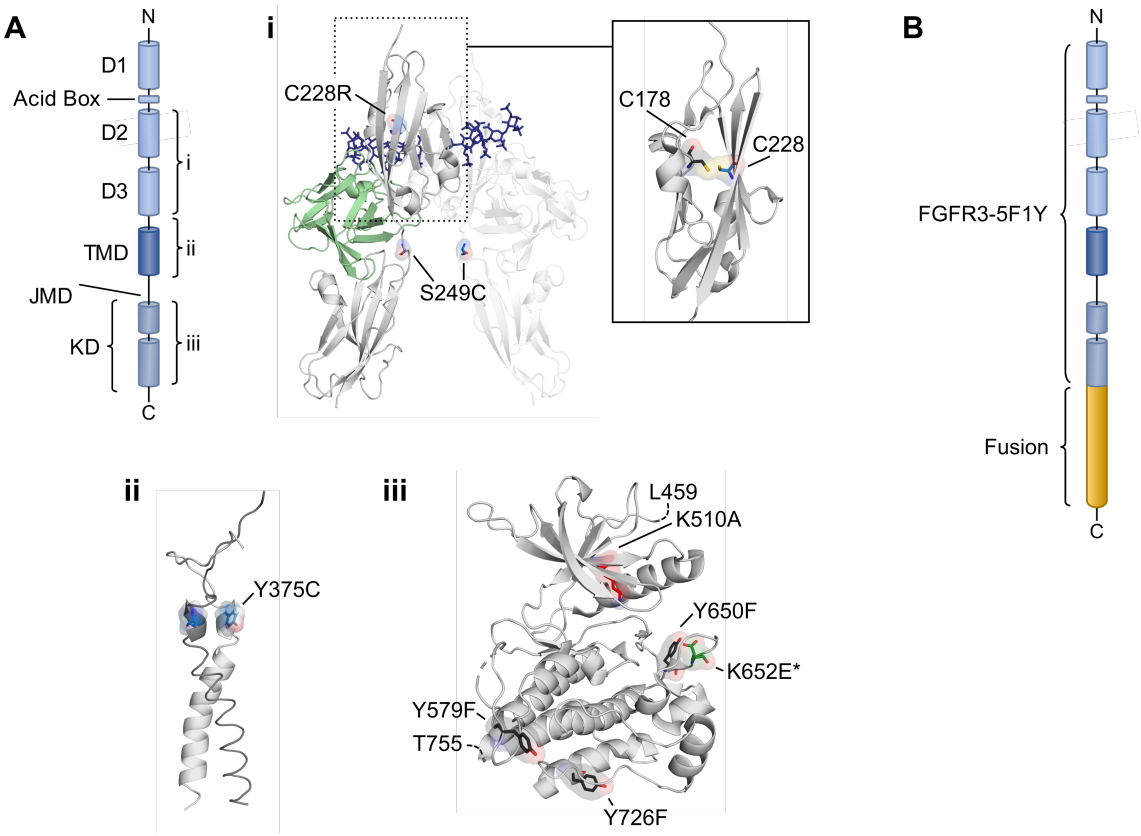


Figure 3.1. Constructs of full-length FGFR3 used in this study.
(A) Schematic illustration of full-length FGFR with subpanels (i-iii) illustrating the location of residue substitutions on the structures of individual FGFR domains. Residues are shown as sticks with transparent surfaces and domain secondary structures shown as grey cartoons (PDB entries 1FQ9, 2LZL and 4K33). Specifically, subpanel (i) illustrates the location of the cysteine substitutions C228R (also shown in an expanded view in context of its disulphide bond with C178) and S249C, (ii) illustrates the location of the cysteine substitution Y375C at the extracellular-transmembrane domain interface, and (iii) illustrates the location of the Tyr>Phe substitutions Y579F, Y650F, Y726F and Y755F, the constitutive kinase-activating mutation K652E, and the kinase-dead mutation K510A. Note that in this kinase domain structure (the only available for FGFR3), K652 has already been substituted to Glu (highlighted by an asterisk), and that the two further Tyr>Phe substitutions Y762F and Y772F introduced into the FGFR3-5F1Y variant lie just outside of the solved region of the kinase domain. (B) Schematic illustration of C-terminal fusion constructs of FGFR3-5F1Y.

3.2.2. Transfection and expression of FGFR3-5F1Y in Sf9 cells

FGFR3-5F1Y constructs were expressed in Sf9 insect cells using the Bac-to-Bac® protocol with minor modifications (as detailed in 2.4). This involved transformation of construct-encoding pFastBac plasmids into DH10Bac™ cells that carry the helper plasmid pMON7124 encoding Tn7 transposase which facilitates the transposition of the gene of interest into a recombinant baculovirus-encoding bacmid; these recombinant bacmids would be used to transfect Sf9 cells. Initially, a variant of DH10Bac™ cells named DH10EmBacY were used in transfections. These generate bacmids which encode baculoviruses that are protease and chitinase deficient and have prolonged cell viability following transfection compared standard DH10Bac cells, enhancing protein expression (329). The EmBacY baculovirus expression system has previously been identified as the best among its counterparts at expressing difficult intracellular proteins (329). The baculovirus also encodes the reporter molecule YFP, allowing transfection to be monitored by bright-field and fluorescence microscopy to assess transfection efficiency and protein production (320,329).

Transfection of FGFR3-5F1Y was successful as indicated by in-gel GFP fluorescence showing the expression of a protein at the expected molecular weight (~119 kDa) corresponding to FGFR3-5F1Y. Additional fluorescent proteins of approximately 58 kDa and 25 kDa were also detected (Figure 3.2A). Anti-His western blotting indicated that of these three fluorescent bands, only that at ~119 kDa carries a His-tag, suggesting that only this band corresponds to FGFR3-5F1Y. The species at 25 kDa corresponds to the YFP transfection reporter molecule or free GFP (Figure 3.2A).

In addition to allowing rapid determination of expression, the mGFP tag allows identification of the spatial cellular localisation of expressed proteins. This informs on the quality of membrane protein expression through interrogating whether the target protein is expressed to the plasma membrane, indicating that it has progressed through protein quality control procedures in the cell. While there is some indication that this may be the case for FGFR3-5F1Y, the high expression of the YFP reporter molecule in the cytoplasm of transfected Sf9 cells limited what could be concluded from this analysis (Figure 3.2B). Therefore, while the YFP reporter of DH10EmBacY-derived bacmids was useful to monitor transfection efficiency, as YFP and GFP have similar excitation and emission spectra (Figure 3.3), its presence was found to be detractive to the use of the mGFP tag as a high-throughput reporter on the magnitude and quality of construct expression.

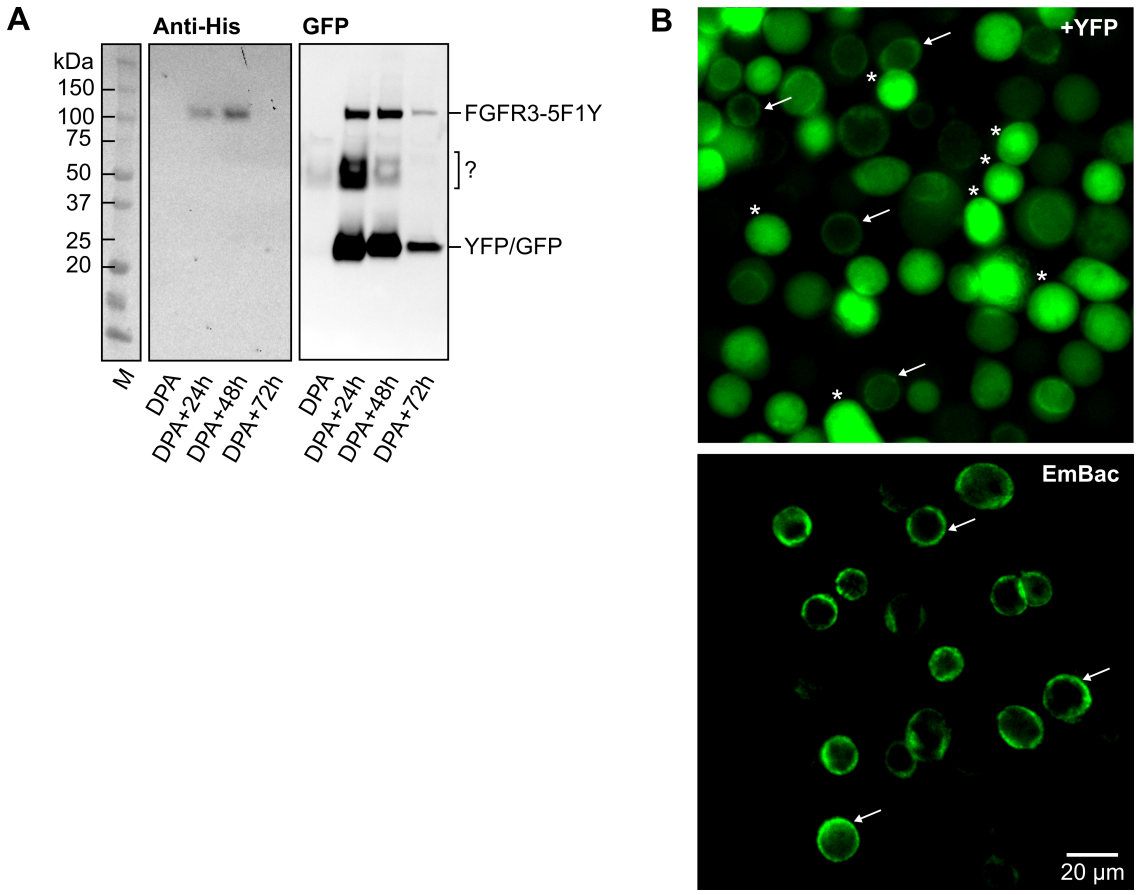


Figure 3.2. Expression of FGFR3-5F1Y in Sf9 cells.
(A) SDS-PAGE analysis of recombinant FGFR3-5F1Y expression in Sf9 cells indicating expression of full-length FGFR3-5F1Y. After infection of Sf9 cells with V0 virus, daily samples corresponding to 1 million cells (as determined by cell counting) were collected and stored at -20 °C. Following cell lysis as detailed in 2.4.4, clarified cell lysates were separated by SDS-PAGE, then fluorescent proteins imaged using in-gel GFP fluorescence (right), then proteins transferred to a nitrocellulose membrane for anti-His western blotting as detailed in 2.2.7 (left). DPA = day after proliferation arrest, M = molecular weight marker. (B) Example fluorescence microscopy images of Sf9 cells following transfection with the baculovirus encoding FGFR3-5F1Y derived from DH10EmBacY cells (top) and from DH10EmBac™ cells (bottom), each 48 hours after the 'day after proliferation arrest' (i.e. DPA+48h). Arrows indicate the plasma membrane-localised fluorescence signal likely corresponding to full-length FGFR3-5F1Y, while asterisks indicate cytoplasmic fluorescence signal likely deriving from the YFP reporter molecule of EmBacY bacmids. Fluorescence images were recorded using the GFP fluorescence mode (λ_{ex}: 470 nm, λ_{em}: 525 nm) on an EVOS FL Auto 2 Imaging System (Thermo Fisher Scientific).

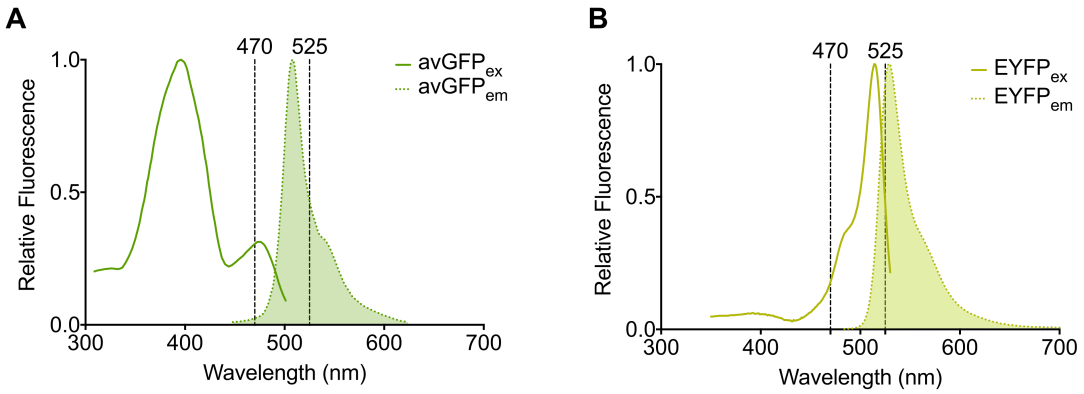


Figure 3.3. Excitation and emission profiles of GFP and YFP.
Panels show the excitation and emission profiles of avGFP (A, *Aequorea victoria* GFP) and EYFP (B, 'YFP'). In each, the excitation profile is shown with a solid line, and the emission profile as a dotted line with colour fill beneath the curve. Wavelengths used for EVOS fluorescence imaging (excitation of 470 nm, emission of 525 nm) are highlighted, illustrating the ability to detect both mGFP-tagged FGFR3-5F1Y and YFP using these wavelengths. mGFP used in the C-terminal tagging of FGFR3-5F1Y constructs corresponds to avGFP containing Q80R, F99S, M153T, V163A and A206K mutations (Appendix 2). These improve the stability and monomerise GFP but do not alter its absorbance properties. Curves were obtained from FPbase (330).

To resolve being unable to confidently conclude whether recombinantly-expressed FGFR3-5F1Y was localised at the cell membrane, transfections were repeated using bacmids generated from standard DH10EmBac™ cells. These lack the YFP reporter molecule, protease and chitinase gene deletions, but nevertheless also showed successful protein expression. Moreover, without interference from YFP, it was evident that FGFR3-5F1Y was predominantly located in the plasma membrane (Figure 3.2B), though fluorescent punctae were noted in some cases. This plasma membrane localisation indicated that full-length FGFR3-5F1Y had generally successfully passed protein folding quality control mechanisms in the cell. The identity of the punctae is currently unclear; these could be physiological clusters of FGFR3-5F1Y or those arising due to overexpression and consequent high local concentrations of the receptor. Alternatively, they could correspond to protein aggregates of FGFR3-5F1Y indicating incorrect processing of the receptor.

To quantify expression levels of FGFR3-5F1Y, viruses obtained from transfections (V0 viruses) were amplified through infection of small-scale Sf9 cultures, then used in titer testing to prepare BIICs for large-scale expression. At this stage, daily samples could be collected following the infection of cultures to determine the day of best recombinant protein expression (Figure 3.2A). Quantification of protein expression was completed at this stage rather than following transfections to overcome the potential underestimation of expression levels due to poor transfection efficiency and possible detrimental effects arising from the use of cytotoxic transfection reagents. Analysis of daily fractions

indicated that DPA+48h corresponded to the day of best expression for FGFR3-5F1Y (Figure 3.2A), with an estimated expression yield of 200-250 µg/L culture, determined by densitometry through comparing the intensity of the GFP fluorescence of the full-length FGFR3-5F1Y band and GFP standards following SDS-PAGE and in-gel GFP imaging. Quantification was completed on clarified total cell lysates to eliminate aggregated recombinant protein and provide an estimation of protein that is likely available to purification for structural characterisation.

3.2.3. The influence of kinase activity on expression levels

While total expression levels of 'wild-type' FGFR3-5F1Y were suitable for further investigations by EM, as losses are made during purifications, it is typically preferable to pursue characterisation of proteins with as high an expression as possible. To explore whether alternative constructs of FGFR3-5F1Y had improved expression levels with respect to 'wild-type' FGFR3-5F1Y (namely those with altered receptor kinase activity), V0 viruses (bacmids derived from DH10EmBacY cells) were amplified, and daily samples were collected for analysis. The constructs explored comprised FGFR3-5F1Y harbouring a K652E substitution (a constitutively active variant of the receptor), or FGFR3-5F1Y with cysteine substitutions in the extracellular domain, additionally with or without the K652E substitution. As noted before, these substitutions in the extracellular domain are thought to result in receptor activation through promotion of obligate receptor dimerisation in the absence of ligand stimulus.

Analysis of daily samples through SDS-PAGE of clarified total cell lysate and the subsequent densitometry analysis of fluorescence intensity of the full-length FGFR3 band indicated that, generally, the introduction of these residue substitutions did not greatly alter expression levels (Figure 3.4). With respect to FGFR3-5F1Y, the introduction of the activating mutation K652E increased expression levels 14%, while the cysteine substitutions Y375C and S249C (each with or without K652E) either had no appreciable effect on expression levels or decreased expression by up to 20% (Figure 3.4B). Each of these quantification measurements corresponds to a single experiment; given this lack of replicates, it is impossible to confidently conclude whether these differences in peak protein expression among the constructs are significantly different.

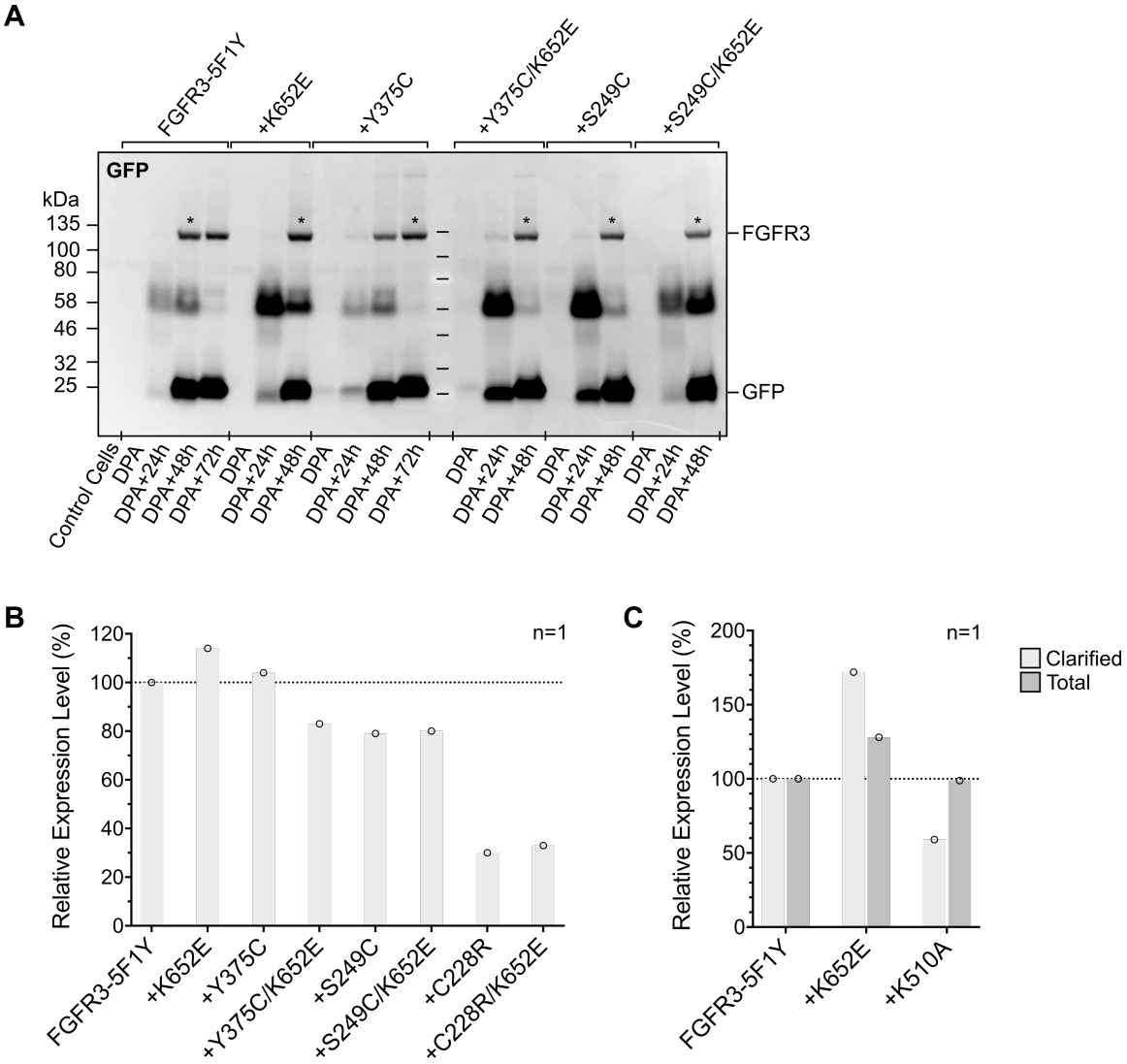


Figure 3.4. A comparison of the expression level of FGFR3-5F1Y and its variants in Sf9 cells.

(A) The relative expression level of different variants of full-length FGFR3 receptors with respect to FGFR3-5F1Y was assayed through densitometry analyses of the intensity of bands corresponding to full-length receptors following SDS-PAGE separation and in-gel GFP fluorescence imaging. To do so, following infection of Sf9 cells with V0 viruses, samples corresponding to 1 million cells (as determined by cell counting) were collected and stored at -20 °C. Following cell lysis as detailed in 2.4.4, clarified cell lysates were separated by SDS-PAGE, then fluorescent proteins imaged using in-gel GFP fluorescence. For each construct, the intensity of each full-length receptor band was determined by densitometry using ImageJ software. The relative expression level of each construct (as shown in (B)) was calculated from the day of highest expression for each construct with respect to FGFR3-5F1Y (as highlighted by asterisks). (B) An estimation of the relative expression levels of 'wild-type' and cysteine variants of FGFR3-5F1Y (following lysate clarification), each with and without the constitutive kinase activating substitution K652E, determined as from an example gel in (A). (C) An estimation of the relative expression levels of 'wild-type', 'constitutively active' (K652E) and 'kinase-dead' (K510A) variants of FGFR3-5F1Y, before and after lysate clarification. Analyses were conducted as for (A) and the gels used for these analyses provided in Supplementary Figure 1. Values are expressed as a relative measurement with respect to total and clarified amounts respectively of FGFR3-5F1Y and each represents a single measurement (n=1).

Among the substitutions assayed, C228R has the greatest effect on receptor expression levels, reducing expression 70% with respect to FGFR3-5F1Y (Figure 3.4B). The expression levels of this variant with and without the additional K652E substitution are comparable, suggesting that the C228R substitution rather than K652E carries the negative impact. Again, while the lack of replicates prevents confidently concluding that this substitution truly negatively affects expression levels, due to the magnitude of the difference observed in these single experiments, it is at least more likely that the change reflects a real reduction. The apparent difference in available amounts of receptor harbouring the C228R substitution with respect to other receptor variants may reflect the nature of the residue change. Unlike the other cysteine substitution variants which introduce an additional cysteine residue into the extracellular domain, the C228R mutation substitutes one of the two cysteine residues of a characteristic disulphide bond with C178 of the Ig-like extracellular domain D2 of the receptor (Figure 3.1A). This disulphide bond contributes to the stability of the Ig-fold (331), thus its destruction, while generating a free cysteine for potential receptor dimerisation through inter-domain crosslinking, likely has a negative effect on receptor stability as a whole.

The data presented thus far suggest that receptor kinase activity level does not generally alter expression levels in Sf9 cells. To explore this further, a kinase-dead variant of FGFR3-5F1Y harbouring a K510A substitution was expressed and analysed with respect to ‘wild-type’ and ‘constitutively active’ variants. In these experiments, baculoviruses prepared from DH10Bac™ cells were used, and both total cell lysate and clarified cell lysates were analysed by SDS-PAGE with anti-His western blotting. These suggested that total FGFR3-5F1Y expression was comparable among ‘wild-type’, ‘constitutively active’ and ‘kinase-dead’ variants (Figure 3.4C). However, differences in protein amount were observed following lysate clarification; the ‘constitutively active’ variant showed ~70% more protein with respect to ‘wild-type’ FGFR3-5F1Y, while the ‘kinase-dead’ variant showed a ~40% reduction (Figure 3.4C). Once more, the lack of replicates here limits what can be confidently concluded, but these findings indicate that the activity status of FGFR3-5F1Y may influence the amounts of protein available for characterisation, though the total recombinant protein expression levels may not be affected. It is noted that in both instances analysed, the ‘constitutively active’ form of FGFR3-5F1Y showed higher protein amounts in the clarified lysates compared to ‘wild-type’ receptor, though the magnitude differs (14% versus 73%). This difference could be the result of using baculoviruses generated from DH10Bac™ or DH10EmBacY cells respectively in each experiment; as previously stated, the latter was developed for improved recombinant protein expression and perhaps minimises differences between expression constructs. Alternatively, these differences may simply highlight the variability in expression trials, particularly in the absence of replicate data.

3.2.4. Investigating the dimerisation of cysteine variants

As discussed in 1.1.7.1, cysteine substitutions in the extracellular domain of FGFR3-5F1Y are thought to recapitulate receptor activation in the absence of activating ligands through constitutive dimerisation of the receptor, mediated through disulphide bonding. To confirm this notion and its utility for structural determination of receptor dimers, the clarified total cell lysates of Sf9 cultures expressing 'wild-type', S249C or Y375C variants of the receptor, each with and without the constitutive kinase activating mutation K652E, were analysed by SDS-PAGE in the presence and absence of the reducing agent dithiothreitol (DTT). While all receptor constructs displayed a monomeric receptor population in the presence of DTT (evidenced by a species at ~119 kDa), no dimeric populations were observed in the absence of DTT, neither for 'wild-type' nor the cysteine variants (Figure 3.5). Instead, in non-reducing conditions the full-length receptor did not enter the SDS-PAGE gel but remained in the gel wells, indicating that the receptor was of a considerably higher molecular mass and was most likely aggregated. Crucially, this phenomenon was observed with both 'wild-type' and 'constitutively active' variants of the receptor, indicating that it was not a consequence of introducing a free cysteine residue in the extracellular domain of the receptor. Instead, it pointed to a more general instability problem of FGFR3-5F1Y in the absence of reducing agent.

It is worth noting that the kinase domain of 'wild-type' FGFR3 contains two free cysteine residues; these residues are often substituted for recombinant expression of the isolated kinase domain (49), but were not substituted in FGFR3-5F1Y. While it is unclear from these data whether receptor aggregation occurs during cell lysis or during SDS-PAGE sample preparation (i.e. during protein unfolding by SDS which will also expose any buried free cysteine residues), the reducing agent TCEP was introduced into all future buffers to minimise undesirable protein crosslinking. However, in doing this, the probing of disulphide-bond meditated dimerisation was no longer possible.

A similar experiment to this one performed here has previously been performed on full-length FGFR3 in wild-type and C228R forms following expression in HEK293 cells (332). Curiously, in this instance, dimers of the C228R variant of the receptor were observed under non-reducing conditions and reducing agents were not used during cell lysis. While the wells of the SDS-PAGE gels were not visible in these published data, precluding confirmation that aggregation does not occur in this case also (though there is no suggestion that this does occur in the article text), the fact that dimers were observed in this instance following expression in a mammalian system but not following expression of FGFR3-5F1Y in Sf9 cells may instead indicate that the insect cell-expressed receptor is of poor quality and potentially incorrectly folded.

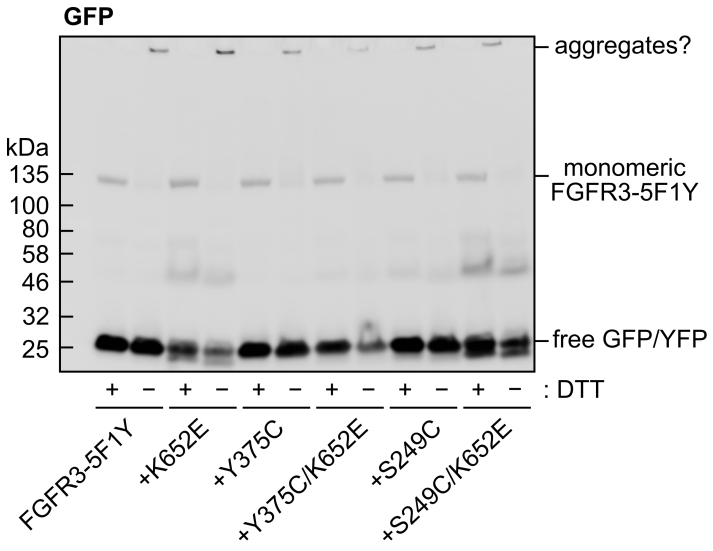


Figure 3.5. Analysing the intrinsic dimerisation of FGFR3-5F1Y cysteine variants under reducing and non-reducing conditions.
To investigate whether the cysteine variants of FGFR3-5F1Y were dimeric, samples corresponding to 1 million cells of the day of best expression for each receptor were lysed and clarified as in 2.4.4, then SDS-PAGE samples prepared with SDS-PAGE sample buffer either containing or lacking DTT. 'Wild-type' (i.e. FGFR3-5F1Y) and constitutively active forms (i.e. +K652E) of the receptor were also analysed as controls. Following protein separation by SDS-PAGE, the full-length mGFP-tagged FGFR3 receptors were visualised using in-gel GFP fluorescence. This analysis indicated that, in all cases, the receptor was aggregated in the absence of DTT. The location of full-length monomeric FGFR3-5F1Y receptor and probable aggregates are indicated.

3.2.5. Solubilisation trials

To purify FGFR3-5F1Y for structural characterisation, the receptor first requires solubilisation from the membrane. As a very wide range of detergents are commercially available and sampling of all of these would be laborious and expensive, literature reviews were first completed to select a small panel of detergents for testing. Two studies were particularly useful here, describing the characterisation of the receptor tyrosine kinases EGFR and PDGFRβ solubilised in a panel of different detergents (124,125). In the PDGFRβ study, using fluorescence size exclusion chromatography (FSEC), the authors found that the receptor was monomeric and monodisperse after solubilisation in the detergents LMNG, DDM and CHAPS (125), suggesting that these may be good candidates for FGFR3-5F1Y solubilisation. In the EGFR study, in addition to information about the solubility and monodispersity of the solubilised receptor from FSEC, information was also gathered regarding the kinase activity of solubilised receptors and their response to EGF ligand stimulus (124). While they found that EGFR was soluble in a broad range of detergents including members from the Triton, Tween, polyoxyethylene alkyl ether (CnEm), maltoside, glucoside, and Anzergent families, this did not always correlate with monodispersity and kinase activity. Notable detergents which solubilised monodisperse receptors with ligand-dependent kinase activity were Triton X-100,

Anzergent 3-8, DDM, and C12E8. The detergents LDAO and other members of the Anzergent family also isolated monodisperse receptors with kinase activity; however, in these instances, this activity could not be stimulated by EGF, indicating that these receptors were not fully functional. The findings from these two studies were used to inform on a starting panel of detergents for solubilisation trials of FGFR3-5F1Y. Unpublished data show that the receptor tyrosine kinase RET can be efficiently solubilised by LMNG and Cyclofos-6 each in the presence of cholesteryl hemisuccinate (CHS) (Dr Heidi Kaljunen, personal communications); these findings were incorporated into the detergent panel also.

Solubilisation trials of FGFR3-5F1Y were first completed using whole cells; the panel of detergents chosen for this study (DM, DMNG, DDM, LMNG, Cyclofos-6, C12E8, LDAO and Anzergent 3-8) were each trialled with and without the additive CHS to assess solubilisation efficiency and determine whether CHS provided any solubilisation benefits. Fos-choline-12, a harsh zwitterionic detergent, was used as a positive control to approximate the upper solubility limit of the receptor, found to be approximately ~60% of the total expressed receptor (Figure 3.6). A wide range of solubilisation efficiencies were observed across the detergent panel ranging from between 40-45% for Cyclofos-6 with and without CHS, LDAO alone and DDM with CHS, to 0-10% solubilisation by C12E8 and Anzergent 3-8, each with and without CHS (Figure 3.6). Detergents DM, DMNG, and LMNG, each with CHS, also solubilised the receptor reasonably well given the upper solubilisation limit; efficiencies of ~28-36% compared to ~60% for Fos-choline-12. Generally, these results indicated that the inclusion of CHS was beneficial to solubilisation efficiency, as has been found for mammalian proteins more generally previously (232–234).

In parallel, solubilisation of FGFR3-5F1Y by the co-polymers SMA and DIBMA was tested. These exhibited lower solubilisation efficiencies compared to most detergents (13% for SMA and 20% for DIBMA) (Figure 3.6). However, it is important to note that neither SMA nor DIBMA remove the receptor from the membrane environment. Therefore, their solubilised fractions may contain higher quality and more functional protein, or at least protein in a state that better mimics how it is found in cells. The retention of some of the membrane may, for example, stabilise the juxtamembrane domain of the receptor by allowing electrostatic interactions between its cluster of positively charged residues and any retained anionic lipids. These interactions may be important for receptor function and may not be possible following solubilisation in detergents. Thus, despite their lower solubilisation efficiencies, both SMA and DIBMA were considered good candidates to take forward.

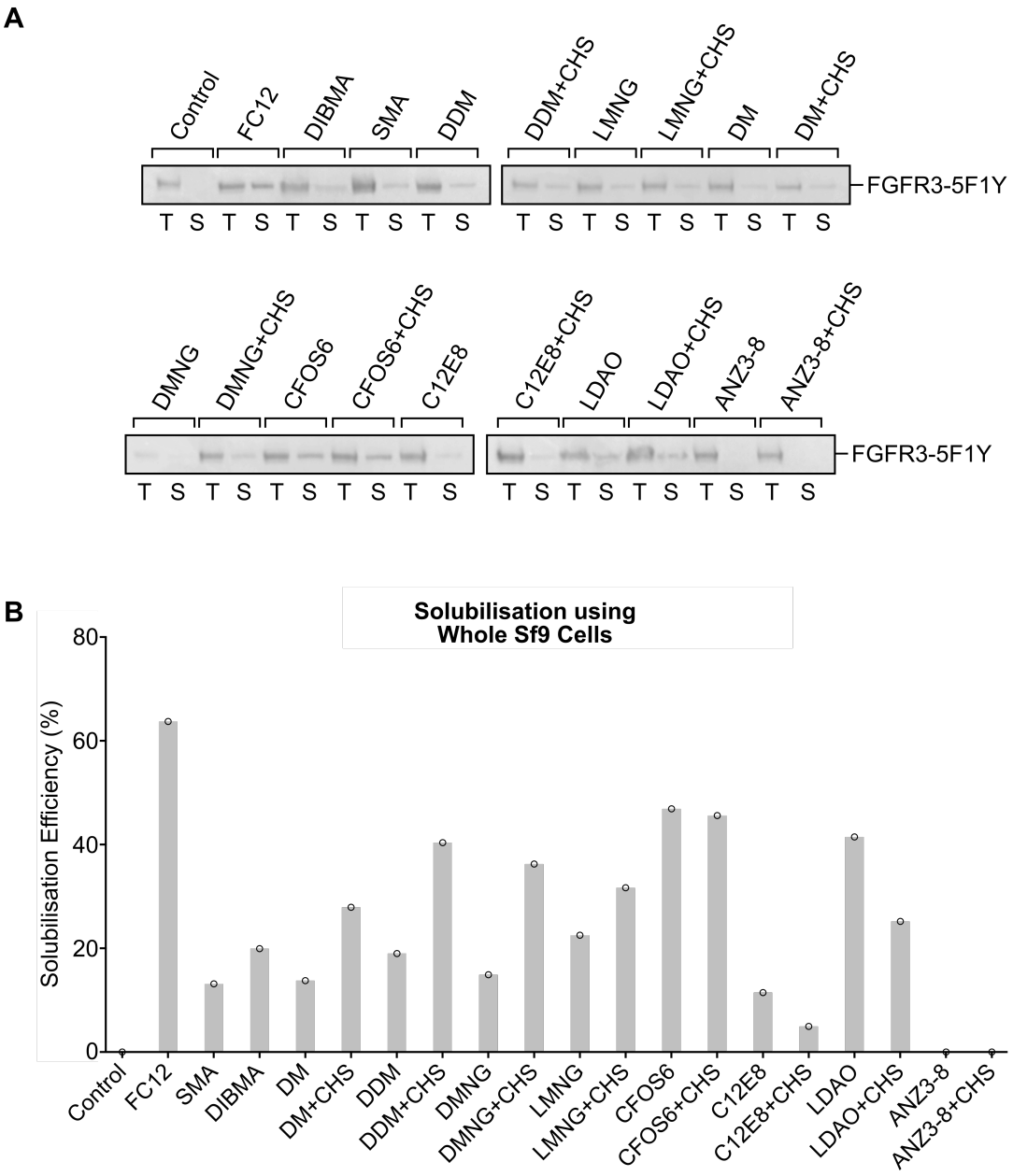


Figure 3.6: Trial solubilisation of FGFR3-5F1Y using whole cells.
(A) The solubilisation efficiency of FGFR3-5F1Y using a detergent panel, SMA and DIBMA, was first assessed using whole Sf9 cells. As detailed in 2.6.1.2, Sf9 cells that had recombinantly expressed FGFR3-5F1Y were diluted to an OD_{600} of 2.0 in FGFR3 homogenisation buffer in the presence of the solubilising agent in question, then incubated overnight to allow solubilisation to proceed. Detergents were used at a 1% *w/v* final concentration and SMA/DIBMA at a 2% *w/v* final concentration. Where used, CHS was present at a 0.2% *w/v* concentration. Following overnight incubation, samples were spun to pellet the insoluble fraction as in 2.6.1.2. Solubilisation efficiency was assessed through densitometry analysis of the intensity of the full-length receptor band in total (T) and soluble (S) fractions of each sample following SDS-PAGE separation and the detection of in-gel GFP fluorescence. The total and soluble fractions correspond to the sample before and after centrifugation to remove the insoluble fraction. FC12 and untransfected Sf9 cells alone were used as positive and negative controls respectively. (B) The solubilisation efficiency of each agent, as estimated from densitometry analyses in (A). The efficiency is expressed as the percentage of receptor present in the soluble fraction with respect to the total fraction. Each estimation corresponds to a single measurement (n=1). Abbreviations: FC12 (Fos-choline-12), CFOS6 (Cyclofos-6), ANZ3-8 (Anzergent 3-8).

As purification of FGFR3-5F1Y would first start with the isolation of the membrane fraction following cell lysis, solubilisation trials were next repeated using FGFR3-5F1Y-containing membranes following membrane preparation. Here, the detergent panel was further reduced through exclusion of C12E8, LDAO and Anzergent 3-8. These detergents either showed low solubilisation efficiencies in whole cell trials (Figure 3.6), or, as was the case for LDAO, was excluded as it is a relatively harsh detergent. In this instance, all detergents were tested in the presence of CHS given that this improved solubilisation efficiency in whole cell trials. The co-polymers SMA and DIBMA were also tested. As shortening the length of purification protocols would help maximise the quality of the final purified protein, solubilisation was allowed to proceed for both 2 hours and overnight to investigate whether shorter incubation times were possible. This indicated that receptor solubilisation for 2 hours was comparable to solubilisation allowed to proceed overnight (Figure 3.7).

Following membrane preparation, the upper solubilisation limit of the receptor indicated by Fos-choline-12 was found to be slightly higher than in whole cell trials (~70% versus ~60% of the total protein) (Figure 3.7). This was a trend seen across a number of detergents (DM, DMNG and Cyclofos-6), and may reflect the removal of aggregated, insoluble receptor during the membrane preparation stage. The solubilisation efficiency by DDM and LMNG remained relatively unchanged, although there was considerable variability between replicate experiments (Figure 3.7). Reproducibility was a common issue in these assays with the exception of when using SMA and DIBMA at the two-hour timepoint (note that there is only one measurement for Cyclofos-6 due to insufficient amounts of this detergent for replicates). This may be partly attributable to the method of quantification (as anti-His western blotting provided better signal to noise for quantification by densitometry, these data was used for quantification in all cases bar SMA and DIBMA as the receptor did not blot well in the presence of these), or may possibly reflect minor inconsistencies in the concentration of CHS in the detergent solutions given that CHS does not dissolve readily but was found to have a considerable impact on solubilisation efficiency in whole cell trials (Figure 3.6). The solubilisation efficiency of SMA and DIBMA was also found to be improved compared to that observed in whole cell trials at 47% and 30%, respectively (Figure 3.7).

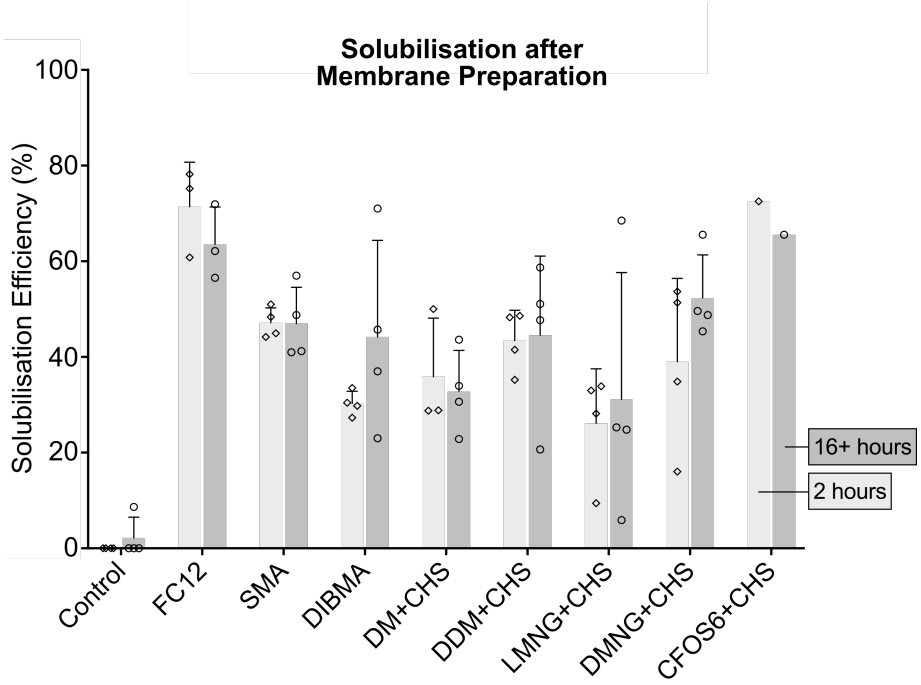


Figure 3.7. Solubilisation trials of FGFR3-5F1Y following membrane preparation. After membrane preparation (see Figure 3.9 and 3.2.6.1), the solubilisation efficiency of each detergent, SMA and DIBMA was assessed once more. The efficiency was assessed as detailed in 2.6.1.2 and as described briefly in Figure 3.6A except that membranes were used at a final concentration of ~0.15 mg/mL and solubilisation allowed to proceed for either 2 or 16+ hours. Solubilisation efficiencies were estimated by densitometry after SDS-PAGE as in Figure 3.6A except that anti-His western blotting rather than in-gel fluorescence was used for all samples except SMA and DIBMA. This discrepancy arises from the observation that SMA and DIBMA-solubilised receptors were not detected by anti-His western blotting but were observed with in-gel GFP fluorescence. As western blotting provided better signal to noise for densitometry analyses, these data were used for all other measurements. The estimated solubilisation efficiency for each agent is expressed as the percentage of receptor present in the soluble fraction with respect to the total fraction. In the graph, the bars correspond to the mean value and error bars to the standard deviation of this data. Individual data points are also provided (circles and diamonds). FC12 and membranes alone were used as positive and negative controls respectively. Replicate values are: 1 (CFOS6+CHS); 3 (Fos-choline-12); and 4 (all others). Trials using CFOS6+CHS could not be repeated due to a lack of available detergent. Abbreviations used: FC12 (Fos-choline-12), CFOS6 (Cyclofos-6), ANZ3-8 (Anzergent 3-8).

Following analyses by SDS-PAGE, to assess the quality of receptor solubilisation, the solubilised receptor in the presence of each agent was also assessed by Native-PAGE. In all cases, this suggested that no receptor aggregation was observed with any of the solubilisation agents (Figure 3.8); however, the results are otherwise difficult to interpret, particularly as in-gel GFP fluorescence imaging and anti-His western blotting of the same samples gave conflicting results. From both imaging methods, it is clear that receptor solubilised with each agent migrates differently in Native-PAGE and, perhaps with the exception of DIBMA (which appears smeared on the gel), each solubilised receptor also migrates as more than one species. As Native-PAGE separates by size, shape and charge of molecules, these differences in the migration pattern with each detergent, SMA and DIBMA likely reflect differences in the size and charge of the micelle that the receptor is solubilised in, or differences in co-polymer chemistry and nanoparticle disc size. In the case of SMA- and DIBMA-solubilised receptors, migration differences may be further influenced by any heterogeneity in the lipids that have been solubilised with the receptor. It is currently unclear what each of the multiple species detected for each detergent may correspond to; though they could reflect different dimerisation states of the receptor, it is not possible to conclude this from Native-PAGE alone but would require analyses such as fluorescence size exclusion chromatography. Curiously, anti-His western blotting doid not consistently detect two species for each detergent, despite one of these species (marked with an asterisk, Figure 3.8) appearing to be shared among detergents in GFP imaging. This second species is detected in DDM+CHS, and to some extent in LMNG+CHS solubilised samples, but not with DMNG+CHS. While western blotting suggested that DMNG+CHS produced the most monodisperse sample, as no detergent led to receptor aggregation, all those analysed were likely good candidates to pursue further.

Combining both SDS-PAGE and Native-PAGE analyses, all four of the tested detergents (DM, DDM, LMNG and DMNG) solubilised FGFR3-5F1Y with similar average efficiencies of between 30-40% in the presence of the cholesterol substitute CHS, and yielded solubilised receptors that were not aggregated and thus likely suitable for further characterisation. Solubilisation with DIBMA occurs at a comparable efficiency (~30%), and though it too does not seem to yield aggregated soluble receptors, its smearing on Native-PAGE may suggest that DIBMA lipid-protein nanoparticles are quite heterogeneous. While solubilisation efficiency by SMA is slightly higher, Native-PAGE suggests that SMA lipid-protein nanoparticles may also be rather heterogeneous, possibly even more so.

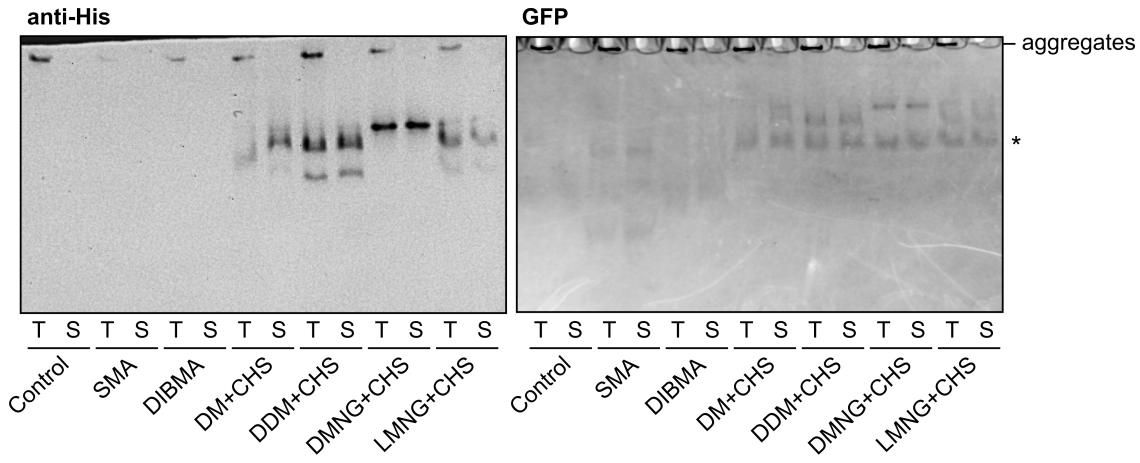


Figure 3.8. Analysis of FGFR3-5F1Y solubilisation by Native-PAGE.
To assess the monodispersity and quality of FGFR3-5F1Y solubilised by each detergent, SMA and DIBMA, solubilisation trials were repeated as before (Figure 3.7), then analysed by Native-PAGE. After protein separation, FGFR3-5F1Y was visualised by in-gel GFP fluorescence (right) then proteins transferred to a nitrocellulose membrane for anti-His western blotting (as in 2.2.7). To assess the effect of protein solubilisation on receptor quality, both total (T) and soluble (S) fractions were assessed. The total and soluble fractions correspond to the sample before and after centrifugation to remove the insoluble fraction. These analyses indicated that receptor solubilisation in no detergent, SMA nor DIBMA appeared to lead to receptor aggregation. However, solubilised FGFR3-5F1Y does not appear to be monodisperse in any case, except perhaps following solubilisation in DMNG+CHS. The asterisk (*) highlights a common species shared by all detergents using in-gel fluorescence analysis that is not observed by western blotting. Note that these data were not used to assess solubilisation efficiency but the quality of the solubilised sample only.

3.2.6. Purification of FGFR3-5F1Y

3.2.6.1. Preparation of membranes

Following large-scale expression of FGFR3-5F1Y using BIICs, Sf9 cells were lysed and the membrane fraction containing FGFR3-5F1Y isolated. These prepared membranes would be the starting material of all protein purifications. The C-terminal mGFP tag of FGFR3-5F1Y conveniently allowed the tracking of receptor retention through this procedure through densitometry of the band corresponding to full-length receptor following SDS-PAGE separation of proteins and imaging of in-gel GFP fluorescence (Figure 3.9). This indicated that among different membrane preparations, at most only ~30% of the total expressed receptor was retained after cell lysis by dounce homogenisation on ice. A second round of dounce homogenisation did not recover much further target protein (Figure 3.9). This suggested that the majority of the recombinantly expressed receptor was aggregated, either in the cells, or becoming so during the process of cell lysis. This loss may indicate that Sf9 cells are unable to correctly process and fold the receptor, potentially as a consequence of recombinant overexpression. Alternatively, protein loss could be consistent with a general instability of the receptor.

Of the protein recovered from cell lysis, an estimated 90% of this was retained to the final membrane fraction following washing (Figure 3.9). In addition to the full-length receptor, two further fluorescent species at ~80 kDa and 25 kDa were present in the membrane fraction. These likely correspond to a truncated form of the receptor (80 kDa) and free mGFP or YFP (25 kDa). Notably, the species at ~80 kDa is also detected by anti-His western blotting while the species at 25 kDa is not. This would suggest that the receptor is truncated at the extracellular domain (as the C-terminal mGFP-His8 tag is retained) and that some free YFP/GFP is brought through the membrane preparation despite these proteins being soluble. It is noted that the additional ~80 kDa species that is retained and concentrated during membrane preparation does not correspond to those observed previously at ~58 kDa during transfections (Figure 3.2) and are present in the total Sf9 cell lysate but are lost during membrane preparation (Figure 3.9).

Using in-gel GFP densitometry analysis of the final concentrated membrane sample with comparison to GFP standards, it is estimated that FGFR3-5F1Y can be recovered at ~200 µg/L culture following membrane preparation. This yield estimation is in line with that estimated from small-scale cultures previously (see 3.2.2).

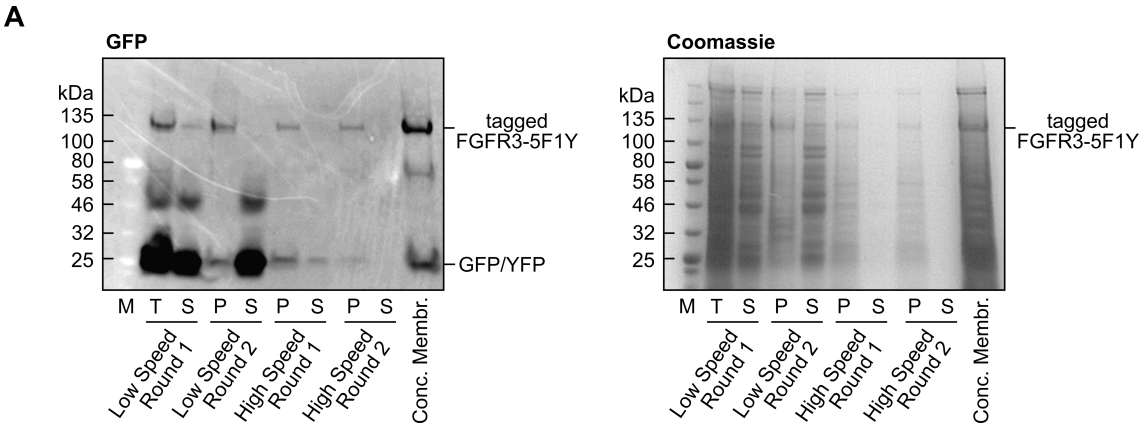


Figure 3.9. Preparing membranes containing FGFR3-5F1Y from Sf9 cells.
As a primary purification step, membranes were isolated from Sf9 total cell lysate (T). As described in detail in 2.6.1.1, following cell lysis by dounce homogenisation, cell debris and intact cells were pelleted by a low speed centrifugation (3,000 g), then this pellet (P) lysed once more and subjected to a second low speed centrifugation. The supernatant (S) of each of these low speed spins was combined, then membranes pelleted by spinning at 42,000 rpm for 1 hour in a Type 45 Ti rotor (High Speed Round 1). As detailed in 2.6.1.1, these membranes were washed in FGFR3 homogenisation buffer twice, then the pellet fraction of the centrifugation 'High Speed Round 2' resuspended in a small volume of FGFR3 homogenisation buffer to obtain the final concentrated membranes (Conc. Membr.). Samples from each of these stages was retained for analysis by SDS-PAGE with in-gel GFP fluorescence imaging followed by Coomassie staining. Densitometry of the band corresponding to tagged full-length FGFR3-5F1Y (labelled) in the GFP image was used to estimate recovery of the receptor during membrane preparation. M = molecular weight marker.

3.2.6.2. Purification of detergent-solubilised receptors

Purification of FGFR3-5F1Y was first tested following solubilisation with LMNG+CHS (Figure 3.10A). While not showing the highest solubilisation efficiency of the receptor among the detergent panel, LMNG was chosen for these first purification trials as it has been reported that desorption of the detergent following protein solubilisation is at such a slow rate that the detergent is no longer required above its CMC value (333). This property could prove to be useful for electron microscopy to minimise the number of empty micelles present in the sample which can be hard to distinguish from the solubilised protein in raw EM micrographs (334).

In practice, in these first purification trials, LMNG was used at a minimum concentration of 10 x CMC (0.01% LMNG and 0.002% CHS). A solubilisation efficiency of ~40% was achieved following overnight incubation, consistent with that observed in solubilisation trials (Figure 3.7). After solubilisation, the soluble fraction was incubated with nickel resin for 2 hours at 4 °C to immobilise the His-tagged FGFR3-5F1Y, then the resin washed with buffers containing increasing concentrations of imidazole (up to 20 mM) to remove unbound and contaminant proteins. Through comparison of the amount of receptor present in soluble fraction following solubilisation to the amount present in the flow through fraction after incubation with the nickel resin, densitometry analysis indicated that only ~30-40% of the total solubilised receptor was captured by the resin (Figure 3.10A). After washing of the resin (during which no further protein was lost), on-column cleavage of the C-terminal mGFP-His8 tag was completed through incubation of the resin with HRV 3C protease overnight. Theoretically, if tag cleavage were successful, untagged FGFR3-5F1Y would be expected to elute from the column when washed with buffer lacking imidazole; however, this was not observed. Instead, while tag cleavage had completed (indicated by the disappearance of GFP fluorescence at ~120 kDa by SDS-PAGE), untagged FGFR3-5F1Y did not elute from the column until incubation with elution buffer containing 250 mM imidazole, indicating that the receptor binds non-specifically to the resin (Figure 3.10A). To confirm that the major protein present in the elution fraction with imidazole was indeed untagged FGFR3-5F1Y, a sample was sent for protein identification by mass spectrometry following in-gel digestion (performed by Dr Rachel George, FBS Mass Spectrometry Facility) (Figure 3.10B). As untagged FGFR3-5F1Y did not elute where expected (i.e. without imidazole), the final sample contained impurities such as the cleaved mGFP-His8 tag and other contaminant proteins which were not lost during 20 mM imidazole washes. The purity obtained was insufficient for structural studies.

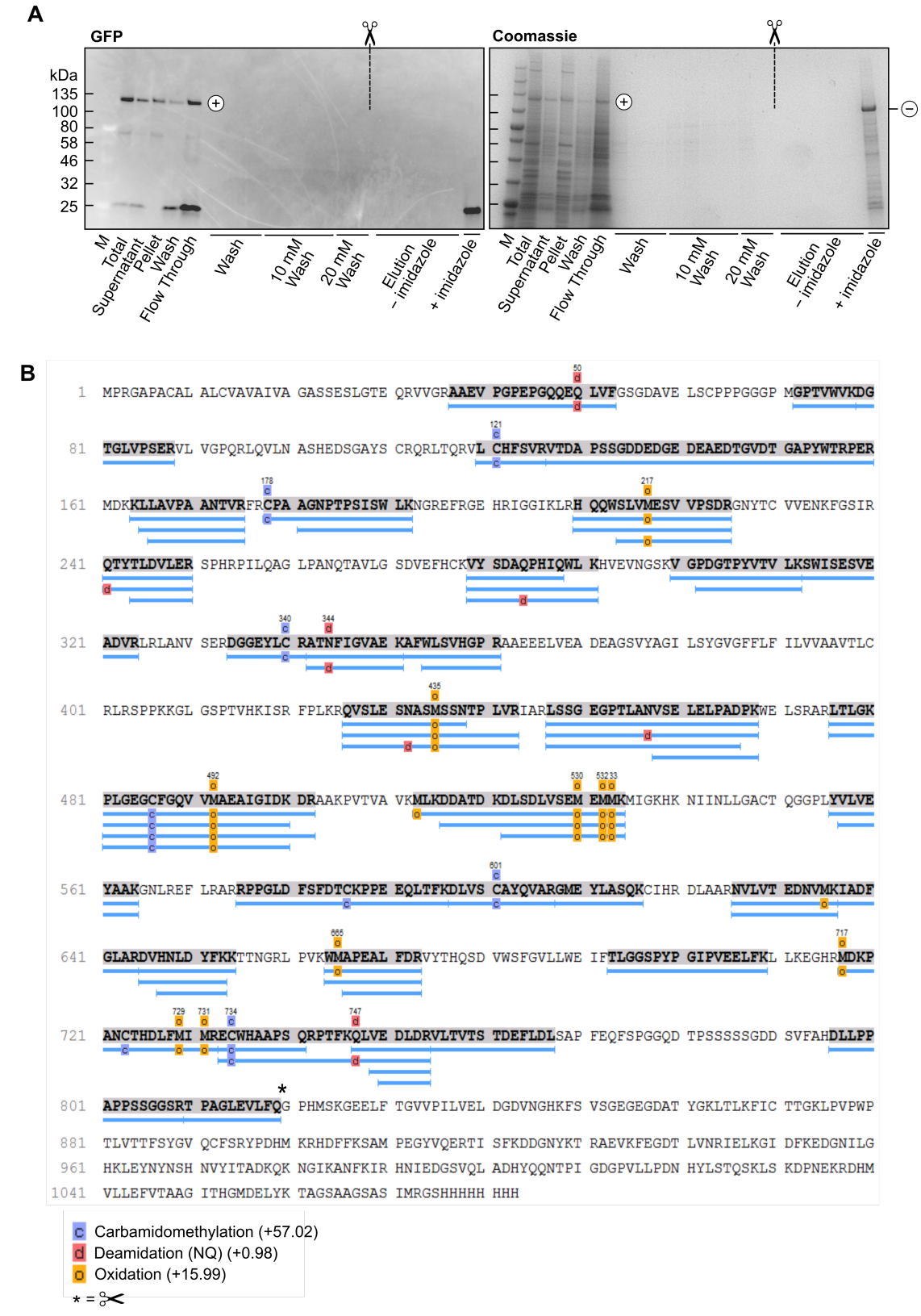


Figure 3.10. Purification of FGFR3-5F1Y in LMNG+CHS.
This caption to this figure follows on the next page.

(A) Purification of FGFR3-5F1Y was first attempted following receptor solubilisation in LMNG+CHS as detailed in 2.6.1.3. Briefly, after solubilisation overnight at 4 °C, the solubilised receptor was incubated with Nickel Sepharose Fast Flow 6 resin for 2 hours, then unbound proteins collected in the 'Flow Through' and 'Wash'. After washing the resin with buffers containing increasing concentrations (10 and 20 mM) of imidazole to remove contaminant proteins, on-column tag cleavage was completed through addition of His6-tagged HRV 3C protease with overnight incubation. The following day, unbound proteins (which was expected to contain untagged receptor) were collected ('Elution – imidazole'), then bound proteins eluted with imidazole ('+imidazole'). Samples from each stage were analysed by SDS-PAGE with in-gel GFP fluorescence imaging followed by Coomassie staining. Tagged and untagged FGFR3-5F1Y are marked with a '+' and '–' respectively, M = molecular weight marker. The scissors indicate tag cleavage through incubation with HRV 3C protease. (B) To confirm the identity of the major protein present in the imidazole elution fraction of (A), a Coomassie-stained gel sample was sent for in-gel digestion and protein identification by mass spectrometry. Blue bars show peptide coverage obtained over the sequence of mGFP-His8 tagged FGFR3-5F1Y, confirming the protein to be the untagged receptor. The asterisk indicates the cut site of HRV 3C protease.

To explore the non-specific binding of untagged FGFR3-5F1Y to the nickel resin further, another purification was completed, again using LMNG+CHS, but allowing solubilisation to proceed for 2 hours only. As a further adaptation, the detergent concentration was diluted to ~0.8% LMNG and ~0.017% CHS during the resin binding stage (compared to 2% LMNG and 0.4% CHS used previously). Additionally, after receptor immobilisation onto and the washing of nickel resin, untagged FGFR3-5F1Y was obtained by one of two methods; either through on-column cleavage as previously, or through first eluting tagged FGFR3-5F1Y from the resin and completing tag cleavage in solution during overnight dialysis. Following in-solution tag cleavage, the dialysed sample was re-applied twice to nickel resin, any unbound proteins washed away and collected, then bound proteins eluted with 250 mM imidazole as before. On this occasion, solubilisation efficiency was estimated to be ~70%, higher than typically observed, and approximately 65-70% of the solubilised receptor bound to the nickel resin (Figure 3.11). This was a considerable improvement from before, suggesting that detergent presence has a detrimental effect on protein immobilisation. Like previously, when using on-column cleavage, untagged FGFR3-5F1Y bound non-specifically to the nickel resin. Non-specific binding was also observed following tag cleavage in solution with >50% loss of receptor onto the resin; however, some untagged receptor was recovered in the flow through fraction of this reverse-IMAC step, resulting in a higher purity sample than from on-column tag cleavage (Figure 3.11). The recovered, partially purified samples were used to assay the kinase activity of purified FGFR3-5F1Y (see 3.2.8).

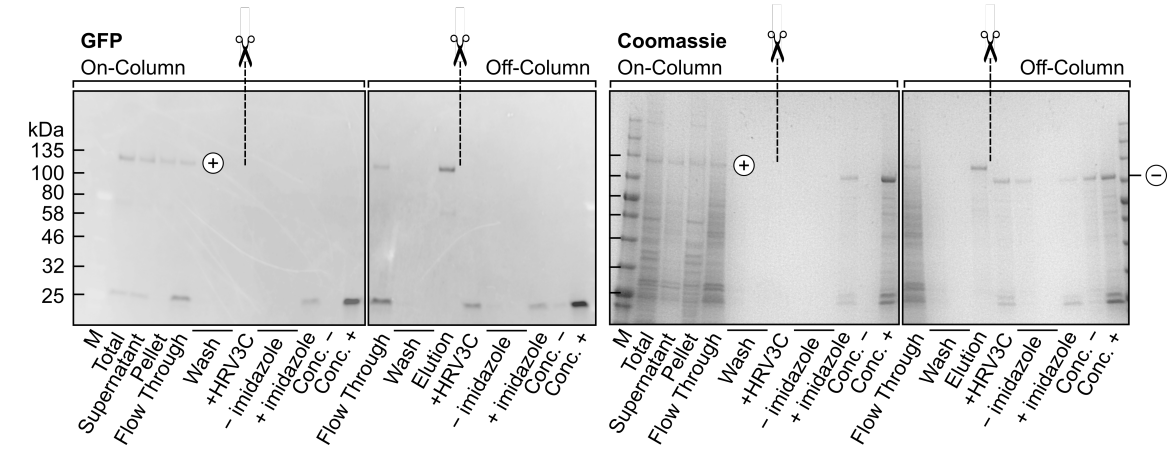


Figure 3.11. Probing non-specific binding of FGFR3-5F1Y to nickel resin.
To probe non-specific binding of untagged FGFR3-5F1Y to Nickel Sepharose Fast Flow 6 resin, purification of the receptor was repeated in LMNG+CHS except that, on this occasion, tag cleavage through addition of His6-tagged HRV 3C protease was either completed 'On-Column' through incubation with the receptor while immobilised to the nickel resin (as in Figure 3.10) or 'Off-Column' through incubation during dialysis after elution of the tagged receptor from the nickel resin. Both samples were then reapplied to nickel resin and unbound proteins recovered ('–imidazole') and bound proteins eluted ('+imidazole'). Proteins from both fractions were concentrated to allow for purity estimation ('Conc. –' and 'Conc. +', respectively). Samples from each stage were analysed by SDS-PAGE with in-gel GFP fluorescence imaging followed by Coomassie staining. Tagged and untagged FGFR3-5F1Y are marked with a '+' and '–' respectively, M = molecular weight marker. The scissors indicate tag cleavage through incubation with HRV 3C protease.

Independently, purification of FGFR3-5F1Y following solubilisation in DDM+CHS was also tried. In this case, purification generally proceeded as with LMNG except membranes containing FGFR3-5F1Y were solubilised in 1% *w/v* DDM and 0.2% *w/v* CHS for 1.5 hours at 4 °C, and binding to nickel resin occurred in the presence of 10 mM imidazole to minimise the binding of contaminant proteins. The detergent concentration was not diluted during receptor binding to resin. Densitometry analyses as before indicated that a solubilisation efficiency of ~50% was achieved with DDM+CHS, consistent with that observed in solubilisation trials (Figure 3.12A). Approximately 85% of the solubilised receptor fraction bound to nickel resin, an improvement on that observed with LMNG. Following washing of the resin with 20 mM imidazole to remove contaminating proteins, tagged FGFR3-5F1Y was eluted from the resin at a purity of approximately 50% as estimated from densitometry. This sample was used to prepare negative-stain EM grids (see 3.2.7), for kinase activity assays (see 3.2.8) and further purified through in-solution tag cleavage and reverse-IMAC to probe non-specific binding of the untagged receptor in a different detergent. As with LMNG, non-specific binding was similarly observed, indicating that it is likely a property of the protein rather than the detergent that the receptor is solubilised in (Figure 3.12B). Given the non-specific binding, little purity enhancement was gained through tag cleavage (Figure 3.12B).

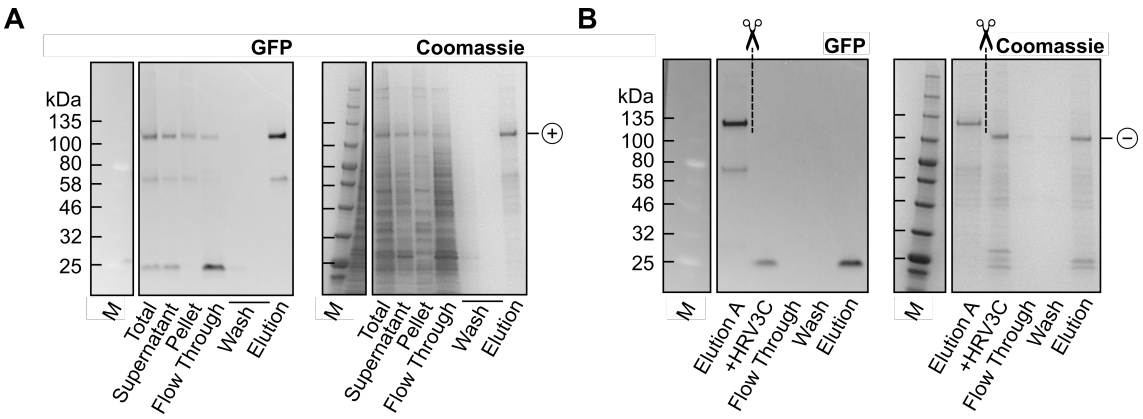


Figure 3.12. Purification of FGFR3-5F1Y in DDM+CHS.
To assess whether non-specific binding of untagged FGFR3-5F1Y could be related to its solubilisation in LMNG+CHS, the receptor was next purified following its solubilisation in DDM+CHS (see 2.6.1.3). On this occasion, following solubilisation and immobilisation of the receptor on to nickel resin, the resin was washed with FGFR3 His Wash Buffer containing 20 mM imidazole, then bound proteins eluted (Panel A). Tag cleavage was completed through addition of His6-tagged HRV 3C protease ('+HRV3C') to the eluted sample ('Elution A') then proteins reapplied to nickel resin, the unbound proteins collected ('Flow Through' and 'Wash') and bound proteins eluted with FGFR3 His Elution Buffer containing 250 mM imidazole ('Elution') (Panel B). Samples were taken at each stage and analysed by SDS-PAGE with in-gel GFP fluorescence imaging followed by Coomassie staining. M = molecular weight marker, and '+' and '–' indicates tagged and untagged FGFR3-5F1Y respectively.

3.2.6.3. An alternative affinity chromatography method

As mGFP-His8 tag removal and reverse-IMAC was generally unproductive during receptor purification due to non-specific binding of the receptor, an alternative affinity chromatography-based method was sought to improve sample purity. GFP-Trap® resin (Chromotek) provided this alternative without requiring the re-cloning of constructs to introduce a different affinity tag. This resin utilises an immobilised anti-GFP nanobody for immunoprecipitation of GFP-tagged proteins, which if compatible with on-column tag cleavage without non-specific binding, should yield highly-pure untagged FGFR3-5F1Y receptor. A small-scale purification trial was completed to test this.

Due to concerns regarding receptor activity following purification in detergents (see 3.2.8), this purification was completed using the co-polymer DIBMA, thus retaining the receptor in a lipid-membrane environment should this be critical for receptor activity. Purification of FGFR3-5F1Y in DIBMA was also expected to be beneficial for electron microscopy through avoiding the presence of empty detergent micelles in electron micrographs, providing a better sample to investigate protein quality and structure.

On this occasion, to probe the ability of FGFR3-5F1Y to form ligand-bound complexes as a proxy for receptor activity, the receptor was additionally purified in the presence of its ligands. To form complexes, membranes were incubated with recombinant human FGF1 (Abcam) and the heparin oligosaccharide dp8 (Iduron) for 1 hour at 4 °C prior to solubilisation with DIBMA (see 2.6.1.3 for details). An unliganded sample was purified in parallel for comparison. After incubation of the DIBMA-solubilised receptor with the GFP-Trap® resin for 2 hours, while binding of the GFP-tagged receptor to the resin was evident, the supernatant was still fluorescent suggesting that the resin had become saturated. Any unbound proteins were thus recovered and incubated with nickel resin overnight for further purification by IMAC (Figure 3.13). Following the washing of unbound proteins from the GFP-Trap® resin, untagged FGFR3-5F1Y was recovered by incubation with HRV 3C protease overnight. Unlike that observed with nickel resin, untagged receptor eluted from the resin generally as expected; however, the obtained sample was not as pure as anticipated, indicating that further resin washing would be necessary prior to tag cleavage and protein elution (Figure 3.13). Nonetheless, the purification trial suggested that GFP-Trap® resin could be an alternative for receptor purification. Further purification of the unbound material by IMAC proceeded as expected from prior purification trials (Figure 3.13).

Interestingly, following solubilisation of the receptor in DIBMA, the presumed truncated receptor species at ~80 kDa as seen in other purification trials did not co-purify with the full-length receptor (Figure 3.13), suggesting that DIBMA solubilisation provides purity enhancements not observed with the detergents LMNG and DDM (Figure 3.10, Figure 3.11 and Figure 3.12). Through comparison of Coomassie-stained SDS-PAGE gels of receptor purification in the presence and absence of its ligands, there is some indication that FGF1 may co-purify with FGFR3-5F1Y, suggesting successful ligand binding by the receptor and complex formation. This may imply that the receptor is in a functionally-relevant, correctly folded state, and is mostly clearly observed in the IMAC purification but is also evident in the GFP-Trap® elution fraction. However, given that FGF1 is also present in the wash fractions of the IMAC purification immediately prior to receptor elution, it will be necessary in the future to demonstrate that co-elution of FGF1 and the receptor is not a consequence of non-specific binding of FGF1 to the nickel resin.

Given their higher purity, the IMAC elution fractions were used in kinase activity assays and the preparation of grids for negative-stain EM to assess the integrity of FGFR3-5F1Y while solubilised in DIBMA.

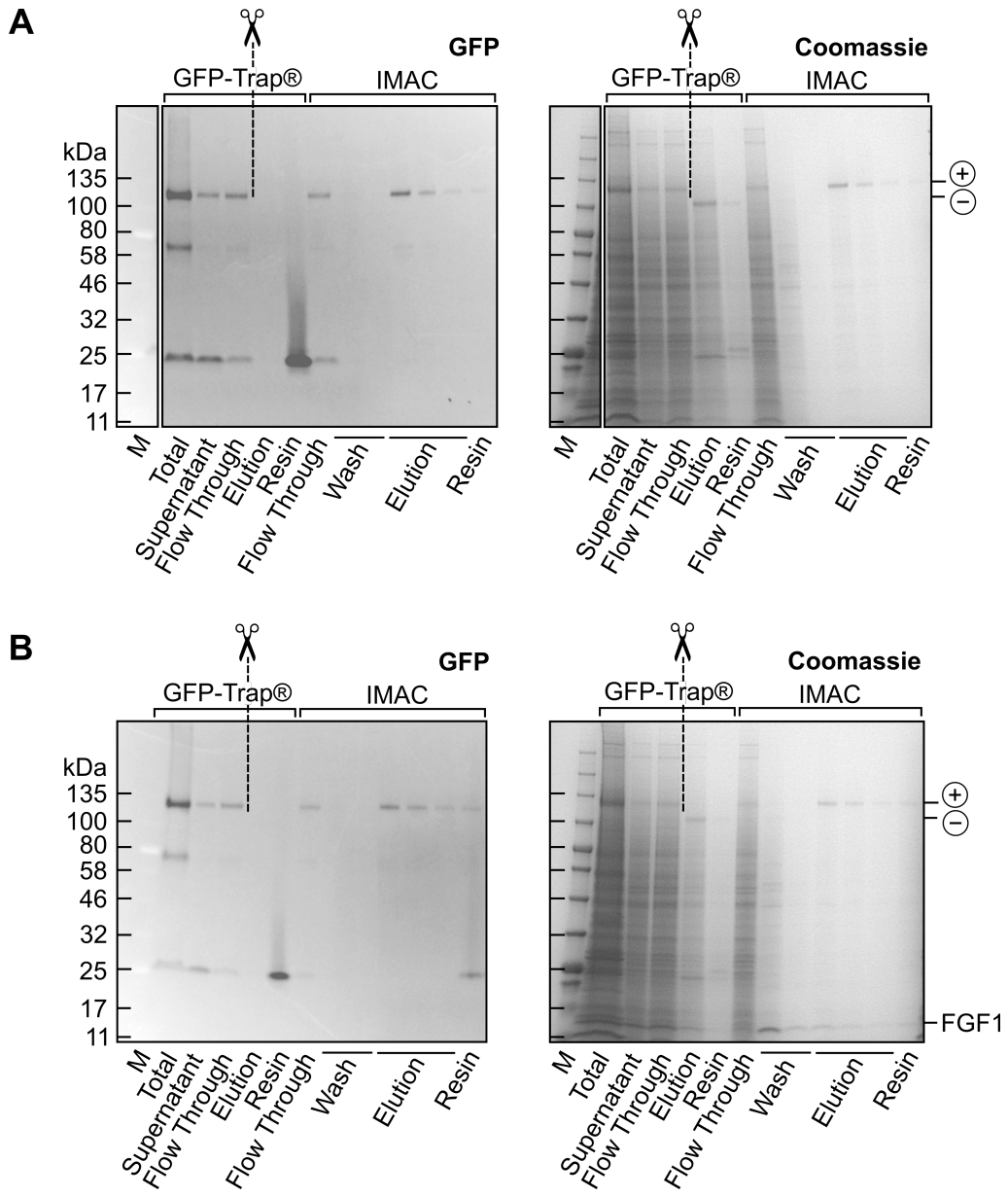


Figure 3.13. Purification of FGFR3-5F1Y in DIBMA in apo and complex forms.
To assess purification of FGFR3-5F1Y through alternative means and to probe ligand binding and complex formation as a proxy of receptor activity, a trial purification of FGFR3-5F1Y following solubilisation in DIBMA was completed. As detailed in 2.6.1.3, following receptor solubilisation, the soluble fraction was incubated with GFP-Trap® resin for 2 hours, then unbound proteins recovered. To obtain untagged receptor, HRV 3C protease was incubated with the GFP-Trap® resin overnight, then cleaved proteins collected ('GFP-Trap® Elution'). As a considerable fraction of the receptor did not bind to the GFP-Trap® resin due to its saturation with GFP, proteins present in the unbound 'Flow Through' fraction of the GFP-Trap® purification were further purified through IMAC. Briefly, this fraction was bound to Nickel Sepharose Fast Flow 6 resin overnight, the resin washed to remove contaminant proteins, then the receptor eluted using FGFR3 His Elution Buffer containing 250 mM imidazole. On this occasion, the C-terminal mGFP-His8 tag was not removed through cleavage with HRV 3C protease as this afforded no purity enhancements in previous trials. DIBMA-solubilised FGFR3-5F1Y was purified both in the absence (A) and the presence (B) of the ligands FGF1 and dp8 heparin oligosaccharide. Tagged and untagged FGFR3-5F1Y are indicated by '+' and '–' respectively. M = molecular weight marker.

3.2.7. Negative-stain EM of FGFR3-5F1Y

To assess the quality of the purified receptor samples and yield low-resolution structural insights of the full-length receptor, negative-stain EM grids were prepared and imaged. Grids were first prepared for unliganded DDM+CHS-purified receptor; these micrographs were crowded with particles that were heterogeneous in size and shape (Figure 3.14A). Protein aggregation was also observed in some grid areas. Given concerns that micrographs could be highly-populated with empty detergent micelles, and that these can be difficult to distinguish from those containing proteins (334), these micrographs were not picked nor processed. Efforts were instead focused on those recorded for the receptor purified in DIBMA that should not suffer this problem.

While some protein aggregation was also observed in micrographs corresponding to DIBMA-purified receptor (both with and without ligands), a better distribution of particles allowed image processing, though some heterogeneity in particle size and shape was also noted, possibly reflecting impurities in the sample (Figure 3.14B). Following the manual picking of particles from micrographs of each sample, 2D classification produced a number of classes with distinct shapes and features (Figure 3.14C). Though interpretation of 2D classes is challenging, especially at this resolution, a number of these obtained classes were similar to one another (highlighted with an asterisk, Figure 3.14C). Given the shape of these classes, it is possible that these could be side views of an FGFR receptor dimer, with the large ring-like density corresponding to the dimeric extracellular domain (ECD) and the smaller density to kinase domain(s) (Figure 3.14C), similar to arrangements observed previously with EGFR (Figure 1.8). However, these classes were observed similarly for both the liganded and unliganded receptors with no distinct differences, suggesting that receptors in both these instances do not differ. Should these classes truly represent dimers of FGFR3 (which have been suggested to exist in the cell when unliganded (39)), it is conceivable that ligand binding may not be identifiable in these data at this resolution because the mass increase provided by FGF1 is too small (each copy of FGF1 is only 16 kDa). Moreover, the conformational changes induced by ligand binding could be too subtle to detect.

Many of the remaining classes are not consistent in size nor shape. While these may correspond to different views of the receptor, it is more likely that these might correspond to impurities in the sample. It is noteworthy here that DIBMA discs are reported to be between 18 and 35 nm in diameter (244). As contributions from DIBMA discs are not observed in the proposed side views of the receptor, it is possible that these classes may also correspond to contaminants, or are overinterpretations of the densities in the obtained 2D classes. Negative-stain EM analyses of the receptor were thus considered inconclusive and will require further exploration in the future.

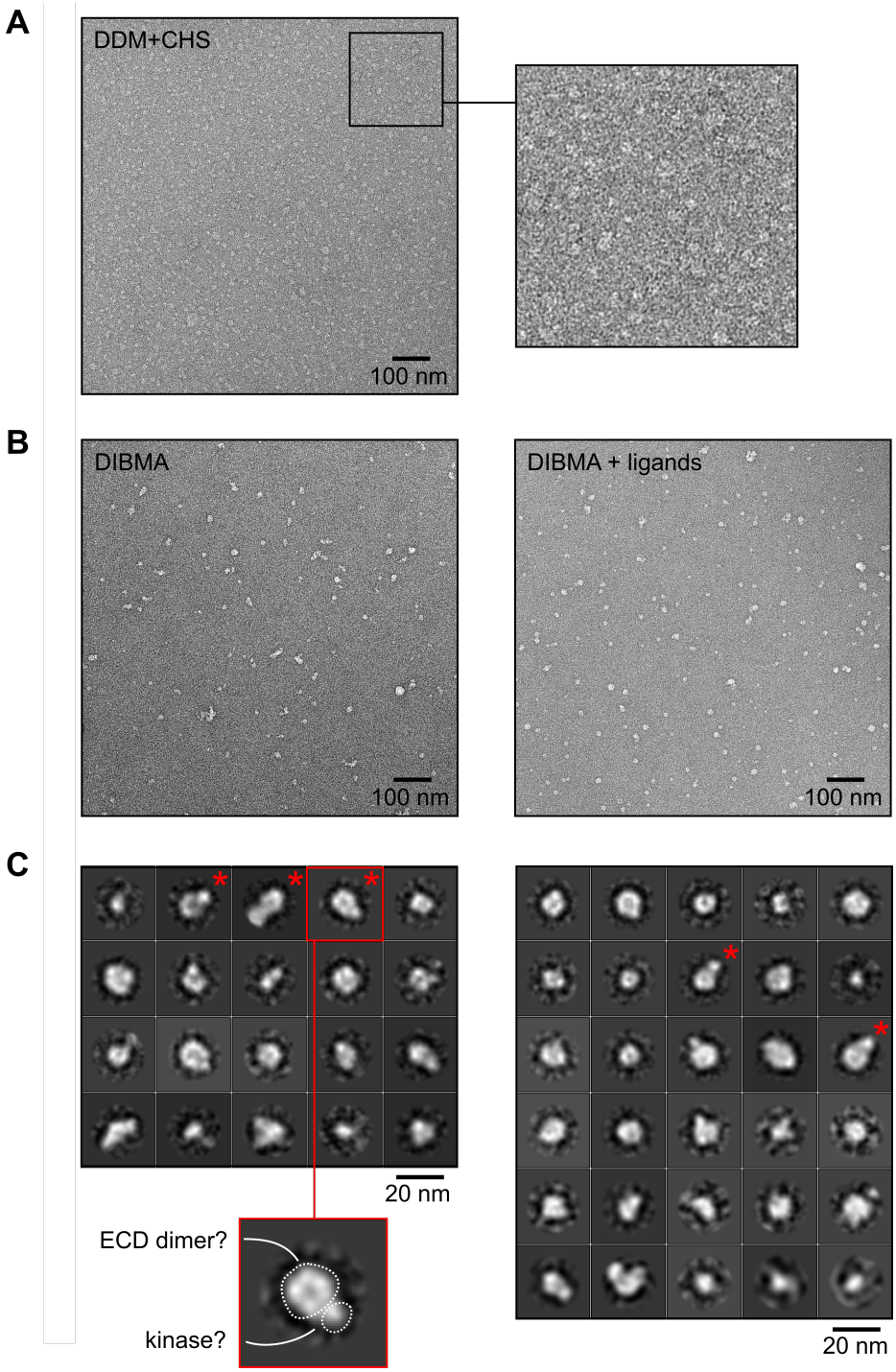


Figure 3.14. Negative-stain EM analysis of full-length FGFR3.
To assess the integrity of purified FGFR3-5F1Y, negative-stain EM grids were prepared of the receptor purified using IMAC following solubilisation in DDM+CHS or DIBMA. Grids were prepared and imaged as detailed in 2.13.1.1. (A) An example micrograph of DDM+CHS-purified FGFR3-5F1Y with an expanded panel illustrating the heterogeneity in size and shapes of the observed particles. (B) Example micrographs of DIBMA-purified FGFR3-5F1Y in the absence (left) and presence (right) of ligands FGF1 and heparin oligosaccharide dp8. (C) 2D classes of DIBMA-purified FGFR3-5F1Y in the absence (left) and presence (right) of ligands. Classes are ordered left-to-right, top-to-bottom by their fractional occupancy. Similar classes identified in each are highlighted by a red asterisk and a speculative annotation of the densities of one example class given. Image processing is described in detail in 2.13.2.1.

3.2.8. Activity assays raise concerns regarding FGFR3-5F1Y kinase activity

To ensure that structural and biochemical findings derived from recombinantly expressed proteins are physiologically relevant, it is important to verify that these proteins are functionally active. To assay the activity of purified FGFR3-5F1Y, phosphorylation of the activation loop in the presence of ATP can be probed. This can be readily achieved through western blotting with a commercially available antibody raised against a peptide encompassing the phosphorylated YYKK motif of the activation loop (Table 2.2).

3.2.8.1. *In vitro* autophosphorylation assays of detergent-purified receptors

The ability of purified FGFR3-5F1Y to autophosphorylate itself was first assayed using LMNG+CHS-purified receptor, tested through incubation of the receptor with ATP for 3 hours on ice. As FGFRs (and other RTKs) have an intrinsic basal level of kinase activity, it was anticipated that incubation of the purified receptor with ATP would result in an increase in phosphorylation at Y649 which could be detected through western blotting with the prior mentioned anti-phosphoFGFR antibody. However, while a slight difference between the start and end samples was observed, it was not significant enough to be quantified (Figure 3.15A). Additionally, several species other than that expected were also detected by the antibody, raising concerns about its specificity (Figure 3.15A).

As the YYKK motif containing the phosphorylatable tyrosine residues to which the antibody was raised against had been mutated to YFKK in FGFR3-5F1Y, conceivably, it was possible that if successful phosphorylation of this motif has occurred, it could be poorly detected by the anti-phosphoFGFR antibody due to technical limitations. To explore this possibility, western blotting using this antibody was tested against a control sample of a variant of FGFR2 kinase domain which also harbours the YFKK motif found in FGFR3-5F1Y (kindly provided by Sasha Evans). This kinase (FGFR2-KD*) was previously confirmed to be phosphorylated at Tyr656 (corresponding to Y653 of FGFR1 and Y649 of FGFR3) of its YFKK motif by mass spectrometry (Appendix 4). It was also successfully detected by the anti-phosphoFGFR antibody (Figure 3.15B), suggesting that mutation of the YYKK motif to YFKK as in FGFR3-5F1Y should not prevent detection of auto-phosphorylation of this motif should it occur. The data therefore suggested that FGFR3-5F1Y does not show strong basal kinase activity following its purification in LMNG+CHS.

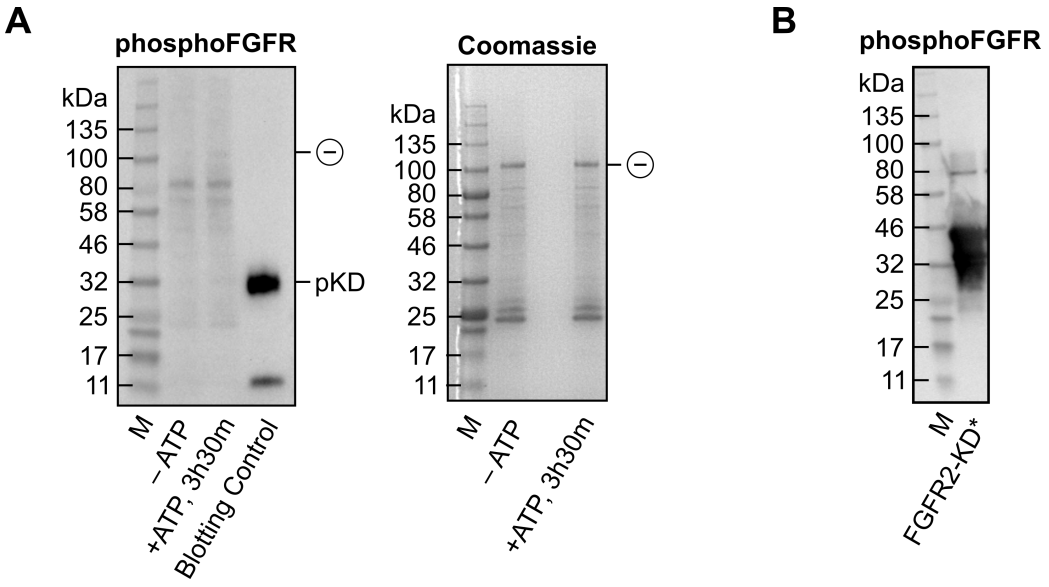


Figure 3.15. Probing basal autophosphorylation activity of FGFR3-5F1Y purified in LMNG+CHS.
(A) To assess the basal autophosphorylation activity of FGFR3-5F1Y following its purification in LMNG+CHS, the receptor was incubated in the presence of 4 mM ATP, 4 mM $NaVO_3$ and 20 mM $MgCl_2$ for 3 hours on ice. Samples were taken before and after ATP addition for SDS-PAGE and western blotting with an anti-phosphoFGFR antibody to detect phosphorylation at Y649 of the YFKK motif of the receptor. Phosphorylated FGFR3 kinase domain (pKD) was used as a blotting control. In parallel, the same samples were analysed by Coomassie to verify the presence of the untagged receptor in each (indicated by '–'). (B) To investigate whether the anti-phosphoFGFR antibody is able to detect phosphorylation of a YFKK motif (given that it is raised against the phosphorylated YYKK motif of FGFR receptors), western blotting was completed using a sample of FGFR2 kinase domain that both harbours this YYKK>YFKK substitution and is phosphorylated at this position (see Supplementary Figure 2). This analysis indicated that the anti-phosphoFGFR antibody can detect phosphorylation of a YFKK motif, implying that the inability to detect phosphorylation of FGFR-5F1Y at this motif with the same antibody suggests that the receptor does not possess basal autophosphorylation activity following purification in LMNG+CHS.

As it has been previously shown that detergents effect the activity of EGFR (124), auto-phosphorylation activity of FGFR3-5F1Y was next explored in DDM+CHS. On this occasion, the purified receptor was again incubated on ice in the presence of ATP but with samples collected at various time points following its addition. However, while FGFR2-KD* (used as a blotting control) was detected by anti-phosphoFGFR, no FGFR3-5F1Y samples were (Figure 3.16A). Curiously, given that Y649 is thought to be the only phosphorylatable Tyr residue retained in the kinase domain of FGFR3-5F1Y, probing of the same samples with a generic anti-phosphoTyr antibody indicated that the receptor is phosphorylated in all samples, both before and after ATP addition; however, no considerable difference in phosphorylation level was detected over the incubation period (Figure 3.16A). It is possible that this apparent contradiction between western blotting with anti-phosphoTyr and anti-phosphoFGFR antibodies may simply reflect differing

sensitivities of each of these (and signal amplification through use of different anti-mouse and anti-rabbit secondary antibodies), with each reporting on a minor population of phosphorylated receptor. However, even so, these data once more suggested that the purified receptor did not show any basal catalytic activity as no changes in signal were observed following incubation with ATP (Figure 3.16A). Equal protein loading was verified through western blotting with an anti-FGFR3 antibody (Table 2.2 and Figure 3.17A).

To explore whether the purified receptor was maximally-phosphorylated at Y649 before ATP addition meaning that no further signal could be expected following incubation with ATP, the auto-phosphorylation assay was repeated but with prior dephosphorylation of the receptor with lambda phosphatase. Additionally, to explore whether kinase activity of the receptor could be stimulated above its basal level through binding of its ligands (a process which should also alleviate any limitations in activity detection due to the receptor having a low basal kinase activity), this assay was also completed in the presence of FGF1 and heparin oligosaccharide dp8 in parallel. Following prior dephosphorylation of the receptor, a minor increase in phosphorylation over the course of the ATP incubation was observed at a molecular weight correlating to the untagged receptor in anti-phosphoTyr blotting; however, once more, no signal was detected using the anti-phosphoFGFR antibody (Figure 3.16B). The presence of the ligands FGF1 and heparin dp8 had no effect on the phosphorylation response (Figure 3.16B). Despite the previous assay indicating that DDM+CHS-purified receptor was phosphorylated at a Tyr residue prior to ATP incubation, no equivalent signal was detected here in the samples of the receptor prior to dephosphorylation (Figure 3.16B).

As the control FGFR2-KD* sample responded to phosphatase treatment and ATP addition as expected (Figure 3.16B), this observed lack of significant receptor autophosphorylation likely implies that the receptor is also inactive following its purification in DDM+CHS rather than an error in the experimental set up. While this is the case, it should be emphasised that this assay was completed on a sample of the receptor which had previously been flash frozen and stored at -80°C. As freeze-thawing is detrimental to protein quality and stability, it is possible that the lack of kinase activity in this instance is at least partly a consequence of this process.

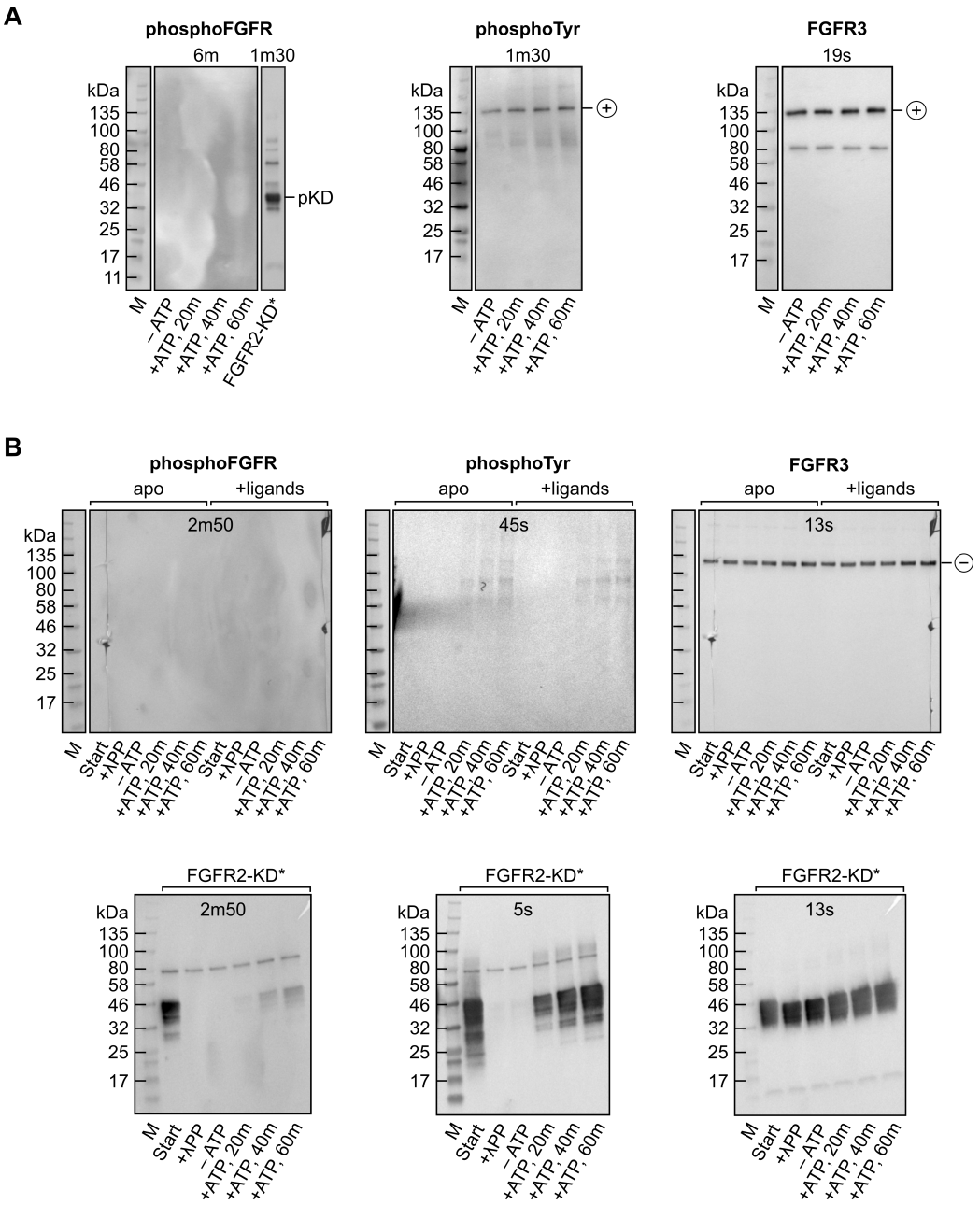


Figure 3.16. Probing autophosphorylation of FGFR3-5F1Y in DDM+CHS.
(A) To assess whether the detergent in which the receptor is purified in could affect receptor activity, the basal autophosphorylation of FGFR3-5F1Y was next assayed following its purification in DDM+CHS. As detailed in 2.10.2, the purified receptor was incubated in the presence of ATP at room temperature and samples taken at various time points for analysis by SDS-PAGE followed by western blotting (see 2.2.7). Following protein transfer, the membrane was probed with an anti-phosphoFGFR antibody, the blot stripped then reprobed with an anti-phosphoTyr antibody, then the blot stripped once more and reprobed for a final time with an anti-FGFR3 antibody. The values above each blot correspond to the exposure time during signal development with ECL reagent. Phosphorylated FGFR2 kinase domain (pKD, also labelled FGFR2-KD*) was used as a blotting control. (B) As in (A), the autophosphorylation assay was repeated except that the sample was dephosphorylated with lambda phosphatase (+λPP) prior to incubation with ATP. On this occasion, the assay was completed in both the absence (*apo*) and presence of ligands FGF1 and dp8 heparin oligosaccharide (+ligands) as detailed in 2.10.2. FGFR2-KD*, shown beneath, was used as a positive control to assess the assay integrity. Blots were processed and reprobed in the same order as in (A). Tagged and untagged FGFR3-5F1Y are indicated by '+' and '–' respectively.

3.2.8.2. Assaying receptors in native-like membrane environments

Thus far, data from activity assays has given only minor but generally unconvincing evidence that FGFR3-5F1Y retains catalytic activity following its purification. As it is not uncommon that membrane proteins are inactivated during the solubilisation and purification process and can require reconstitution into a more membrane-like environment to restore function (though this effect varies between proteins and protein families) (235), the kinase activity of FGFR3-5F1Y was next explored in environments more similar to the cell membranes from which it was extracted. More specifically, receptor activity was assayed following membrane preparation of Sf9 cells expressing FGFR3-5F1Y, and following solubilisation of the receptor into DIBMA lipid nanoparticles.

When repeating the stimulation-based autophosphorylation kinase assay for the receptor present in prepared membranes as was completed following purification in DDM+CHS, upon incubation with ATP, an immediate increase in western blotting signal was detected for several species after blotting with the generic anti-phosphoTyr antibody, including at the expected molecular mass of FGFR3-5F1Y (Figure 3.17A). An increase in blotting signal was also observed, this time in a more time-dependent manner, when blotting with the anti-phosphoFGFR antibody (Figure 3.17A). This occurs despite a slight decrease in the signal detected with the anti-FGFR3 antibody indicating receptor losses over the course of the assay (Figure 3.17A). Though these increases in blotting signal suggest that the receptor may retain kinase activity when present in a more lipid-rich environment prior to solubilisation, as the presence of the ligands FGF1 and heparin do not stimulate this apparent activity in any way, it is possible that these phosphorylation events may instead report on activities of tyrosine kinases that are endogenous to Sf9 cells also present in these prepared membranes.

To explore this possibility further, the assay was repeated on FGFR3-5F1Y purified and isolated from endogenous kinases in DIBMA lipid nanoparticles. These nanoparticles retain the receptor in a membrane-like environment with lipids derived from Sf9 cell membranes, unlike following the solubilisation of the receptor in detergents. Here, while both generic anti-phosphoTyr and anti-phosphoFGFR antibodies indicated that the receptor was phosphorylated prior to treatment with phosphatase, no recovery of phosphorylation was detected over the 40-minute incubation period with ATP, even in the presence of its ligands (Figure 3.17B). Despite appearing marginally phosphorylated by western blotting, submission of the DIBMA-purified receptor to mass spectrometry for post-translational modification identification did not identify any tyrosine phosphorylation of the receptor (Supplementary Figure 3). Altogether, these results therefore suggested that FGFR3-5F1Y did not possess kinase activity after purification.

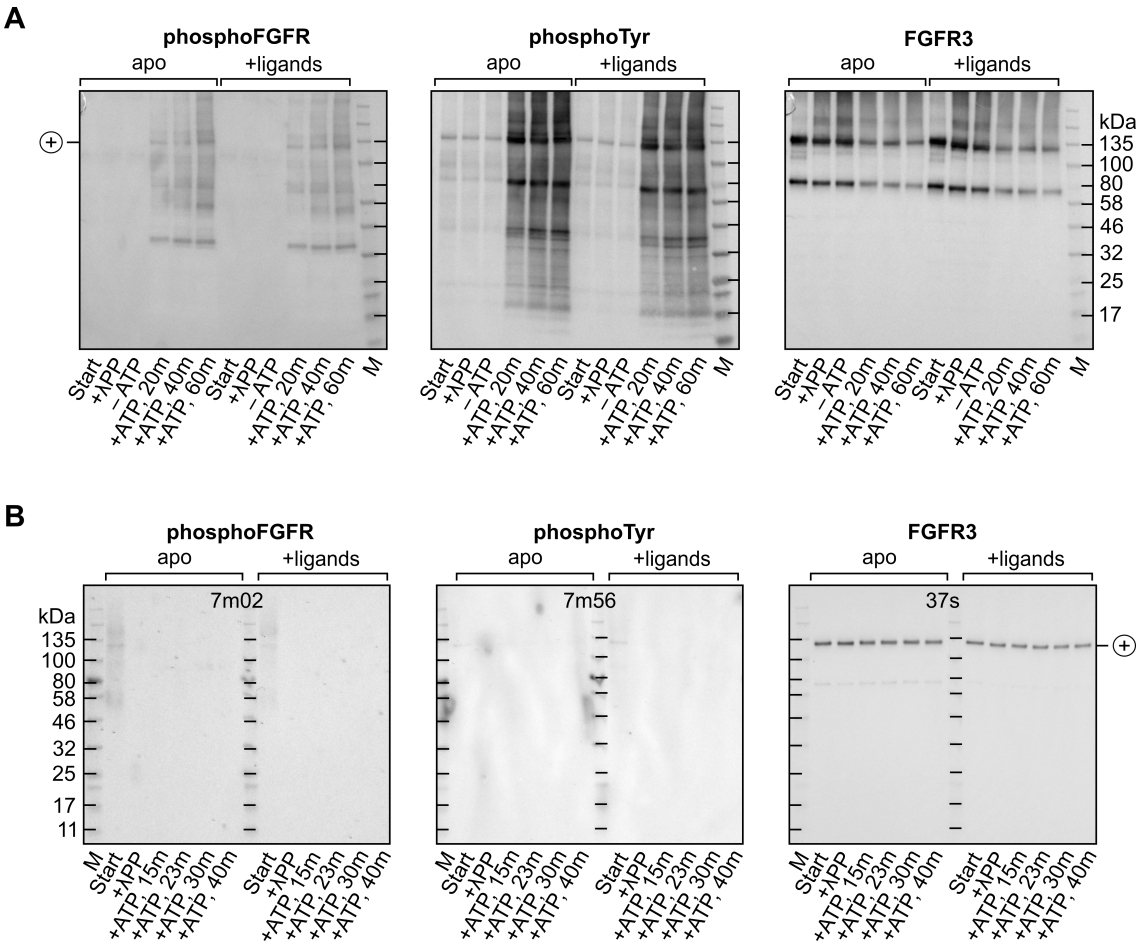


Figure 3.17. Probing autophosphorylation of FGFR3-5F1Y in membrane-like environments.

To assess whether FGFR3-5F1Y was kinase active in a more lipidic, membrane-like environment, the autophosphorylation activity of the receptor was probed following isolation of the membrane fraction from Sf9 cells (panel A, see 3.2.6.1) and following its purification in DIBMA (panel B, see 3.2.6.3). In both instances, autophosphorylation of the receptor at the YFKK motif of the activation loop was assayed both in the absence (*apo*) and presence of ligands FGF1 and heparin oligosaccharide dp8 (+ligands). Prior to incubation with ATP, samples were first dephosphorylated with lambda phosphatase (+λPP). Reaction samples were taken at various timepoints for analysis by SDS-PAGE followed by western blotting (see 2.2.7). Following protein transfer, the membrane was probed with an anti-phosphoFGFR antibody, the blot stripped then reprobed with an anti-phosphoTyr antibody, then the blot stripped once more and reprobed for a final time with an anti-FGFR3 antibody. The values above each blot correspond to the exposure time during signal development with ECL reagent. In each, tagged FGFR3-5F1Y is indicated by '+'.

3.2.8.3. FGFR3-5F1Y autophosphorylation activity in Sf9 cells

Altogether, *in vitro* autophosphorylation assay data suggested that FGFR3-5F1Y was inactive following its purification. Before pursuing further receptor characterisation, an Sf9-based stimulation assay was devised to probe activity of the recombinantly expressed receptor while present in Sf9 cells, and prior to its purification. This would help disentangle whether Sf9-expressed FGFR3-5F1Y is always inactive, or whether it loses its activity as a consequence of the purification process. This assay would use the same anti-phosphoFGFR antibody to detect receptor autophosphorylation despite its possible sensitivity problems, as the generic anti-phosphoTyr antibody would be too non-specific on this occasion due to endogenous Sf9 tyrosine kinases. Cells expressing 'wild-type', 'constitutively active' and 'kinase-dead' variants of FGFR3-5F1Y were assayed. Non-transfected Sf9 cells were used in parallel as a control to assay the non-specificities of each antibody.

Sf9 cells were stimulated with FGF1 and heparin in triplicate, and samples taken at various time points for analysis. Unfortunately, analysis of these data was hampered by a significant non-specificity of the anti-phosphoFGFR antibody given that it detected multiple species in all conditions, even in the non-transfected control cells (Figure 3.18). Interestingly, this finding possibly also indicates that phosphorylation detected by this antibody in earlier *in vitro* experiments need not correspond to Y649 phosphorylation at all. In addition to the non-specificity of the anti-phosphoFGFR antibody, blotting with anti-FGFR3 of these same samples following the stripping and reprobing of the blot indicated that the receptor was present in multiple truncated species (Figure 3.18). Though the receptor truncations observed here were not met in prior expressions of the receptor (Figure 3.17), their presence in these assay samples indicated a possible incapability of Sf9 cells to reliably express full-length receptors.

Due to these issues discussed with antibody non-specificity and receptor truncation, these data were considered unreliable and were not analysed further. Analysis of kinase activity of FGFR3-5F1Y while in Sf9 cells was thus not achieved in the time-frame of the project.

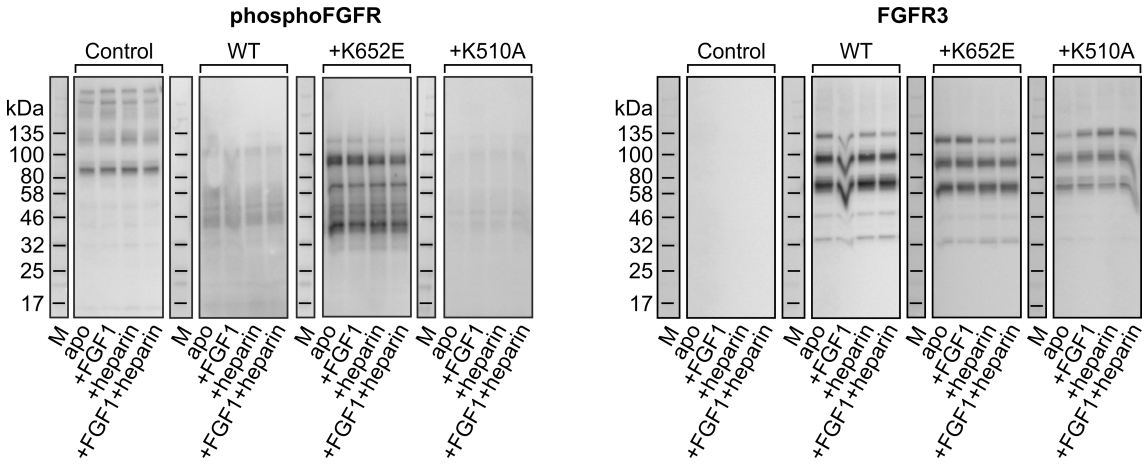


Figure 3.18. Probing autophosphorylation of FGFR3-5F1Y in Sf9 cells.
To probe autophosphorylation activity of FGFR3-5F1Y when present in a cellular environment, Sf9 cells expressing either 'wild-type' (WT), constitutively-active (+K652E), or 'kinase-dead' (+K510A) variants of FGFR3-5F1Y were incubated with FGF1 or heparin alone, or with FGF1 and heparin simultaneously. To measure any non-specific background activity, control non-transfected Sf9 cells were also assayed, and each also treated with no ligands (apo) as controls. As detailed in 2.10.1, at various time points, cells were harvested and lysed, then the total protein content of each cell lysate determined. 20 µg of each sample was analysed by SDS-PAGE followed by western blotting, first with an anti-phosphoFGFR antibody, then after stripping, reprobed with an anti-FGFR3 antibody. For clarity, while each was performed in triplicate, only one of these repeats is shown of samples taken 10 minutes following stimulation with ligands FGF1 and heparin. For each antibody, each blot panel was exposed for the same amount of time. M = molecular weight marker.

3.2.9. Towards an improved sample for the structural determination of FGFR3

The results presented thus far detail efforts towards solving a high-resolution structure of full-length FGFR3. As a consequence of facing multiple challenges including relatively poor-quality protein expression (see 3.2.6), indications of poor or no kinase activity (see 3.2.8), and inconclusive preliminary negative-stain EM analyses (see 3.2.7), this aim was not realised during this thesis project. Looking ahead, there are several avenues which could be explored that could lead to successful structure determination in the future; some of these and preliminary work towards them are discussed here.

3.2.9.1. Fusion constructs of FGFR3

Though technical developments in cryo-EM in recent years have led to substantial improvements in structure determination by this technique, low molecular mass samples still present challenges. This can be particularly true for membrane proteins where in many cases very little of the protein extends out of the membrane mimic environment. At a molecular weight of ~89 kDa as a monomer and ~210 kDa as a liganded dimer, FGFR3-5F1Y lies at the lower end of the molecular weight range of proteins structures solved by cryo-EM (297). However, as the membrane-spanning region of FGFR3 is only a single helix and thus most of the mass of the receptor lies outside of the membrane

environment, it was thought that the receptor would be readily identifiable in EM micrographs for structural characterisation. However, given the difficulties faced during interpretation of preliminary negative-stain EM analysis of the receptor, opportunities to increase the mass of the receptor could be worth exploring. Fortunately, this could possibly be realised in the form of FGFR3 receptor fusions. As detailed previously, these pathologically-relevant fusions introduce predicted coiled-coil domains on the C-terminus of FGFR3, and can increase the mass of the receptor dimer by up to ~146 kDa in the case of FGFR3-RT4 (i.e. a total liganded dimer mass of ~356 kDa). As coiled-coils are rigid domains, they may also minimise conformational heterogeneity of the receptor kinase domains, and by likely projecting away from the membrane environment, could additionally provide a distinctive feature for particle picking and image processing. Moreover, these coiled-coil domains may rigidify the intracellular domain of the receptor, trapping it in a distinct state that may enable its visualisation in EM, as was not possible for the insulin receptor due to flexibilities in this region (133).

To explore the viability of recombinantly expressing FGFR3-5F1Y fusion constructs in insect cells for structural determination, transfections of Sf9 cells were completed for ‘RTT112’, ‘RT4’, ‘SW780’, and ‘deldel’ variants using bacmids derived from DH10Bac™ cells. Though lighter than FGFR3-5F1Y, the ‘del19’ variant of the receptor was also transfected, as structural determination of this oncogenic aberration could also provide mechanistic insights regarding receptor regulation. Analyses of these transfections indicated successful expression of ‘RTT112’, ‘SW780’, ‘deldel’ and ‘RT4’ variants of FGFR3-5F1Y (Figure 3.19); expression of ‘del19’ was either unsuccessful or beyond detection. While fluorescence microscopy showed that the expressed receptor was generally localised to the plasma membrane of Sf9 cells, in some cases, the localisation of fluorescence to discrete punctae or incomplete cell ‘halos’ was observed (Figure 3.19A), as seen previously with ‘wild-type’ receptor (Figure 3.2). At this stage, it is not possible to discriminate whether expression differences between constructs are real or a consequence of poor transfection efficiency. Nonetheless, the observation of expression from these preliminary results is encouraging, as they suggest that fusion constructs of FGFR3 could present alternatives for structural determination in the future.

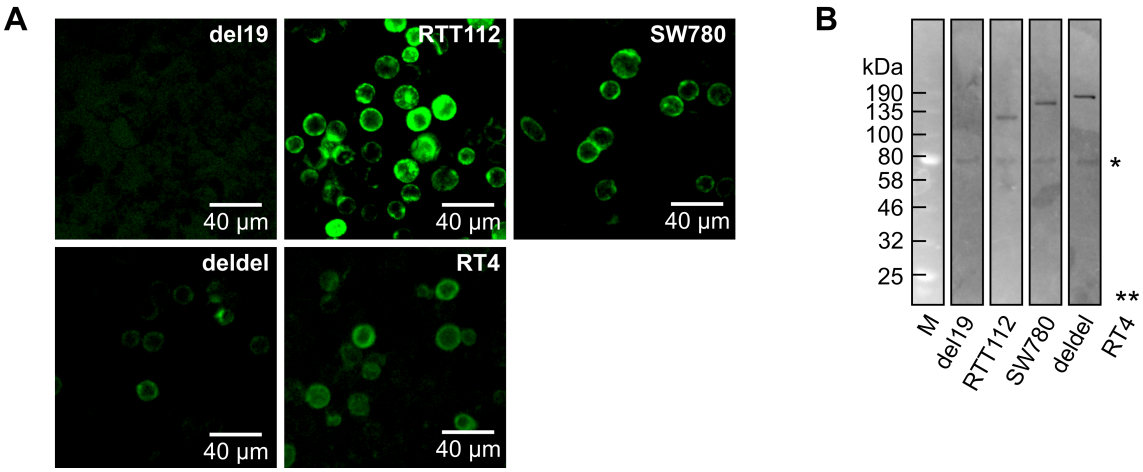


Figure 3.19. Expression of FGFR3-5F1Y fusion constructs in Sf9 cells.
(A) Example fluorescence microscopy images of Sf9 cells following transfections with bacmids derived from DH10EmBac™ cells encoding FGFR3-5F1Y-del19 and RTT112, SW780, deldel and RT4 fusion constructs of FGFR3-5F1Y, imaged 6 days following bacmid transfection (see 2.4.3) Fluorescence images were recorded using the GFP fluorescence mode (λ_{ex}: 470 nm, λ_{em}: 525 nm) on a EVOS FL Auto 2 Imaging System (Thermo Fisher Scientific). (B) In-gel GFP fluorescence analysis of SDS-PAGE separated transfection samples (as in (A)), illustrating expression of full-length RTT112, SW780 and deldel fusion constructs. Note that the Sf9 insect cells express an endogenous protein that fluoresces at approximately 80 kDa (indicated by an asterisk (*)), and that the corresponding analysis for the RT4 fusion was not completed (**).

3.2.9.2. Exploring the binding of adaptor protein FRS2α to FGFR3

The formation of larger complexes of FGFR3 bound to functionally-relevant binding partners (e.g. signalling adaptor proteins) may be a further avenue to improve receptor samples for structural determination. Adaptor proteins could be useful in two ways: by increasing the molecular mass of the protein complex thus improving signal to noise in EM, and through potential rigidification effects should they bind in a flexible region of the receptor. FRS2α, an adaptor protein linking FGFR and MAPK signalling, is a candidate protein in this regard that binds constitutively to the juxtamembrane domain of FGFR1 through its phosphotyrosine binding (PTB) domain in a non-canonical, phosphotyrosine-independent manner (59,335). As the juxtamembrane domain of FGFRs could be a source of conformational flexibility in the receptor, binding of FRS2α at this region may help to rigidify this region. However, this interaction between the PTB domain of FRS2α and the juxtamembrane domain of FGFRs has only been explicitly demonstrated *in vitro* and *in cellulo* for FGFR1. Although it has been implied that FGFR3 may not bind FRS2α due to differences in the sequence of its juxtamembrane domain with respect to that of FGFR1 (336,337), FRS2α-mediated FGFR3 signalling has been reported in cells (338). For rigidification of the juxtamembrane domain by FRS2α to be an option for FGFR3 structural studies, it thus first needed to be demonstrated that FRS2α PTB domain does indeed interact with the juxtamembrane domain (JMD) of this receptor.

To yield *in vitro* binding data to demonstrate whether this was the case, FRS2α PTB domain was expressed in and purified from *E. coli*, based on protocols from previous studies (59,321,339). Expression trials of three different constructs of the PTB domain with different domain boundaries and purification tag locations (see 2.5.1) identified that a construct corresponding to residues 9-122 of FRS2α with a C-terminal His6 tag ($FRS2\alpha^{C9\text{-}122}$) expressed well (Figure 3.20A). Large-scale expression and purification of $FRS2\alpha^{C9\text{-}122}$ yielded ~150 µg/L of culture at >95% purity, a yield which was sufficient for preliminary biophysical analyses. Though expressing well, significant protein losses were met at several stages, limiting the final protein yield that could be obtained and thus the number of experiments that could be completed. These losses occurred primarily at three steps: at cell lysis, where very little target protein was available in the soluble fraction; during IMAC purification, where a high amount of target protein remained bound to the nickel resin despite extensive washing with 250 mM and 500 mM imidazole; and during gel filtration, where only an estimated 20% of total $FRS2\alpha^{C9\text{-}122}$ was present in its monomeric form (Figure 3.20B and C). Mass spectrometry confirmed that the purified protein was of the expected mass (Figure 3.20D).

Isothermal titration calorimetry (ITC) was used to probe the binding between FRS2α PTB domain and FGFR3 JMD *in vitro*. Preliminary experiments confirmed that a peptide corresponding to the juxtamembrane domain of FGFR3 (synthesised by Dr Zsofia Hegedus (University of Leeds)) unequivocally binds to FRS2α PTB domain, with a reported affinity of interaction in the low micromolar range (K_d of 1.6 µM) and a 1:1 binding stoichiometry (Figure 3.21). These findings are in the same range as has been reported for the binding of FGFR1 juxtamembrane domain to similar FRS2α PTB domain constructs (339), suggesting that binding to this region is indeed conserved among FGFR1 and FGFR3 despite minor sequence differences between these two receptors at the juxtamembrane domain. Due to limiting protein amounts, further control experiments with FGFR1 juxtamembrane domain and $FRS2\alpha^{C9\text{-}122}$ have not yet been completed, precluding a direct comparison between the affinities of each JMD peptide.

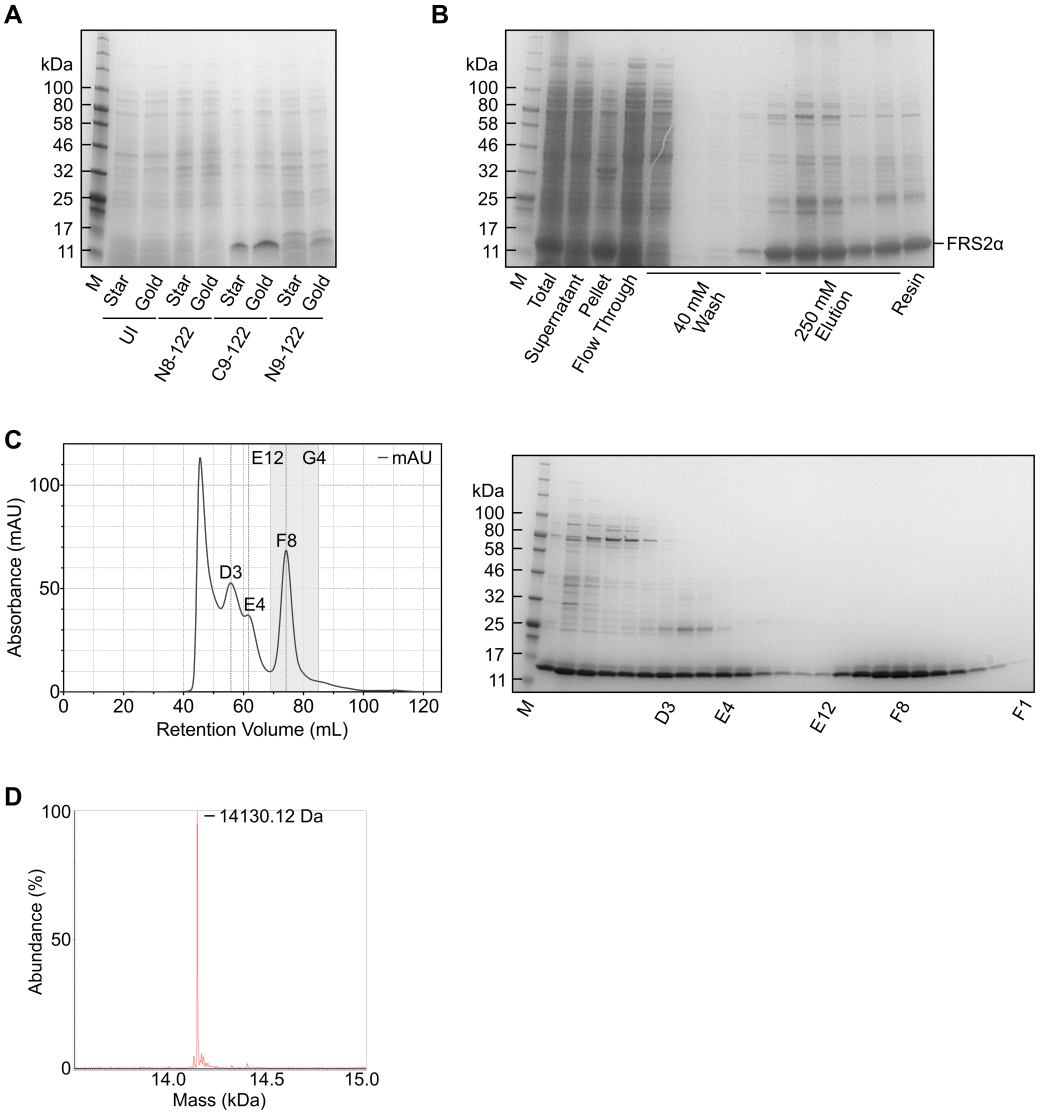


Figure 3.20. Purification of FRS2α PTB domain.
(A) To optimise expression of FRS2α PTB domain, three constructs (FRS2α^{N8-122}, FRS2α^{N9-122} and FRS2α^{C9-122}, as detailed in 2.3.2) were expressed in BL21 Star pRARE2 and BL21 Gold pRARE2 *E. coli* cells. To assess expression, samples were taken before (UI = uninduced) and after expression induction, then analysed by SDS-PAGE with Coomassie staining. This indicated that FRS2α^{C9-122} was the best expression construct. (B) Following large-scale expression (see 2.5.1.2), FRS2α^{C9-122} was purified by IMAC using Nickel Sepharose Fast Flow 6 resin (as detailed in 2.6.2). Briefly, following cell lysis, proteins were immobilised onto nickel resin, then the resin washed with FRS2α wash buffer containing 40 mM imidazole to remove contaminant proteins. Bound proteins were subsequently eluted with buffer containing 250 mM imidazole. Samples from each stage were analysed by SDS-PAGE followed by Coomassie staining. The presence of FRS2α PTB domain is highlighted. (C) To further purify FRS2α PTB domain, the IMAC elution fractions from (B) were applied to a HiLoad™ Superdex 75 16/600 column to separate the PTB domain from contaminant proteins. As illustrated in the SEC chromatogram, FRS2α PTB domain eluted in multiple species. Fractions belonging to monomeric FRS2α PTB domain (highlighted grey) were combined and concentrated to obtain the final purified protein sample. (D) To assess their protein composition, fractions from SEC (panel C) were analysed by SDS-PAGE followed by Coomassie staining. The labelled fractions correspond to those labelled in the chromatogram in panel C. (E) The final purified FRS2α PTB domain sample was assessed by intact mass determination (see 2.2.12). The detected mass of 14130.12 Da matched the expected mass of the PTB domain (14130.88 Da).

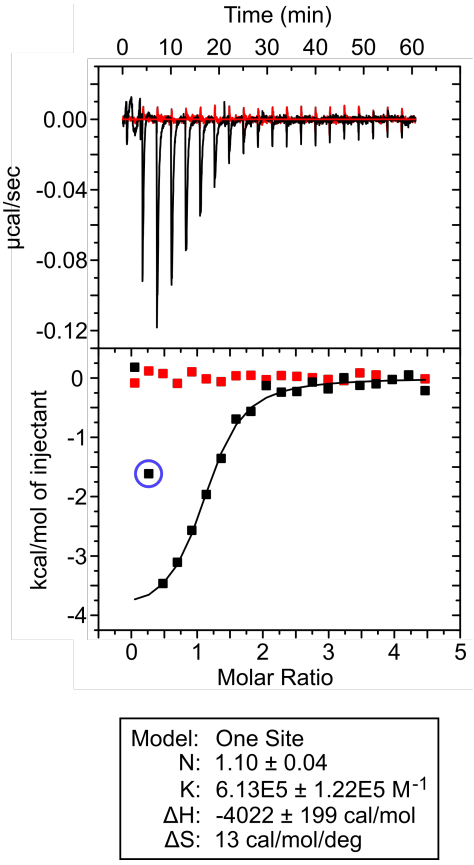


Figure 3.21. Preliminary ITC data demonstrates the binding between FRS2α PTB domain and FGFR3 juxtamembrane domain.
ITC thermogram showing raw data of heats released (top panel) following injection of a peptide corresponding to FGFR3 juxtamembrane domain (residues 409-428) into FRS2α PTB domain (black trace) or buffer (red trace), and integrated data for thermodynamic fitting (bottom panel). Each trace corresponds to a single measurement. A probable outlier, circled in blue, was not used during curve fitting. Provided beneath these traces are the values calculated from this curve fitting detailing the number of binding sites (N), the association constant (K), the change in enthalpy (ΔH), and change in entropy (ΔS) of the binding interaction. The dissociation constant K_d (~1.6 µM) can be calculated by 1/K.

3.3. Discussion

3.3.1. Progress and challenges faced in the structure determination of FGFR3

The data presented here show that a variant of FGFR3 named FGFR3-5F1Y can be successfully recombinantly expressed in Sf9 insect cells, obtaining approximately 200-250 µg of the receptor per litre of culture after membrane preparation (Figure 3.2). The receptor can also be solubilised in a number of detergents and the co-polymers SMA and DIBMA in a non-aggregated form, a pre-requisite for protein purification and structural studies (Figure 3.6 and Figure 3.7). Though only small-scale purification trials have been completed to date, using the C-terminal HRV 3C-cleavable mGFP-His8 tag, the receptor can be purified to at least 50% purity using nickel-based resin by immobilised metal affinity chromatography (Figure 3.12). It is possible that further purity enhancements could be gained from more extensive washing prior to target protein elution. Alternatively, GFP-Trap® resin was shown to be a possible alternative for receptor purification, though protocol optimisation would be required (Figure 3.13). Though only small-scale, the yields from these purifications were amenable to activity testing and preliminary negative-stain electron microscopy analyses, essential steps to evaluate protein sample quality prior to moving to high-resolution structure determination.

Though initial protein structural insights can often be gained from surrogates with impaired biological function, ideally, the kinase activity of FGFR3-5F1Y should be demonstrated to ensure that any gained structural insights have physiological relevance. The observation of possible co-purification of FGF1 with FGFR3-5F1Y following their incubation with one another prior to solubilisation implied that the receptor may be able to bind its ligands, an indicator that it may be correctly folded and functional (Figure 3.13). However, attempts to demonstrate kinase activity of FGFR3-5F1Y purified in either LMNG+CHS, DDM+CHS or DIBMA through the receptor's ability to autophosphorylate itself at Y649 of the activation loop were generally unsuccessful. While the receptor appeared to be phosphorylated to some degree following purification, incubation of the receptor with ATP did not lead to further increases in receptor phosphorylation, either alone or in the presence of its activating ligands FGF1 and dp8 heparin oligosaccharide (Figure 3.15, Figure 3.16 and Figure 3.17). Mass spectrometry was also unable to detect phosphorylation at this residue by post-translational modification identification following protein digestion (Supplementary Figure 3). Though a possible increase in phosphorylation was detected with the receptor purified in DDM+CHS, the effect was small and was once more unaffected by the presence of FGFR-activating ligands.

Though discouraging, it is possible that this apparent lack of receptor kinase activity following purification of FGFR3-5F1Y could partly be a consequence of experimental

limitations of the system and the detection method being used. For example, it is known that the membrane mimic within which RTKs are solubilised in can considerably influence kinase activity and the ability of receptors to respond to ligand treatment, as has been demonstrated for EGFR and the insulin receptor (122,124). For detergents, this inability of a receptor to respond to its ligands could potentially be rationalised by an incapacity of this membrane mimic to allow the transmembrane helices to conformationally rearrange upon ligand binding. While two detergent systems were tested for FGFR3-5F1Y, others may also need to be assessed. To address the possibility that the membrane environment of the receptor is vital to its kinase activity, the ability of the receptor to autophosphorylate itself was also evaluated in DIBMA lipid nanoparticles. However, this system introduces a possible limitation for the detection of activity should there be restrictions in the proximity of kinase domains to allow autophosphorylation. This could be a particular concern if the receptor happens to be solubilised in a monomeric form, or if in a dimeric form but with each transmembrane helix located in its own DIBMA disc. Such an arrangement has been observed by negative-stain EM following the solubilisation of the insulin receptor in nanodiscs (122), and may restrict the ability of kinase domains to dimerise in order to phosphorylate one another. In the future, to confirm that purified FGFR3-5F1Y indeed lacks kinase activity as the current *in vitro* data suggest, an alternative assay that is independent of kinase domain proximity should be used. Such an assay could instead probe the phosphorylation of a model peptide substrate, as has been shown previously for EGFR (124). Unfortunately, time constraints in this project meant that this alternative could not be tried.

As discussed previously, given that it is not uncommon that membrane proteins can lose their activity as a consequence of protein purification, a lack of kinase activity could also indicate that purification procedures need optimisation. To assess whether this is the case, it is useful to evaluate whether the recombinantly-expressed construct is active when in its expression host and prior to purification. This was attempted for FGFR3-5F1Y through assessment of the receptors ability to autophosphorylate itself when in Sf9 cells, but an observed non-specificity of the anti-phosphoFGFR antibody critical to this assay and receptor expression complications meant that these assays were not informative (Figure 3.18). In the future, alternative assays could be explored to assess the activity of FGFR3-5F1Y in its Sf9 host. This is essential to determine whether it is appropriate to further pursue its structural characterisation. The activity of the receptor could be indirectly evaluated through assessment of the receptor's ability to stimulate downstream signalling cascades in the cell (e.g. the MAPK cascade).

3.3.2. Evaluating the appropriateness of Sf9 cells as an expression host

Sf9 insect cells and the baculovirus expression vector system (BEVS) were chosen here for expression of FGFR3-5F1Y as a compromise between expression in a eukaryotic system which would faithfully replicate expression of the receptor in its native mammalian host, versus the ease of use and experimental costs. While the BEVS system has proven successful for other proteins (20,47,223–225,126,216–222), there are indications here that this may not be the case for FGFR3-5F1Y. One indicator that this may be the case comes from the observation of additional fluorescent and anti-His western blotting-sensitive species other than the expected full-length receptor and free GFP/YFP in SDS-PAGE analyses following its expression (Figure 3.2). Though not demonstrated, these could correspond to truncated forms of the receptor, likely from the amino-terminal end (given that these species retain their C-terminal tag). Further probable indicators of poor-quality Sf9 expression of FGFR3-5F1Y include the observation of receptor aggregation in the absence of DTT in non-reducing SDS-PAGE gels (Figure 3.5), the presence of an aggregated population in Native-PAGE prior to solubilisation (Figure 3.8), and the loss of at least 70% of total expressed receptor during membrane preparation (Figure 3.9). Thus, while fluorescence microscopy had indicated that the recombinantly-expressed receptor was plasma membrane localised, suggesting that the receptor had passed all protein-folding checkpoints in the cell, this produced protein did not behave well. It is possible that many of these above observations could be rationalised by an inherent instability of the receptor, but the inability to conclusively demonstrate receptor kinase activity (see 3.2.8 and 3.3.1) provides reason to be concerned whether the expression host is to blame.

To validate the quality of FGFR3-5F1Y recombinantly expressed in Sf9 cells and its appropriateness for structural studies, additional experiments to biochemically and biophysically characterise the receptor could be performed. These could include SEC or FSEC analyses to probe protein aggregation and dimerisation status in the absence and presence of ligands, and the use of circular dichroism to assess the ‘foldedness’ of purified receptor and provide an estimation of secondary structure content to compare against what would be expected from the published structures of individual receptor domains. These assays could also be used to probe whether the receptor is simply unstable, and (in combination with further activity testing as above) determine whether protocol adaptations could be made to improve stability and functional viability of the receptor.

3.3.3. Alternative strategies for FGFR structure determination

Given the possibility that FGFR3-5F1Y expressed in Sf9 insect cells is inappropriate for further structural characterisation, alternative approaches will be required to solve a high-resolution structure of FGFR3, or indeed any member of the FGFR family.

Using mammalian cell culture rather than insect cells as the host for recombinant expression is an obvious strategy here. This move to mammalian cell lines should alleviate possible problems related to incorrect processing of receptors in insect cells by moving to a more-native host, though problems originating recombinant protein expression overwhelming protein folding machinery could still remain. Many studies have reported mammalian-based expression of FGFR3 and variants thereof in the past as a means to study receptor signalling pathways, including cysteine variants and fusion constructs. While the expression levels in these cell-based assays are likely insufficient for structural studies, they demonstrate that functional FGFR3 receptors can be recombinantly expressed in mammalian cell culture where insect cells may be failing. To explore this avenue as an option for FGFR3-5F1Y, the receptor was successfully cloned for BacMam expression, a variation of the BEVS system designed for simple and high-yield expression of recombinant proteins in mammalian tissue culture (319). However, time limitations meant that expression trials and *in cellulo* activity testing could not be completed within the timeframe of the project.

Given the problems with demonstrating receptor activity thus far, it is essential that autophosphorylation of the receptor is assayed prior to large-scale purification and characterisation from a mammalian host. It is worth reiterating here that the receptor used in these attempts of structural characterisation is not the ‘wild-type’ FGFR3 receptor but a designed variant. While the five Tyr>Phe substitutions introduced into the receptor kinase domain in FGFR3-5F1Y are not expected to have detrimental effects on receptor activity, activity testing following mammalian recombinant expression should be used to confirm that this is indeed true. It is possible that this variant design may provide an alternative explanation for the challenges faced during the efforts in this project.

As an alternative to mammalian expression, a variation of insect cell BEVS called Kinase Factory™ (Geneva Biotech) could also be tested. This commercially available system introduces co-expression of a panel of chaperone proteins with roles in the Hsp90-based chaperone system to improve the solubility, yield and activity of recombinantly-expressed human protein kinases (340), and conveniently would not require the re-cloning of constructs that have been amassed. Using Kinase Factory™ may be particularly beneficial in this project’s aims as FGFR3 is a known strong client of Hsp90 (164), and this system co-expresses the kinase-specific co-chaperone Cdc37 to improve kinase interaction with endogenous Sf9 Hsp90.

As a final alternative, a cell-free based expression could be explored. A system using *E. coli* cell lysate for expression of functional RTKs directly into nanolipoprotein particles, though at low yields, has previously been described (341). It is important, however, to consider that these systems typically lack post-translational machinery to introduce modifications such as glycosylation which may be important in receptor functionality.

If changing expression systems, it would also be recommended to revisit construct design of FGFR3 at this stage to test expression of genuine 'wild-type' receptor and other variants, rather than those with a FGFR3-5F1Y backbone. In this work, only variants of FGFR3 have been investigated rather than those of all FGFR family members. Expanding the panel of expression constructs to include other members and their respective pathological variants may improve the likelihood of identifying a receptor which is tractable to structural studies. As high-resolution structural information detailing any of the FGFR family would be a significant step forward and highly informative to understanding RTK activation mechanisms in their full-length context, restricting efforts to FGFR3 only may limit opportunities for success. As discussed previously, it may also be necessary to explore constructs and complexes of FGFRs which are larger in mass to build upon challenges faced during negative-stain EM analyses of FGFR3-5F1Y (Figure 3.14). These could come in the form of fusion constructs of FGFRs (such as those of FGFR3 demonstrated to be expressed in Sf9 insect cells (Figure 3.19)), or through the binding of adaptor proteins such as FRS2α at the receptor juxtamembrane domain (possible for FGFR1 and likely also FGFR3 (Figure 3.21)). It may also be worth considering introducing another affinity tag or upgrading the His8-tag used thus far for an alternative epitope tag to improve purification efficiency, particularly if expression levels are expected to be poor, such that protein recovery can be maximised. An alternative tag may also help minimise non-specific binding interactions with the purification resin as observed with Nickel Sepharose Fast Flow 6 resin (3.2.6).

3.3.4. Summary

In this chapter, the expression of FGFR3-5F1Y and variants thereof was successfully established in Sf9 insect cells to isolate full-length receptors for structural characterisation. These receptors could be purified following solubilisation in either detergent or DIBMA, and showed indications of probable ligand binding, though receptor activity could not be conclusively demonstrated. In finding that preliminary negative-stain EM data of the receptor were difficult to interpret, groundwork was established to prepare larger and less heterogeneous samples that may be more suitable to structural determination by EM. Although no structural data were obtained of full-length FGFR3 in this project, work detailed in this chapter lays the foundations for future structural studies of full-length RTKs.

Chapter 4. The recruitment of kinases by co-chaperone Cdc37

4.1. Aims

The next two chapters explore the regulation of kinases through their chaperoning by Hsp90. Over 60% of the human kinome has been proposed to interact with Hsp90 (155), and for many, this interaction is suggested to be critical for proper kinase function (170). In this system, the Hsp90 co-chaperone Cdc37 both recruits client kinases and remodels them before their interaction with Hsp90. As discussed previously (see 1.3.6), prior studies of kinase-Cdc37 interactions have identified that kinase recognition is a two-stage process. Specifically, a prior NMR spectroscopy study investigating the binary interaction between Cdc37 and the kinase sb-Raf^{V600E} found that kinases are first approached by the amino-terminal region of Cdc37 (N-Cdc37), followed by a secondary interaction with its carboxy-terminal domain (C-Cdc37) (173). This secondary interaction is only stable if the kinase is considered a client of Hsp90 (173). Several studies have attempted and failed to identify a common structural motif or hallmark among client kinases that separates them from non-clients. Instead, the current model of kinase dependency on Hsp90 suggests that all kinases lie somewhere on a client to non-client continuum, with their dependency relating to their global thermal instability (141,155,165). In context of kinase interaction with Cdc37, it has been proposed that the ability to form a stable binary complex that involves interfaces with both N- and C-Cdc37 relies on the ability of Cdc37 to sense this instability. Moreover, it has been suggested that the co-chaperone Cdc37 actively remodels and unfolds client kinases (173). This is supported by the observation of Cdc37-dependent structural changes of sb-Raf^{V600E} in methyl-based NMR studies, including the appearance of sharp peaks in the ‘unfolded’ region of the kinase spectrum, and an increased susceptibility to proteolytic digestion of the kinase in the presence of the Cdc37, implying structural remodelling (173). However, as these methyl-NMR data of the kinase were not assigned, those regions of the kinase experiencing remodelling and unfolding could not be precisely defined (173).

In this chapter, I aimed to better characterise the interaction between Cdc37 and kinases, using the isolated FGFR3 kinase domain as a representative client. As the Cdc37 portion in this interaction is reasonably understood (173), I would instead focus on assessing the proposed structural changes of the kinase in this binary complex. A number of aspects would be explored including how kinases and Cdc37 interact with one another, better defining which regions of a kinase are remodelled by Cdc37, and how these two aspects of this protein-protein interaction relate to one another. To explore whether these effects on kinase structure are conserved or differ among kinases that lie in different regions of the client to non-client continuum, I would study this kinase-co-chaperone interaction using both ‘wild-type’ and an oncogenic variant of FGFR3 kinase domain that

differ in client strength. Given the dynamic nature of the kinase-Cdc37 interactions, NMR spectroscopy would be employed. Building on these data, I would also use these two variants of FGFR3 kinase domain to explore whether the structural origin of kinase thermal stability and client strength could be defined. As it has been observed that ATP-competitive inhibitors stabilise client kinases and compete with Cdc37 (175), I would also explore the effects of the FGFR inhibitor PD173074 on kinase structure, in hope to determine how kinase inhibitor binding can lead to Cdc37 independence, and how this may inform on the process of recruitment of kinases to Hsp90.

4.2. Results

4.2.1. Titration of FGFR3 kinase domain with Cdc37 induces kinase unfolding

To understand the effects of Cdc37 binding on kinases during binary complex formation and recruitment to Hsp90, non-labelled Cdc37 was titrated (at molar ratios of 0.75, 1.90 and 3.00) into uniformly ^{15}N-labelled wild-type FGFR3 kinase domain (FGFR3-KD^{WT}) and 2D ^{1}H-^{15}N TROSY spectra acquired (by Dr Gary Thompson). As only the kinase domain was labelled in these data, any observed differences in these spectra would be indicative of structural changes in the kinase as a consequence of Cdc37 binding. FGFR3-KD^{WT} was previously found to be a weak client of Hsp90, estimated to show a binding affinity with Cdc37 in the millimolar range (Tom Bunney, personal communications; now published in (322)). Titration of this weak interaction was therefore expected to occur in the fast exchange regime with respect to NMR chemical shift changes, a feature that would have allowed analysis of structural changes induced by Cdc37 binding through the tracking and mapping of chemical shift changes between the free and bound states of the kinase.

Comparing spectra of FGFR3-KD^{WT} in the absence and presence of Cdc37 revealed two major effects: a dramatic overall reduction in peak intensities due to signal broadening, and the appearance of a group of sharp peaks in the centre of the spectrum (Figure 4.1). The observed reduction in signal intensity across the spectrum was a good indicator of complex formation between the kinase domain and Cdc37, reflecting the incorporation of the kinase into an 82 kDa object which is largely NMR-invisible with the ^{15}N labelling strategy used in these experiments. This arises due to rapid relaxation of nuclear magnetisation as a consequence of slow tumbling of the complex in solution, and means that FGFR3-KD^{WT} in complex with Cdc37 is unfortunately largely undetectable in these spectra.

A very clear exception to this phenomenon was the appearance of 14 new complex-dependent peaks during titration with Cdc37. Directly opposing the global broadening of peaks across the spectra, many of these peaks were very intense, suggesting that the kinase residues they corresponded to do not experience the same relaxation phenomena of the complex as a whole and are more mobile. This feature and the central location of these peaks in the TROSY spectrum is consistent with probable localised unfolding of the kinase, consistent with prior findings from methyl-NMR experiments that Cdc37 partially unfolds the unrelated kinase sb-Raf^{V600E} at an undefined region, despite these two studies each using two different kinase domains (173). These data therefore further highlighted that the co-chaperone Cdc37 has roles not only in sorting through kinases to detect clients for Hsp90, but also seems to remodel these client kinases prior to their delivery to Hsp90.

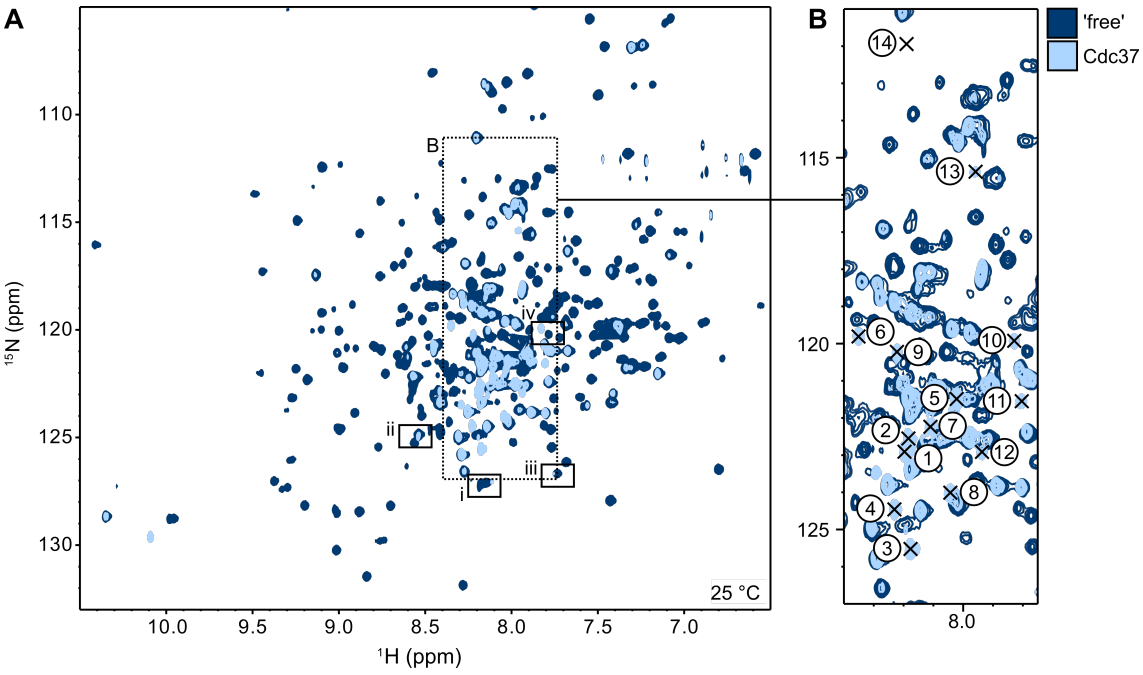


Figure 4.1. NMR analysis of Cdc37 binding to wild-type FGFR3 kinase.
(A) ^{1}H-^{15}N HSQC TROSY spectra of 'free' ^{15}N-labelled FGFR3-KD^{WT} (dark blue) and in the presence of 3 molar equivalents of unlabelled Cdc37 (light blue), each acquired at 950 MHz and at 25 °C for an identical number of scans. The boxed regions labelled i-iv correspond to those that are shown with expanded views in Figure 4.2. (B) An expansion of the overlaid spectra of panel A, highlighting the location of 14 new complex-dependent peaks located in the centre of the spectrum.

4.2.2. Attempts to define the interface between FGFR3 kinase and Cdc37

To understand how the binding of co-chaperone Cdc37 to kinases causes their remodelling, it is necessary to describe how these two proteins interact with one another. In protein NMR, binding interfaces between proteins (or between proteins and other objects) can be identified by tracking the movements of peaks in the presence of an interacting partner, and these interpreted by the mapping of these moving peaks onto the protein of interest (307). To extract the most information from this approach (called chemical shift perturbation (CSP) analysis), it is essential to have assigned NMR spectra. Fortunately, to allow interpretation of these Cdc37 binding data, the backbone amide assignments of free, *apo* FGFR3-KD^{WT} were available (~70% of non-Pro residues, kindly provided by Dr Domenico Sanfelice and Dr Gary Thompson, BMRB entry 27082) (342).

As noted previously, due to the weak binding affinity between FGFR3-KD^{WT} and Cdc37, it was expected that exchange between the free and bound states of the kinase would occur in the fast exchange timescale, enabling CSP analysis. However, upon visual inspection of spectra from the titration, it was clear that very few perturbations were observable (Figure 4.2A). Moreover, these perturbations did not show strict fast exchange behaviour, as they both moved in peak position and intensity simultaneously (Figure 4.2A). Of these visually-identified perturbed peaks, only one of these was assigned in the free state of the kinase, corresponding to residue T504 of strand β3 in the kinase N-lobe.

To explore whether there was a pattern of more subtle peak shifts across the protein, CSPs were calculated for each assigned peak of the kinase domain, using the position of peaks in the titration spectra closest to those that were assigned in the free kinase. This analysis found that the backbone amide peaks of the kinase rarely shifted at all, and rather simply decreased in intensity (Figure 4.2B). These findings were consistent with complex-induced global broadening of peaks, and implied that, generally, the positions of peaks used in this analysis likely corresponded to the remaining population of free, unbound kinase rather than the Cdc37-bound kinase; peak positions of the latter could likely not be defined owing to their poor signal to noise as a consequence of peak broadening. Notably, those CSPs that could be identified were all located close to the centre of the TROSY spectrum, presumably observable as they reported on regions of the kinase with fewer structural constraints and therefore faster effective correlation times and narrower linewidths. Similar to the new complex-dependent sharp peaks, these regions are likely more mobile than the complex as a whole.

4.2.3. Defining the Cdc37-induced local unfolding of FGFR3 kinase domain

While the binding interface of the binary complex could not be determined by CSP analysis, an understanding of the complex could nonetheless be obtained if those regions of the kinase that appear to be remodelled in the presence of Cdc37 were defined. Efforts were thus next focused on assigning the complex-dependent sharp peaks of the kinase that were induced by Cdc37 binding. As described in the relevant methods (see 2.11.5), these peaks were assigned in a corroborative manner through standard triple resonance approaches (Dr Domenico Sanfelice), prediction of residue chemical shifts in unstructured polypeptides (Dr Gary Thompson), and with selective unlabelling (313) of complexes to help resolve ambiguities in assignments from the triple resonance data (described here).

In these selective unlabelling experiments, five samples were prepared of the binary complex at a 1:1 molar ratio, each formed from non-labelled Cdc37 and ^{15}N-labelled FGFR3-KD^{WT} which had been selectively unlabelled at either Gln and Ile (-QI), Asn and Arg (-NR), Lys and Leu (-KL), Phe and Val (-FV) or Trp (-W) residues. This combinatorial unlabelling strategy has been reported previously (313). These samples were prepared by Dr Tom Bunney as detailed briefly in 2.11.1 and in more detail in (322) and kindly provided for NMR analysis. ^{1}H-^{15}N TROSY-HSQC spectra were recorded for each sample and compared to a spectrum acquired similarly for a complex containing uniformly ^{15}N-labelled kinase domain. To use these data to aid in peak assignment, it was first necessary to assess the efficiency of selective unlabelling achieved with each sample. This was accomplished by comparing the intensity of the relevant assigned peaks in spectra acquired of the uniformly and selectively-unlabelled samples (i.e. Lys and Leu peaks for the -KL sample), then determining whether those of the unlabelled sample showed intensity reductions significant enough to pass confidence thresholds that account for stochastic noise in the data (for details, see 2.11.5).

Analysing the unlabelling efficiency of each sample revealed that only the –KL sample saw effective unlabelling, with 63% of assigned Lys residues and 21% of Leu residues seeing peak intensity reductions that passed the calculated confidence thresholds (Figure 4.3A and B). All other samples showed very poor unlabelling with little to no residues passing the confidence thresholds, also reflected in their generally higher mean peak intensity ratios (Figure 4.3A). Even in the case of the –KL sample, a clear difference was observed between the unlabelling efficiencies observed between Lys and Leu residues, with the former showing approximately three-fold higher unlabelling. This difference is likely a consequence of a metabolic scrambling of the unlabelled Leu used as a supplement in the expression media into Ile and Val residues during recombinant expression of the kinase (313), as assigned peaks for each of these residues in this

spectrum also see a global reduction in intensity (Figure 4.3A). Lys residues, on the other hand, do not suffer from this metabolic scrambling effect and thus higher selective unlabelling efficiencies are observed (311). Instead the incomplete unlabelling of Lys residues in this kinase sample likely reflects an insufficient pool of unlabelled Lys being available during protein expression, and the subsequent incorporation of alternatively-available ^{15}N-labelled Leu that was synthesised by the *E. coli* culture. Due to insufficient selective unlabelling by all other kinase samples, only data of the –KL selectively unlabelled sample were used to help assign the new complex-dependent peaks, with greater emphasis placed on Lys unlabelling.

Analysis of the unlabelling efficiency observed for the Cdc37 complex-dependent peaks with the -KL sample indicated that peaks 1 to 7 experienced intensity reductions that passed the calculated confidence thresholds (Figure 4.3C). Of these, peaks 6 and 7 showed almost complete unlabelling, an effect when considering metabolic scrambling observed across the assigned residues is more consistent with these peaks corresponding to Lys residues rather than Leu residues (Figure 4.3B). Using standard triple resonance approaches (by Dr Domenico Sanfelice), peaks 2 to 6 were assigned as E458, L459, A461, D462 and K464 respectively (Figure 4.4A), thereby also identifying peak 6 as a Lys residue. The two approaches were found to be less consistent elsewhere, as while the selective unlabelling data suggested that peaks 1 to 5 should also be either Lys or Leu residues (or possibly Ile or Val due to metabolic scrambling), only at peak 2, assigned as L459, does this agree with the triple resonance data.

It is likely that imperfect data normalisation of the unlabelling spectra rationalises this discrepancy, as many of the complex-dependent peaks are more intense than the Gly residues that were used as reference peaks for intensity normalisation between samples. Note that Gly residues were used for normalisation of these data as they are unaffected by metabolic scrambling for all of the five combinations of selective unlabelling used in the study (311,313), and thus this methodology of intensity normalisation could be applied consistently among all five samples. The suggestion that imperfect normalisation may explain discrepancies between amino acid typing by selective unlabelling and triple resonance approaches is supported by the correlation between the absolute peak intensity of complex-dependent peaks, and whether these peaks passed the calculated confidence thresholds (see 2.11.5); the intensity of these peaks decreases from peak 1 through to peak 14 (Figure 4.4A), and only the first half of these peaks are deemed to pass the defined threshold (Figure 4.3C). Moreover, complex-dependent peaks 1 to 5 are all more intense than at least one of the Gly reference peaks (G481, G553, G549 and G574). As both peaks 6 and 7 show almost complete intensity reduction, they likely correspond to Lys residues experiencing true unlabelling, as described above.

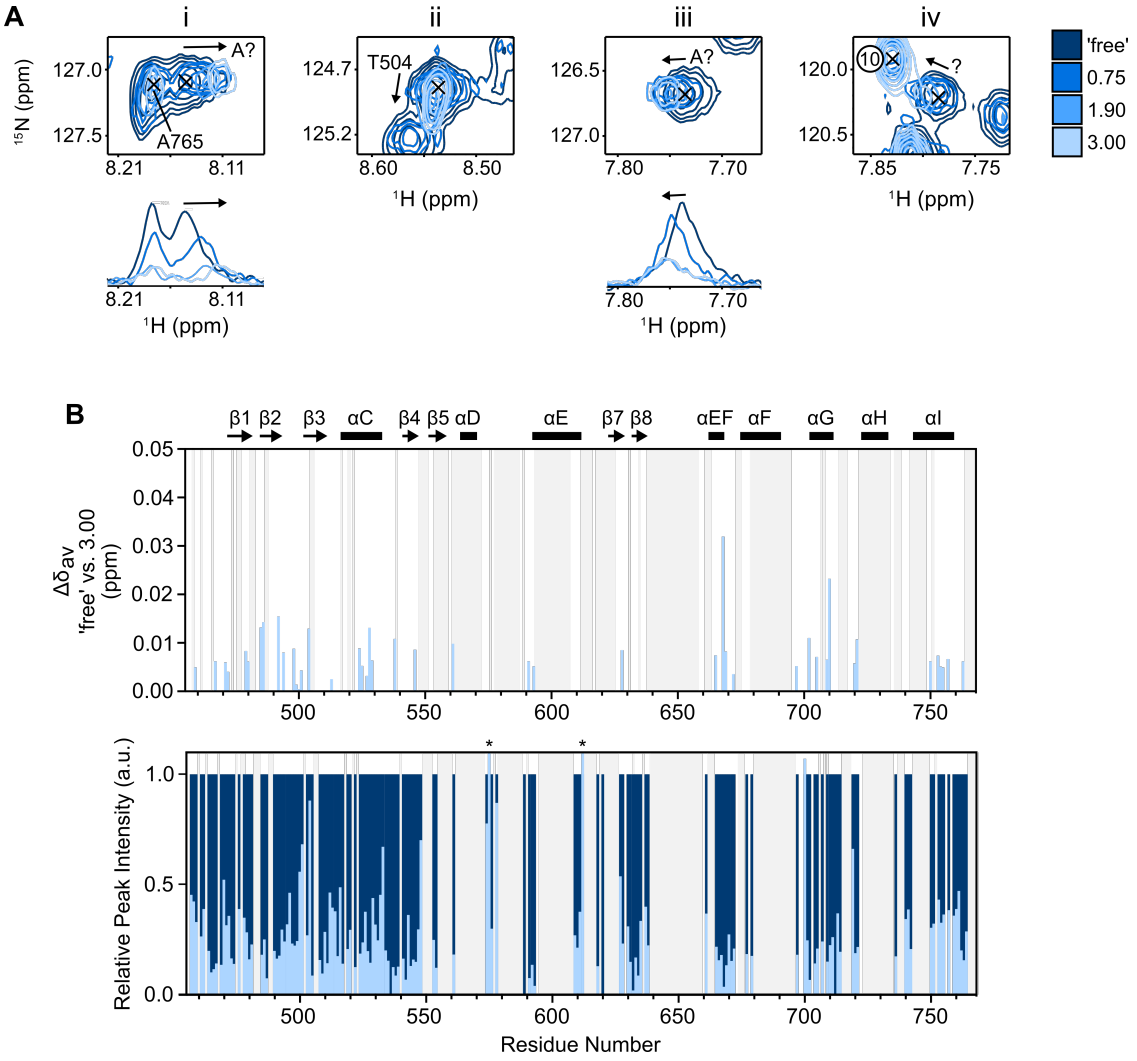


Figure 4.2. Chemical shift perturbation and peak intensity analysis of Cdc37 binding to wild-type FGFR3 kinase.
(A) Expanded regions of ^{1}H-^{15}N spectra of ^{15}N-labelled FGFR3-KD^{WT} titrated with various molar amounts (0.75, 1.90 and 3.00) of non-labelled Cdc37, showing visual movement of peaks. Below two of these, 1D ^{1}H spectra of the peaks are shown, illustrating concurrent peak movement and intensity changes. Where known, the assignments of the peaks concerned are given. In three instances, peaks with observed perturbations in chemical shift are not assigned (labelled with '?'). (B) Plot of observed chemical shift perturbations ($\Delta\delta_{av}$, calculated as in 2.11.4) (top) and the relative peak intensities (bottom), comparing FGFR3-KD^{WT} alone and with a 3-molar excess of Cdc37 (coloured as in A). In both, grey bars indicate unassigned regions of the kinase domain. The secondary structure of FGFR3-KD is given above the charts. In the bottom chart, asterisks indicate relative peak intensity values of greater than 1. Given the magnitude of the average perturbation in chemical shift ($\Delta\delta_{av}$) and the generally uniform reduction in intensity of these peaks, all perturbations were deemed insignificant and likely reflect minor inaccuracies in peak picking across the spectra, as discussed in the main text. For examples of significant peak perturbations, see Figure 4.7 and Figure 4.10.

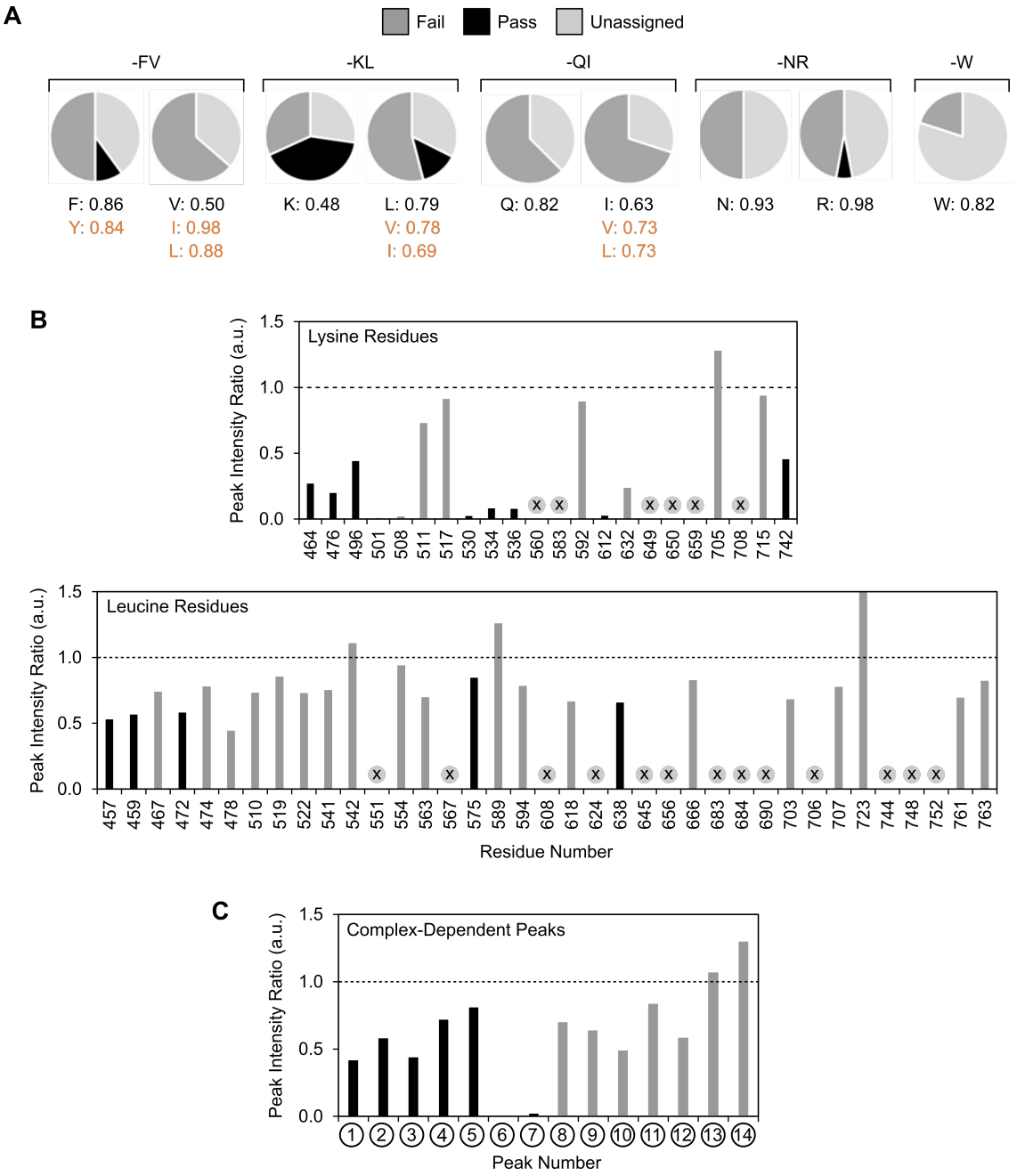


Figure 4.3. Selective unlabelling of wild-type FGFR3 kinase.
(A) Pie charts summarising the efficiency of selective unlabelling with -FV, -KL, -QI, -NR and -W samples of FGFR3-KD^{WT}, assessed through comparison the intensity of assigned peaks in spectra of a uniformly ^{15}N-labelled kinase and the relevant selectively unlabelled kinase, as detailed in 2.11.5. Light grey, dark grey and black indicate residues that are unassigned (and thus not used in the estimation of unlabelling efficiency), and those that failed or passed the calculated confidence thresholds (see 2.11.5). Below each chart, the mean peak intensity ratio of target residues and of metabolic scramblers (as defined in (313)) are shown in black and orange, respectively. (B) Example analysis of unlabelling efficiency, shown for all Lysine and Leucine residues in the kinase domain of the –KL selectively-unlabelled FGFR3-KD^{WT} sample. Bars are coloured as in panel A, reporting on whether peak intensity reductions in the unlabelled sample passed the calculated confidence thresholds. Unassigned residues in the kinase backbone are highlighted by crosses. (C) Analysis of –KL selective unlabelling of the Cdc37 complex-dependent peaks which were used to aid in peak assignment. Peaks are labelled as in Figure 4.1.

Despite this phenomenon, analysis of the intensities of the complex-dependent peaks and the unlabelling data could be used to tentatively assign a final peak that was not assigned using triple resonance approaches. As there is a clear correlation between peak intensity and the assignments of peaks 2 to 6 whose residues all lie sequentially in the kinase's amino acid sequence, peak 1 was identified to likely correspond to L457 (Figure 4.4A). The unlabelling efficiency of this peak is consistent with that of peak 3 (assigned as L459), providing additional, though speculative, support that it corresponds to a Leu residue.

Together, these assignments define the remodelling effect on kinases by Cdc37 as an apparent induction of localised unfolding of an 8-residue stretch (including interspersed Pro residues) at the immediate N-terminus of FGFR3-KD^{WT} (Figure 4.4B). When mapping this region of localised unfolding on the crystal structure of FGFR3 kinase domain, this polypeptide is found to lie atop the αC helix in the N-lobe, making a number of polar contacts with residues of αC, the β4-β5 loop, and residue E466 (Figure 4.4B). While only the N-terminus of the kinase domain is observed to unfold in these data, these very limited changes probably report on more extensive structural changes occurring elsewhere in the kinase, likely in the N-lobe, but which cannot be detected due to the binary complex being largely unobservable under these experimental conditions, as discussed previously. This is supported by the observation by NMR that residues D462 and K464 are unfolded in the presence of Cdc37 binding, yet these residues are involved in a hydrogen bonding network between the side chains of E527 and K530 of the αC helix and the backbone of E466 in the free kinase (Figure 4.4B). Likewise, though this region has higher B factors in the crystal structure than the rest of the kinase, the backbone carbonyl of L459 partakes in a hydrogen bonding network extending to the side chain of E466 (Figure 4.4D). It is plausible that rearrangements elsewhere in the kinase might destabilise these interactions, allowing this short polypeptide to locally unfold in response to the binding of Cdc37. It is notable that E446, a residue that appears to be central to this network of interactions, is the site of a substitution (E466K) that enhances the client status of FGFR3-KD for Hsp90 (322).

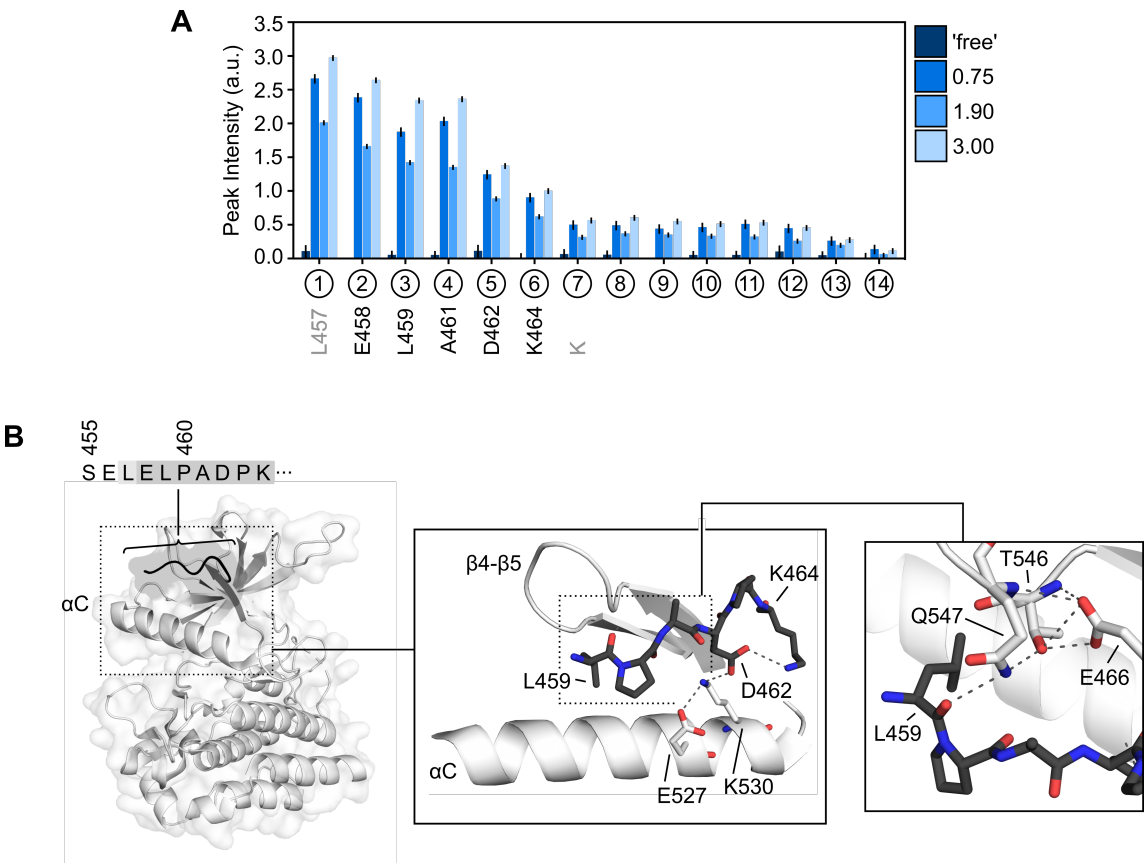


Figure 4.4. Defining Cdc37-induced local apparent unfolding of FGFR3 kinase. (A) Measured peak intensities of the complex-dependent peaks from TROSY-HSQC titration spectra of wild-type FGFR3 kinase domain with Cdc37, with peak assignments identified by triple resonance and selective unlabelling approaches given below each peak number. Error bars are given on each peak corresponding to the noise measured in each spectrum to illustrate uncertainty in peak intensity estimation. Peaks are numbered as in Figure 4.1 and bars are coloured according to the molar equivalents of non-labelled Cdc37 (0.75, 1.90, and 3.00) present at each titration point, as in the key. The kinase concentration was kept constant between samples. Assignments given in grey correspond to those which have not been validated from standard triple resonance experiments. (B) Mapping of the assigned complex-dependent peaks from panel A (dark grey) on the crystal structure of FGFR3 kinase domain (PDB entry 4K33). Panels show the interaction network of these residues with others in the αC helix and residues of the N-lobe of the kinase.

4.2.4. Probing the formation of a binary complex with a stronger client

To demonstrate whether the apparent structural remodelling of kinases is a specific effect of Cdc37 that is conserved among client kinases of different strengths, Cdc37 titrations were next repeated using a variant of FGFR3 kinase domain harbouring the I538F substitution (FGFR3-KD^{I538F}). This substitution, observed in hematopoietic cancer (84), is located within the DFG-latch of the kinase and an allosteric network of residues at the kinase N-lobe and C-lobe interface encompassing the 'molecular brake' and activation loop, both which play roles in kinase activity (65). While the pathological benefit of this mutation is currently unclear (unlike mutations in these elements such as N540K at the molecular brake, I538F slightly reduces kinase activity (322)), its introduction into the isolated kinase domain of FGFR3 reduces its global thermal stability by ~10 °C (322). Consistent with previous reports on kinases more generally (141), this reduction in thermal stability correlates with an increased affinity for Cdc37 (with an estimated K_d of 1.8 μM compared to a K_d of greater than millimolar for FGFR3-KD^{WT}) and thus it being a stronger client of Hsp90 (322).

Titrations of FGFR3-KD^{I538F} with Cdc37 revealed the same effects as observed with wild-type FGFR3-KD that indicated binary complex formation: an extensive broadening of peaks, and the appearance of new, complex-dependent peaks located at the same or similar positions as those observed in TROSY spectra of Cdc37-bound FGFR3-KD^{WT} (Figure 4.5). These suggested that Cdc37 does indeed induce the same effects in both versions of the kinase, and that the remodelling of kinases by Cdc37 is conserved among clients of different strengths. Curiously, while peaks corresponding to localised unfolding of the N-terminus were only present in the wild-type kinase spectra when Cdc37 was present, three of these peaks were already present in spectra of *apo* FGFR3-KD^{I538F}, prior to Cdc37 addition (Figure 4.6A and B). This implies that free FGFR3-KD^{I538F} samples a partially unfolded state, a feature consistent with its lower thermal stability, possibly also indicating that the kinase is metastable. Upon binding, Cdc37 enhances the sampling of this locally-unfolded state (Figure 4.6).

As more structurally-malleable kinases likely require less energy input to be remodelled, the detection of 'metastability' could describe how Cdc37 discriminates between client and non-client kinases. Presuming that the interaction between Cdc37 and kinases is dominated by residues and motifs that are (semi-)conserved among kinases, such that they can be detected by the N-terminal end of Cdc37 (173), speculatively, the energy input required to remodel kinases may drive the affinity differences observed in client binding. A proper understanding of the details of client kinase sorting will remain unresolved until high-resolution structures describing the binding interface between Cdc37 and kinases of different strengths are determined.

As the interaction between FGFR3-KD^{I538F} and Cdc37 is of low micromolar affinity, it was expected that exchange between the bound and unbound forms of the kinase would occur in the slow exchange regime in NMR. For an interaction of 1:1 stoichiometry in this regime (as is the case for FGFR3 kinase and Cdc37 (322)), one would expect to observe two states in spectra, with the population distribution between these states being reflected in the intensities of peaks corresponding to each. As the 82 kDa binary complex of kinase and Cdc37 is largely invisible under these experimental conditions, and thus bound and unbound states of the kinase are (generally) not simultaneously observable, the intensity of peaks corresponding to the unbound fraction of kinase during the titration should report on the population distribution in the sample and (as the concentration of the kinase was in excess of the affinity of the interaction in these experiments) directly reflect the ratio of kinase and co-chaperone present in the system. Therefore, when Cdc37 was present at half the concentration of FGFR3-KD^{I538F}, at least half of the kinase was expected to be observable in the unbound state in the spectra. Surprisingly, this was not found to be the case, as in spectra of complexes with both 0.25 and 0.50 molar amounts of Cdc37 with respect to FGFR3-KD^{I538F}, a lower fraction of unbound kinase was observed than expected (Figure 4.6C). Moreover, in cases where both a bound and unbound state were observable (i.e. for residues of the region observed to apparently locally unfold in the presence of Cdc37), changes in occupancy were unequal here also (Figure 4.6D). These suggest that despite its low micromolar affinity, the interaction between kinase and Cdc37 occurs in the intermediate exchange regime (i.e. on the µs-ms timescale) and that it is not a simple on-off binding interaction but may involve conformational changes, possibly even implying that more than one free or bound states of the kinase are in exchange. In these data, peak broadening in the presence of the co-chaperone is thus likely caused both by the formation of an NMR-invisible 82 kDa complex (under these experimental conditions) but also as a consequence of exchange between free and bound forms of the kinase.

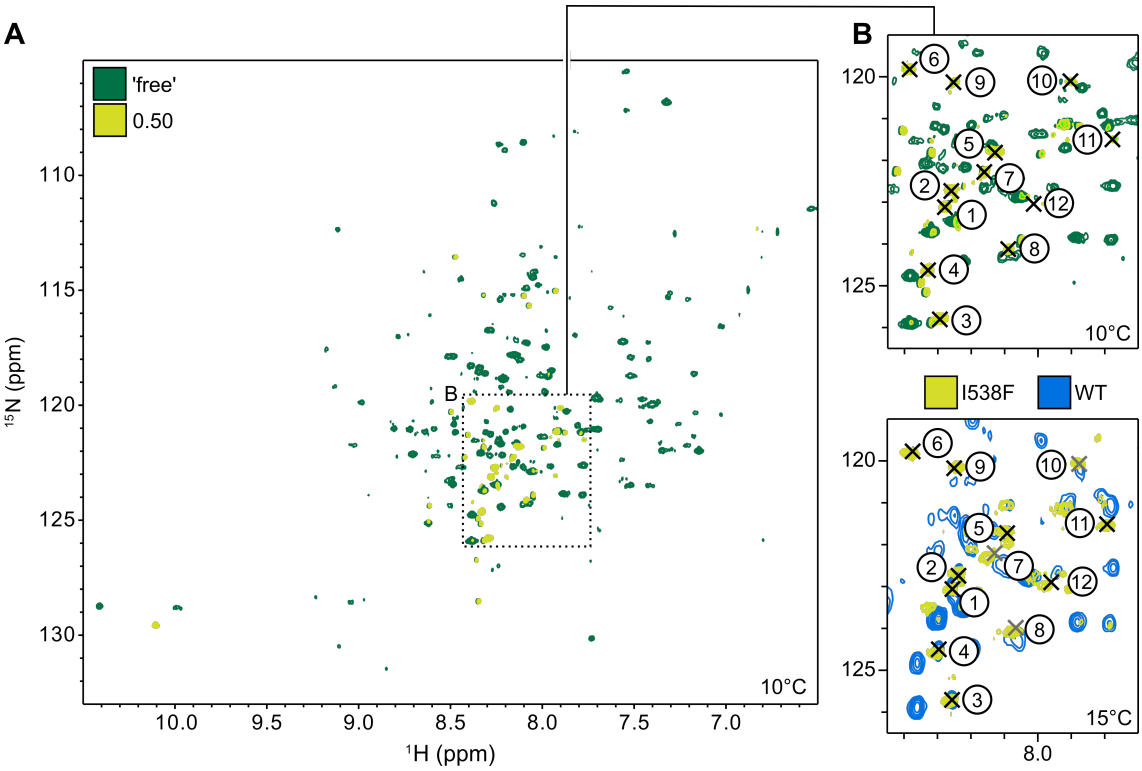


Figure 4.5. NMR analysis of Cdc37 binding to the I538F variant of FGFR3 kinase domain.
(A) ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-labelled FGFR3-KD^{I538F} alone (dark green) and in the presence of a 0.50 molar equivalent of non-labelled Cdc37 (light green), each acquired at 950 MHz and at 10 °C for an identical number of scans. (B) Expansion of panel A showing the appearance of 12 complex-dependent peaks (top) which are present in similar locations to those observed with wild-type FGFR3-KD (bottom). Crosses in the top and bottom panels indicate the location of these complex-dependent peaks in spectra of the I538F variant of the kinase domain and of wild-type FGFR3 kinase domain, respectively. To compare wild-type and I538F variants of the kinase domain in the presence of Cdc37, spectra were acquired of the former with 1 molar equivalent of unlabelled Cdc37 and the latter with 0.5 molar equivalents of unlabelled Cdc37, both at 950 MHz and at 15 °C. Differences in the amount of molar equivalents of Cdc37 were used to reflect the differences in affinity of each of these kinase domains for the co-chaperone.

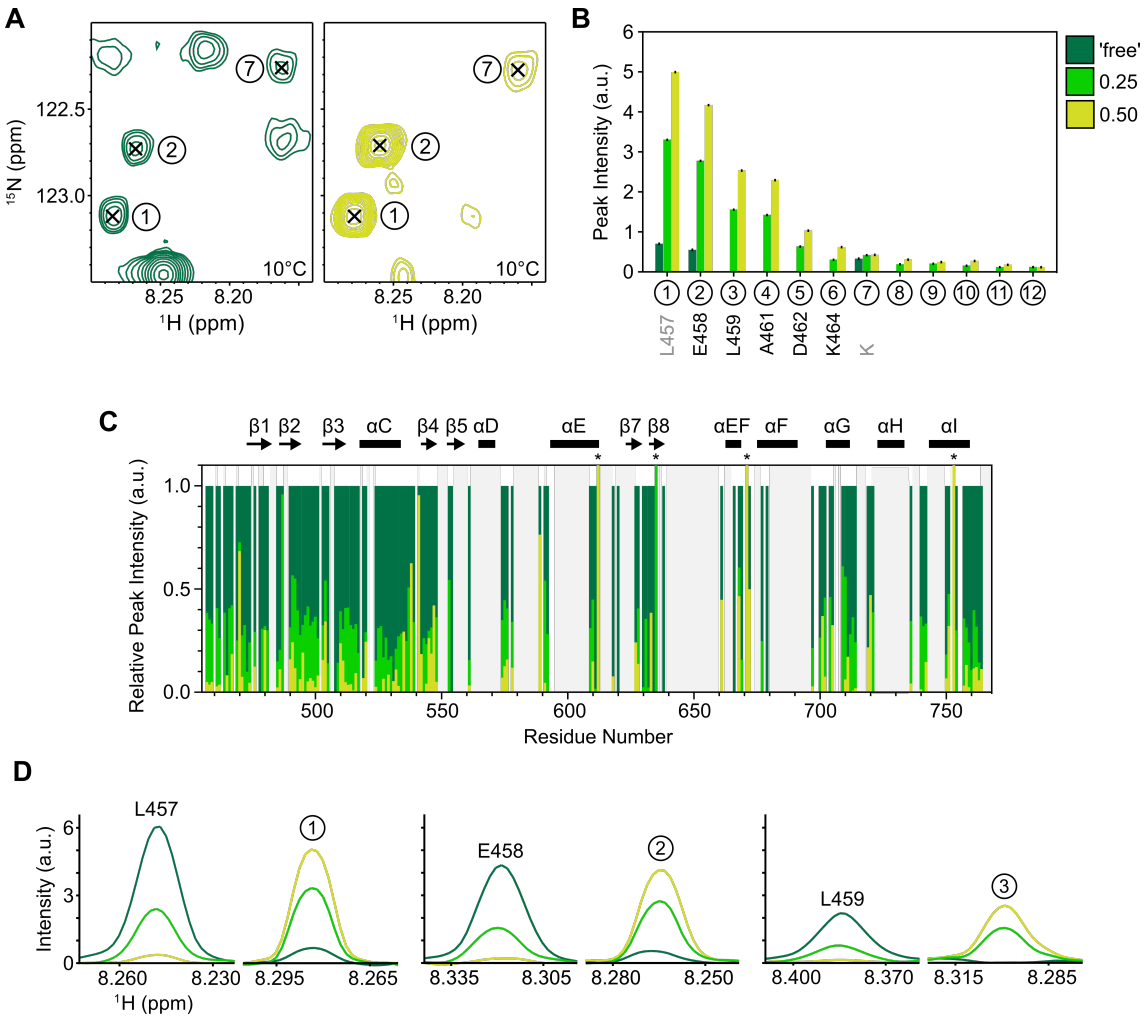


Figure 4.6. Peak intensity analysis of Cdc37 binding to the I538F variant of FGFR3 kinase domain.

(A) An expanded region of ^{1}H-^{15}N TROSY spectra of FGFR3-KD^{I538F} alone and in the presence of a 0.50 molar ratio of Cdc37 (as in Figure 4.5), illustrating the presence of 'complex-dependent' peaks in free FGFR3-KD^{I538F} prior to Cdc37 titration. (B) Measured peak intensities of complex-dependent peaks from the titration series of FGFR3-KD^{I538F} with Cdc37, with the peak assignments as determined with wild-type FGFR3 kinase domain shown below, as in Figure 4.4. Error bars are given on each peak corresponding to the noise measured in each spectrum to illustrate uncertainty in peak intensity estimation. Peaks are numbered as in Figure 4.1 and Figure 4.4, and bars are coloured according to the molar equivalents of non-labelled Cdc37 (0.25 and 0.50) present at each titration point, as in the key. (C) Relative peak intensities of all assigned residues of FGFR3-KD across the titration series, coloured as in panel B, illustrating that reductions in peak intensities do not reflect the fraction of kinase domain that is expected to become invisible (under these experimental conditions as discussed in the main text) given the molar equivalents of Cdc37 present in the sample. Grey bars indicate those regions which are unassigned, and asterisks indicate values that exceed 1. The secondary structure of FGFR3 kinase domain is given above the charts. (D) 1D ^{1}H spectra of three of the 'complex-dependent' peaks (peaks 1 to 3) and their corresponding peaks in the 'free' unbound kinase state across the titration series. Both panels C and D illustrate that peak intensities do not change in a linear, ratio-dependent manner, as would be expected in the slow-exchange regime.

4.2.5. Clues to the structural basis of client kinase strength

In the previous section, the malleability of kinases to Cdc37 was proposed to be the main underlying factor that contributes to the client strength and status of each kinase. However, the structural features underlying this 'malleability' remain unclear. As wild-type and I538F variants of FGFR3 kinase domain correspond to two client kinases of different strengths, a comparison of the HSQC spectral fingerprints of each variant could highlight structural differences introduced by the I538F substitution that transform FGFR3 into a better client for Hsp90, revealing the features which dictate the client strength of kinases on the client-to-nonclient continuum.

CSP analysis of the free, unbound forms of FGFR3-KD^{WT} and FGFR3-KD^{I538F} showed that structural differences introduced by this substitution were mostly observed in and around the mutation site as would be expected, but that it also induced changes (though less strongly) at more distant sites in the kinase N-lobe up to 20 Å away (Figure 4.7). Interestingly, a group of these perturbations (residues 519, 522, and 525-529) stretch along the αC helix of the N-lobe and interact with the N-terminal region of the kinase that is observed to be unfolded in the presence of Cdc37 (Figure 4.4 and Figure 4.7). Residues that are perturbed by the I538F mutations also include N540 of the molecular brake, the site of a further mutation (namely N540K) that also reduces kinase thermal stability and enhances the client status of FGFR3 kinase domain (322). While a large portion of the C-lobe is unassigned in these spectra, it is noted that many perturbations induced by I538F occur in the kinase N-lobe, consistent with prior findings that features of the N-lobe are responsible for differences in client strength (164).

This observation of a stretch of perturbations in the kinase N-lobe may highlight that the interactions of a residue network in the N-lobe underlies the origin of kinase 'clientness', with the strength of interactions in this network influencing the global thermal stability of the kinase and the energy input required by Cdc37 to induce its remodelling to dictate client strength. Perturbation of this network, either through its strengthening or weakening, could, by extension, alter the energy required for kinase remodelling and shift a kinase across the client-to-nonclient continuum. This proposal is consistent with prior findings that no particular structural motif is responsible for enhanced binding of Cdc37 (141). However, as all eukaryotic kinases share the same overall fold, they likely all share similar networks of residues across the N-lobe that are subtly different and that affect their client status through allosteric means.

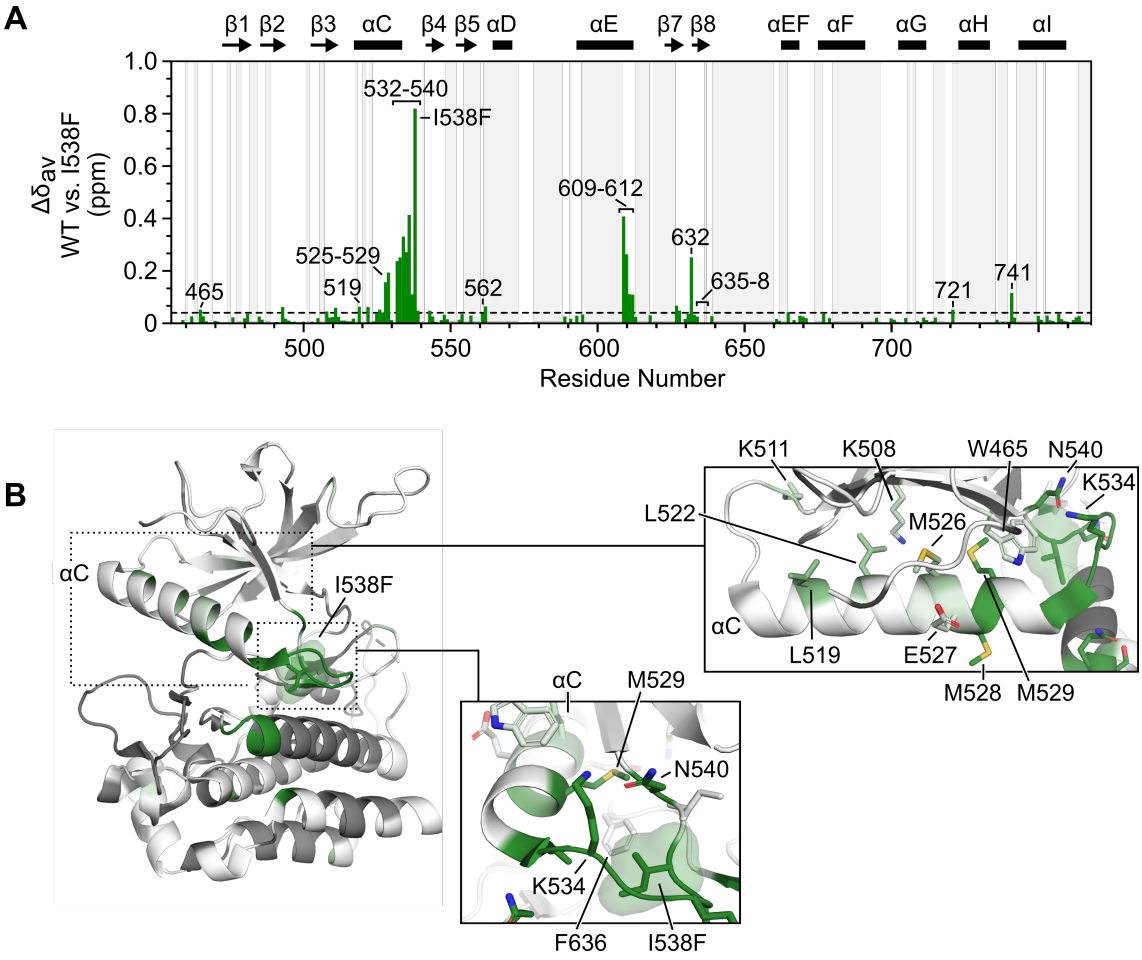


Figure 4.7. Chemical shift perturbation analysis of the I538F substitution in FGFR3 kinase domain.

(A) Plot of observed chemical shift perturbations ($\Delta\delta_{av}$, calculated as in 2.11.4) from comparison of TROSY-HSQC spectra of 'free' FGFR3-KD^{WT} and FGFR3-KD^{I538F}. Grey bars indicate unassigned residues (35% of non-Pro residues). The secondary structure of FGFR3-KD is given above the chart. A threshold corresponding to the mean $\Delta\delta_{av}$ across all assigned residues was used define those perturbations deemed as significant. (B) Mapping of chemical shift perturbations on FGFR3 kinase domain (PDB entry 4K33), shown with a gradient where darker green equates to larger changes. Unassigned residues are coloured grey. Panels show expanded views of the kinase illustrating a network of residues in the kinase N-lobe affected by the introduction of this mutation. Residues of interest are shown as sticks and the mutation site I538F highlighted with transparent surface representation.

4.2.6. The ATP-competitive inhibitor, PD173074, prevents Cdc37 interaction

Previously, it has been reported that ATP-competitive inhibitors of kinases can compete with Cdc37 *in vitro*, that they can partially release kinases from pre-formed kinase-Cdc37 complexes, and prevent access of kinases to the Cdc37-Hsp90 chaperone system in cells (175). The findings that Cdc37 itself also antagonises ATP binding to kinases (175), and that mutations in and around the ATP binding cleft of kinases are important for Cdc37 interaction (168,343–345), imply that the binding of ATP-competitive inhibitors at this same site could be responsible for their effect on Cdc37 binding. However, it is unclear whether competition between kinase inhibitors and Cdc37 is due to direct steric competition, due to kinase stabilisation leading to Cdc37 independence (given inhibitors are known to thermally stabilise kinases (346)), or a combination of the two.

To explore whether kinase inhibitors also inhibit Cdc37 binding to FGFR3 kinase, and whether these two modes of competition could be distinguished, the effect of the FGFR-selective inhibitor PD173074 on the formation of binary kinase-Cdc37 complexes was probed. Consistent with prior findings, PD173074 was found to successfully compete with Cdc37 for binding to FGFR3-KD^{WT}; when present in a 2-fold molar excess over the kinase, this was found to be the case both when the inhibitor PD173074 was added before or after incubation of the kinase with Cdc37 (Figure 4.8). To investigate the displacement of kinase from Cdc37-bound complexes in more detail, PD173074 was titrated into a pre-formed FGFR3-KD^{WT}-Cdc37 complex, and TROSY spectra acquired at each titration point. These showed a clear population shift between Cdc37-bound and PD173074-bound kinase during the titration, with an almost complete clearance of the binary co-chaperone-bound complex (Figure 4.10).

The more successful depletion of a pre-formed kinase-Cdc37 binary complex by PD173074 than previously reported for vemurafenib and sb-Raf^{V600E} (175) likely originates from differences in the interaction affinity between each kinase and Cdc37. While PD173074 and vemurafenib inhibit FGFR3 and sb-Raf^{V600E} with similar effectiveness (measured *in vitro* IC_{50} values of 29 nM (105) and 31 nM (347)), sb-Raf^{V600E} is a much stronger client of Hsp90 than FGFR3-KD^{WT}, binding to Cdc37 with a reported K_d of 0.2 μM (175) compared to a K_d of greater than millimolar for FGFR3. As FGFR3-KD^{WT} was used at only 130 μM in these experiments, the occupancy of the binary complex is expected to be low in this pre-formed complex sample (estimated to be below 10%), leaving the majority of the kinase available for inhibitor binding. This suggests that the depletion of kinase-Cdc37 binary complexes by inhibitors may occur through inhibitor-mediated consumption of the pool of free kinase (a consequence of these inhibitors showing higher binding affinities than Cdc37), and thus their effects on the kinase-Cdc37 binding equilibrium that result in kinase dissociation from Cdc37.

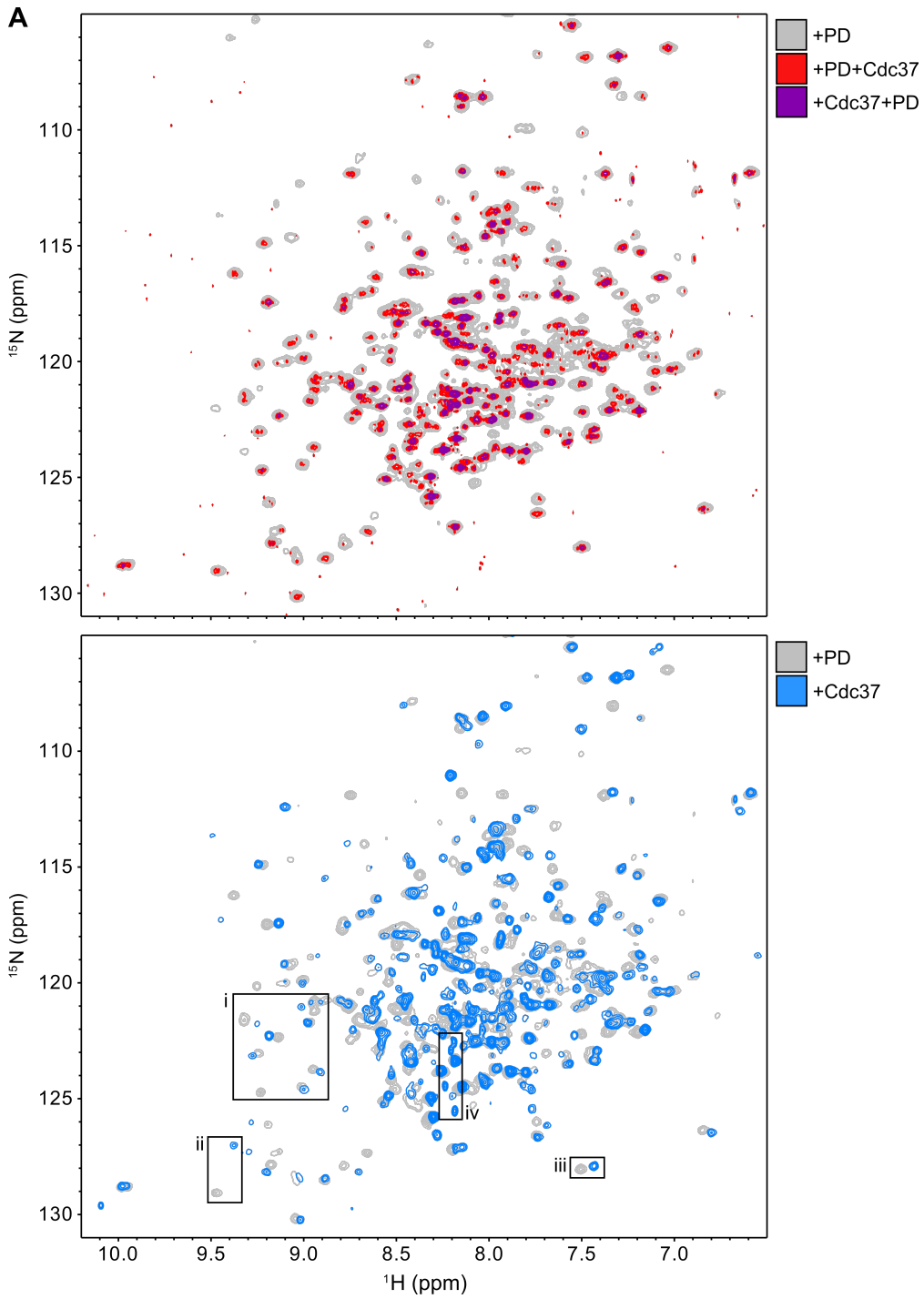


Figure 4.8. The effect of the FGFR inhibitor PD173074 on kinase-Cdc37 binary complex formation.
(A) ^{1}H-^{15}N TROSY-HSQC spectra of PD173074 (PD)-bound FGFR3-KD^{WT}, and of FGFR3-KD^{WT} incubated with PD and Cdc37 in opposing orders (+PD then +Cdc37, or +Cdc37 then +PD). In all cases, the same PD-bound state of the kinase is observed, indicating that the kinase inhibitor PD173074 outcompetes Cdc37 no matter the order of component incubation. (B) For comparison, ^{1}H-^{15}N TROSY-HSQC spectra of FGFR3-KD^{WT} in the presence of PD or Cdc37, illustrating differences between the two bound forms. Panels i-iv correspond to those shown as expanded views in Figure 4.9.

While it is undeniable that the inhibitor PD173074 is able to outcompete Cdc37 for binding to wild-type FGFR3 kinase domain, this data is unable to define the mechanism of competition. Crucially, no evidence of simultaneous binding of both PD173074 and Cdc37 is observed in the data. However, as the kinase-Cdc37 binary complex itself is largely undetectable under these experimental conditions (as discussed previously), an inhibitor-bound tertiary complex would likely also be undetectable. Thus, such a complex containing kinase, Cdc37 and kinase inhibitor cannot confidently be ruled out, and it remains unclear whether this competition mode is based purely on stabilising kinases when they are unbound, or whether the inhibitor is able to interact directly with kinase-Cdc37 complexes and induce co-chaperone dissociation. Insights could be gained here in the future through NMR experiments using ^{13}C methyl-TROSY analyses that allow systems much larger in molecular mass to be studied (310).

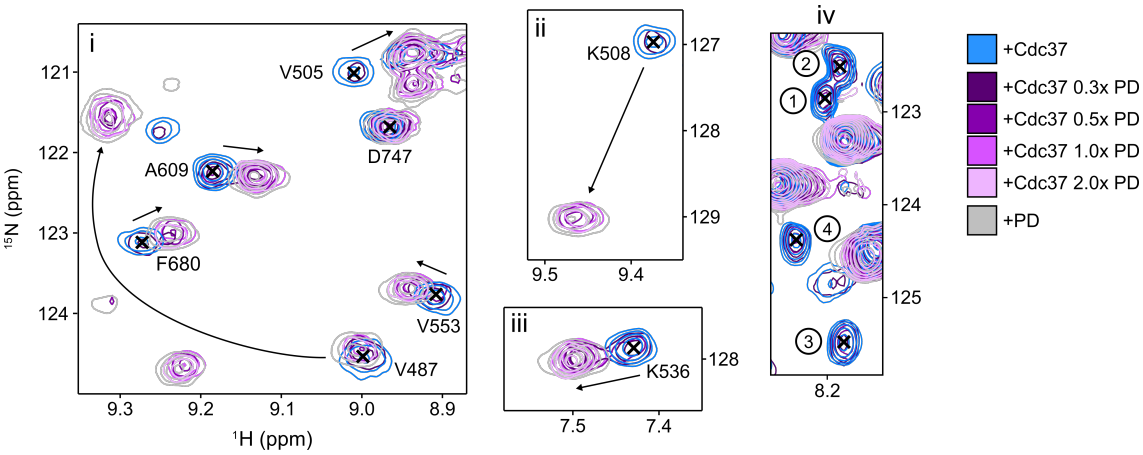


Figure 4.9. PD173074 successfully outcompetes Cdc37 binding to FGFR3 kinase domain.
To further demonstrate that the kinase inhibitor PD173074 outcompetes Cdc37 for interaction with wild-type FGFR3 kinase domain, ^{1}H-^{15}N spectra were acquired of titration of PD173074 into a pre-formed Cdc37-kinase complex (as detailed in 2.11.6). Spectra were acquired of the kinase-Cdc37 complex alone, and of PD173074 added at various molar equivalents (0.3, 0.5, 1.0 and 2.0) with respect to the kinase concentration. A PD-bound spectrum of FGFR3-KD^{WT} (2.0 molar equivalents of PD173074) is provided for comparison to illustrate the shift from a Cdc37-bound to PD-bound state. The panels shown correspond to those regions highlighted in Figure 4.8.

As inhibitor binding improves the thermal stability of kinases, analysing the effects of their binding on kinase structure could provide further insights into features that define client strength, building upon observations from the effect of I538F substitution in FGFR3 kinase domain. To explore this possibility, structural and conformational differences between free and PD173074-bound wild-type FGFR3 kinase domain were determined by CSP analysis, using backbone resonance assignments of each state kindly provided by Dr Hans Koss, Dr Domenico Sanfelice and Dr Gary Thompson (BMRB entries 27082 and 27083) (326). As expected, this analysis identified major perturbations at residues about the nucleotide binding cleft between the kinase lobes, identifying an inhibitor binding site consistent with that which is observed in the crystal structure of PD173074 bound to FGFR1 kinase (Figure 4.10). These perturbations include those at the P-loop and DFG motif, but also those within the αC helix and elsewhere in the N-lobe. In contrast, where assigned, perturbations observed across the C-lobe were largely weak and insignificant (Figure 4.10).

PD173074 improves the global thermal stability of FGFR3-KD^{WT} by ~15°C (322). As there are generally no significant CSPs observed in the kinase C-lobe, these data suggest that these contributions to kinase stability derive from stabilisation of the kinase N-lobe and the interface between the N- and C-lobes alone. Interestingly, there is considerable overlap between regions observed to be perturbed by inhibitor binding and I538F substitution in FGFR3 (Figure 4.7 and Figure 4.10), with perturbations being observed in both instances at residues 508, 519, 522, 528-9, 532-3, 535, 539-40 and 562, 632 and 678. Many of these cluster at the C-terminal end of the αC helix and the αC-β4 loop, the location of the I583F and N540K substitutions that increase the client strength of FGFR3 kinase domain (322). This supports the premise that a network of interactions in the kinase N-lobe may be the underlying origin of client kinase strength. The strengthening of this network may (at least partially) explain how Cdc37 independence can arise in kinases as a consequence of inhibitor binding.

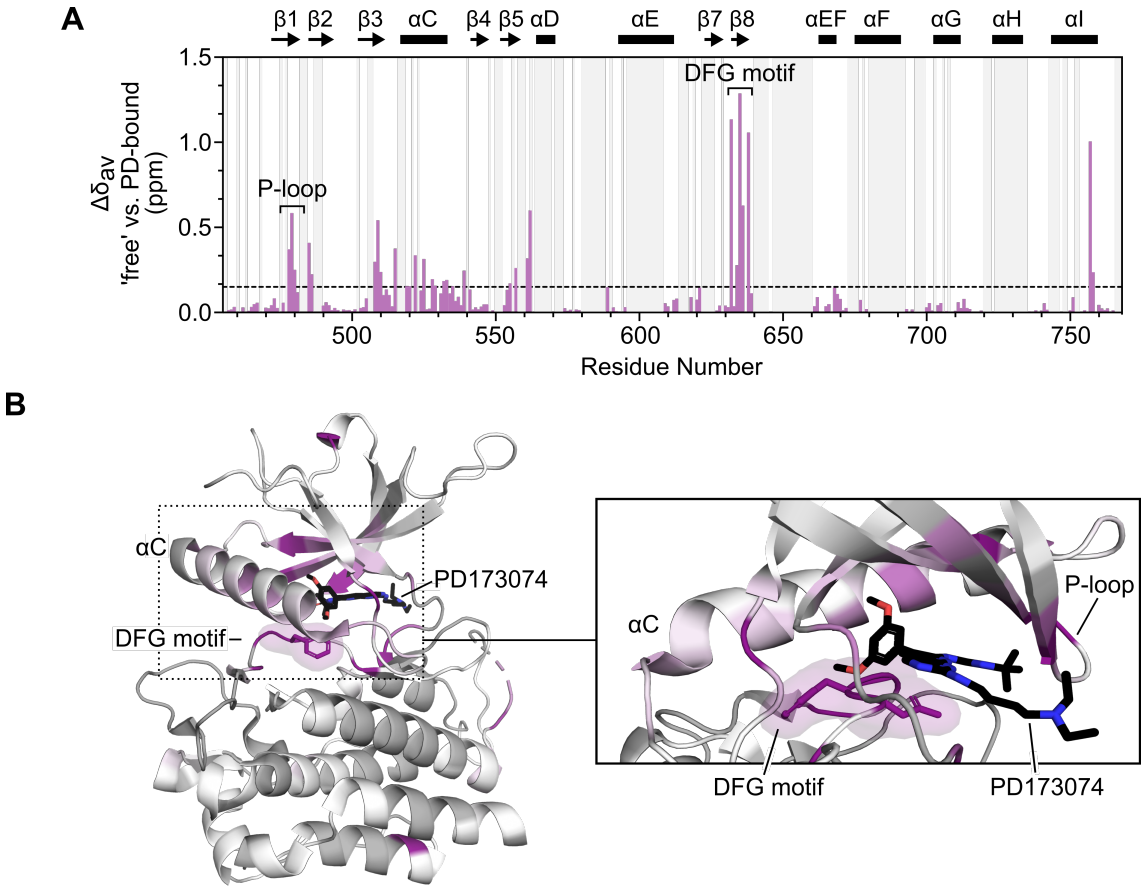


Figure 4.10. Chemical shift perturbation analysis of PD173074 binding to FGFR3 kinase domain.
(A) Plot of chemical shift perturbations ($\Delta\delta_{av}$, calculated as in 2.11.4) from comparison of TROSY-HSQC spectra of FGFR3-KD^{WT} alone and bound to PD173074 (PD). Grey bars indicate unassigned residues either in free FGFR3-KD^{WT} or in PD-bound FGFR3-KD^{WT} (37% of non-Pro residues). The secondary structure of FGFR3-KD is given above the chart. A threshold corresponding to the mean $\Delta\delta_{av}$ across all assigned residues was used define those perturbations deemed as significant. (B) Mapping of chemical shift perturbations on to FGFR3 kinase domain (PDB entry 4K33) with a gradient where dark purple corresponds to the greatest changes. Unassigned residues are coloured grey. The binding site of PD173074 (black sticks) in the cleft between the kinase N- and C-lobes is shown, generated from structural alignment with PD-bound FGFR1 kinase domain (PDB entry 2FGI). The DFG motif, also shown as sticks, is highlighted with a transparent surface.

4.3. Discussion

In this chapter, the recruitment of kinases to Hsp90 for chaperoning was explored. As reported previously, this is mediated by the co-chaperone Cdc37 that both recognises clients of Hsp90 from the kinase pool and remodels them prior to their transfer to Hsp90, including a partial unfolding of the kinase (173). However, the exact regions of kinases that are remodelled by Cdc37 were poorly defined, a consequence of lacking resonance assignments in these data and thus relying on limited proteolysis approaches to identify regions of the kinase that had become more accessible to proteases in the presence of the co-chaperone (173). Here, NMR spectroscopy was used once more to probe the interaction between kinases and Cdc37 in attempt to better characterise this binary complex interaction. The kinase domain of the receptor tyrosine kinase FGFR3 was used in this instance, a departure from model clients typically used in probing Cdc37 interactions.

Through use of two variants of FGFR3 kinase domain that differ in client strength (the weak client of FGFR3-KD^{WT}, and the stronger client of FGFR3-KD^{I538F}), these NMR data show that the remodelling induced by Cdc37 is conserved among different strength clients and results in the conserved apparent partial local unfolding of the kinase (Figure 4.1 and Figure 4.5). Through a collaborative effort, this unfolding event was identified to occur at the immediate N-terminus of the kinase (Figure 4.4). Given that these residues take part in a number of hydrogen bonding interactions in the crystal structure of FGFR3 kinase domain (PDB entry 4K33), it is proposed that the NMR-observed unfolding of the kinase at its N-terminus likely reports on additional changes occurring elsewhere in the kinase, likely in the kinase N-lobe. However, due to the relaxation-based peak broadening effects observed in these NMR experiments as a consequence of the 82 kDa molecular mass of the binary complex and the use of ^{15}N labelling, these more extensive structural changes were not observable. The possible existence of multiple bound and unbound states of FGFR3-KD^{I538F} in the presence of Cdc37, implied from exchange of this interaction occurring in the intermediate regime on the NMR timescale despite the low 1.8 μM affinity of this interaction, is consistent with the premise that Cdc37 binding to kinases is not a simple on-off mechanism. Speculatively, this could suggest that Cdc37 causes more extensive structural changes elsewhere in the kinase. These events could possibly be observable in future experiments if ^{13}C methyl-TROSY based analyses are used (310).

It should be noted that in these studies, FGFR3 kinase domain was studied in isolation of the rest of the receptor. In context of this full-length receptor, the region observed to apparently unfold in the presence of Cdc37 immediately follows the juxtamembrane domain of the receptor that connects the transmembrane domain and kinase domain. It

is important therefore to consider whether this apparent unfolding event observed with the kinase domain alone remains relevant in context of its location in the authentic receptor. As discussed previously (see 4.2.3), in the sole crystal structure of FGFR3 kinase domain, while the region observed to be unfolded in the presence of Cdc37 exhibits higher B factors than the rest of the kinase as a whole, it is nonetheless observed in the structure, whereas residues preceding this in the crystallisation construct (namely P448-E458) are not (49). This implies that residues L459-K464 that are observed as unfolded in the presence of the co-chaperone by NMR, are indeed structurally restrained in the free kinase, and that this apparent unfolding event reflects a real change. In the future, NMR experiments using FGFR3 kinase domain constructs that also contain the juxtamembrane domain (and possibly the transmembrane region) could be carried out to confirm that the same effects are observed with larger constructs truer to the intact receptor.

Curiously, while peaks in HSQC TROSY spectra corresponding to this apparent unfolding event only appeared in the presence of Cdc37 for the wild-type kinase domain, in the case of the less thermally stable and stronger client kinase, FGFR3-KD^{I538F}, a number of these peaks were already present in the free kinase (Figure 4.6). This suggests that free FGFR3-KD^{I538F} samples a partially unfolded state, likely a reflection of its lower thermal stability (322). It has been suggested previously that Cdc37 recognises clients through detecting their global thermal stability (141). Studies have indicated that the kinase N-lobe alone drives client strength (164–169). However, as clients do not appear to contain a specific motif that Cdc37 recognises (while some such as a glycine rich loop at β3-αC (168,169) or electrostatics of the αC-β4 loop (166) have been proposed as important for Cdc37 interaction, their presence does not correlate with client status), it has instead been proposed that all kinases possess the features necessary for interaction with Hsp90, but that some (i.e. stronger client kinases) present these features more readily than others (non-clients) (165,170). By extension, discrimination may arise from how easily they can be remodelled, reflecting their thermal stability; however, the underlying structural characteristics in the kinase N-lobe that dictate this feature were unclear. Here, through CSP analysis of the two FGFR3 kinases of different client strengths, a network of perturbations within the N-lobe and particularly along the αC helix was identified connecting the mutation site, I538F, which strengthens client status, to the region observed to unfold in the presence of the co-chaperone (Figure 4.7). Building on earlier findings that ATP-competitive inhibitors can stabilise kinases and outcompete Cdc37 (175), CSP analysis of the FGFR-selective inhibitor PD173074 binding to FGFR3 kinase found that perturbations induced by the inhibitor overlapped with those induced by I538F substitution (Figure 4.10), further suggesting that interactions between residues of the N-lobe underlies client strength. In strong clients, the global strength of interactions

in this network is presumably weak and more easily structurally manipulated than in weaker clients. This is consistent with observations that mutations in the kinase N-lobe influence the strength of a kinase's client status for Hsp90 (e.g. E466K, I538F and N540K in FGFR3 kinase (322)), and that these mutations and kinase inhibitors that bind at the interface between the two kinase lobes, altering client strength and influencing the thermal stability of kinases, likely do so through subtly perturbing N-lobe structure.

Given that FGFR3-KD^{I538F} appears to be partially unfolded in the absence of Cdc37, it could be argued that the co-chaperone does not actively remodel and unfold kinases, but instead selectively binds to the unfolded state and stabilises it, increasing its population. Such a model has previously been argued for the molecular chaperone Hsp70 (348,349). Further experiments will be required to define which of these two modes of action Cdc37 uses as they cannot be separated by the current NMR data alone.

4.3.1. Building a model of Cdc37 binding to kinases

While NMR data presented here could define that kinases are partially and locally unfolded at their immediate N-terminus in the presence of Cdc37, CSP analysis of spectra acquired in the absence and presence of the Cdc37 were unable to define the interaction interface between the kinase and co-chaperone (Figure 4.2). This stemmed from the binary complex being largely invisible in these ^{1}H-^{15}N TROSY HSQC spectra as a consequence of relaxation-induced broadening due to the high molecular weight of the complex (~82 kDa). Additionally, as described above, this broadening effect meant that probable more extensive structural changes in the kinase that cause this local unfolding event at the kinase N-terminus were not observable.

Remaining questions regarding kinase-co-chaperone interactions can be partially addressed by complementary data available from collaborators. In the first of these datasets, hydrogen-deuterium exchange mass spectrometry (HDX-MS) was used to study the effects of Cdc37 on the exchange of hydrogen and deuterium at the backbone amides of two variants of FGFR3-KD (containing either I538F or E466K mutations) and the strong model client kinase sb-Raf^{V600E} (Roger Williams laboratory, MRC Laboratory of Molecular Biology, Cambridge). These studies found a considerable increase in hydrogen-deuterium exchange (HDX) across the N-lobe of each of the three kinases when in the presence of Cdc37, compared to when they were alone (Figure 4.11A). This finding suggests that Cdc37 remodels the entire kinase N-lobe, consistent with the more extensive remodelling of the kinase predicted from interpretation of the local unfolding observed in the NMR data. These findings are also consistent with more extensive (though undefined) changes in kinase structure observed in the study of sb-Raf^{V600E} interaction with Cdc37 by methyl-NMR (173).

In addition to the increased rate of HDX in the N-lobe when Cdc37 is present, protection from HDX was also observed in the C-lobe of all three kinases (Figure 4.11A). While the extent of HDX protection is not the same in each case, these findings are consistent with the C-lobe being the binding interface between kinase and co-chaperone. Though the sole assigned peak perturbation observed during CSP analysis of Cdc37 binding to wild-type FGFR3-KD was instead located in the kinase N-lobe (residue T504 of strand β3, Figure 4.2), this perturbation can be rationalised as arising from an allosteric effect of kinase remodelling rather than reporting on a binding interface. This suggestion fits well with the premise that structural changes are induced elsewhere in the kinase N-lobe in addition to the N-terminus alone. When analysing Cdc37 by HDX-MS in the same way, protection from HDX was found at the helices of C-Cdc37 (residues 302-341) and N-terminal coil of the co-chaperone (residues 5-28), suggesting that these correspond to the potential kinase-binding regions in this partner (Figure 4.11A).

In the absence of higher-resolution information, to yield insights into the structure of a kinase-co-chaperone complex, additional collaborators (Dmitri Svergun laboratory, EMBL Hamburg) used small angle X-ray scattering (SAXS) analysis to build a low-resolution model of Cdc37 bound to FGFR3-KD^{I538F} using *ab initio* and hybrid rigid body approaches, integrating the NMR findings described in this chapter and the HDX-MS data described above. The model that they generated suggests that the C-terminal helices of Cdc37 approach and bind to FGFR3 kinase at its C-lobe, while positioning the N-terminal coil of Cdc37 such that it can access the N- and C-lobe interface (Figure 4.11B). In addition to fulfilling NMR and HDX-MS data, this model is also consistent with prior published findings that have suggested that the N-terminal portion of Cdc37 (N-Cdc37) is responsible for detection of kinases, and that C-Cdc37-mediated interactions are necessary to stabilise kinase-co-chaperone binding (173). As NMR data presented in this thesis indicate that the same kinase N-lobe localised unfolding occurs with kinases of different client strengths, further supported by shared remodelling of the N-lobe in HDX-MS data of the unrelated kinases sb-Raf^{V600E} and FGFR3, this SAXS-derived model likely represents a model for Cdc37 binding that is true for all kinases, and provides the first structural snapshot of kinases on their way to Hsp90. Together, these data therefore suggest that in addition to recognising client kinases from the kinase pool, the co-chaperone Cdc37 also remodels and partially unfolds kinases at their N-lobe before they are transferred to Hsp90. This process leads to kinase inhibition, given the observation that FGFR3 kinases autophosphorylate themselves less efficiently in the presence of Cdc37 (Dr Tom Bunney, personal communication; (322)). Interestingly, a similar mechanism of client delivery to Hsp90 is observed for the glucocorticoid receptor (GR), where the chaperone Hsp70 partially unfolds GR before it interacts with Hsp90 (157).

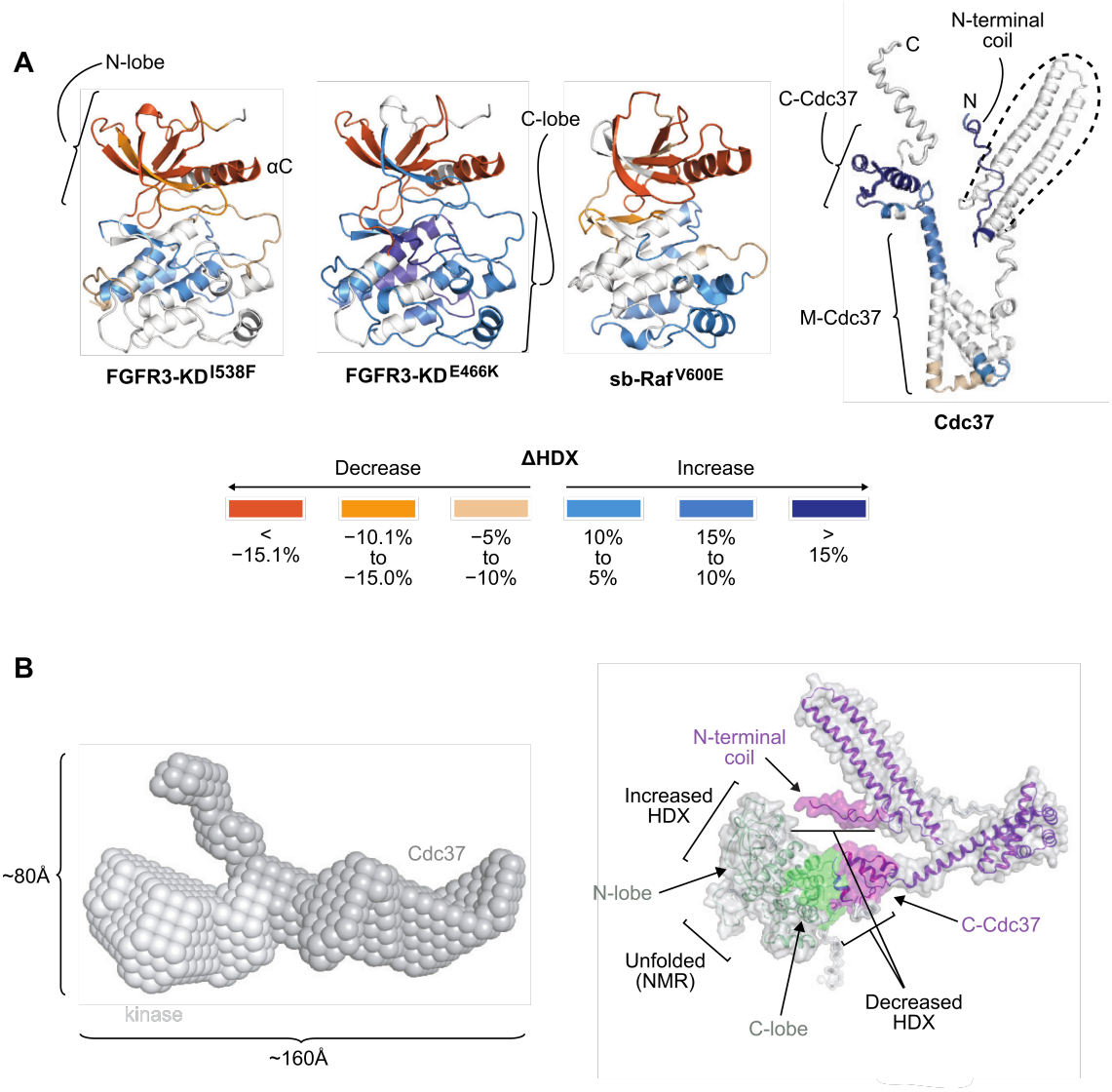


Figure 4.11. Complementary HDX-MS and SAXS data.
(A) The observed changes in the amount of hydrogen-deuterium exchange (HDX) for the kinases FGFR3-KD^{I538F}, FGFR3-KD^{E466K} and sb-RaF^{V600E} in the presence of Cdc37 (left) and for Cdc37 in the presence of FGFR3-KD^{E466K} (right). The region of Cdc37 highlighted by the dotted line did not yield any HDX-MS data. Changes are mapped onto kinase and Cdc37 structures according to the gradient shown, distinguishing regions with increases and decreases in HDX. (B) The SAXS envelope of the FGFR3-KD^{I538F}:Cdc37 complex with those regions attributed to Cdc37 and kinase coloured in dark and light greys, respectively (left). Adjacent, the best fit obtained from modelling of the complex is shown in cartoon with surface representation (right). Regions of the kinase (green) and Cdc37 (purple) which are seen to unfold in NMR or experience changes in HDX are indicated. These figures have been adapted from those kindly provided by collaborators (Roger Williams and Dmitri Svergun laboratories), now published in Bunney *et al.* (322).

4.3.2. Kinases are further remodelled during chaperoning by Hsp90

The published cryo-EM structure of the model kinase Cdk4 bound in a ternary complex to Cdc37 and Hsp90 shows local unfolding of the kinase N-lobe at its β4 and β5 strands (148). While this is consistent with Cdc37-induced remodelling of the kinase N-lobe, the regions of kinases that are observed to be unfolded in binary Cdc37-bound and ternary Cdc37 and Hsp90-bound complexes are not the same, and unfolding of β4 and β5 in the presence of Cdc37 is not observed in these NMR data (Figure 4.12). It is also unlikely that such an extended unfolded state of the kinase as observed in this ternary complex would be energetically favourable without the additional stabilising interactions afforded by Hsp90 (Figure 1.14) (148). Thus, while the data presented here of binary kinase-Cdc37 interactions and the cryo-EM structure of Cdk4 bound to Hsp90 in a ternary complex both tell stories of kinase N-lobe remodelling, it is clear that further structural rearrangements of kinases must occur during their handover from Cdc37 to Hsp90. The remodelling of the kinase N-lobe by Cdc37 presumably makes it more malleable, allowing these further changes to take place.

Additionally, while the SAXS model of Cdc37 binding to FGFR3-KD^{I538F} suggests a rather significant interaction interface between the kinase C-lobe and Cdc37 (Figure 4.11B), only a very small surface of the C-lobe at the N- and C-lobe interface contacts Cdc37 in the ternary complex structure (Figure 1.14) (148), suggesting that kinase-co-chaperone interactions also change during the kinase chaperoning cycle. As the cryo-EM structure captures a kinase bound to the closed nucleotide-bound state of Hsp90 and the Cdc37-bound state corresponds to that before Hsp90 interaction, these structural changes must occur during or near after kinase transfer to the chaperone. The structural and mechanistic details of these early stages of kinase interaction with Hsp90 are currently undefined (170).

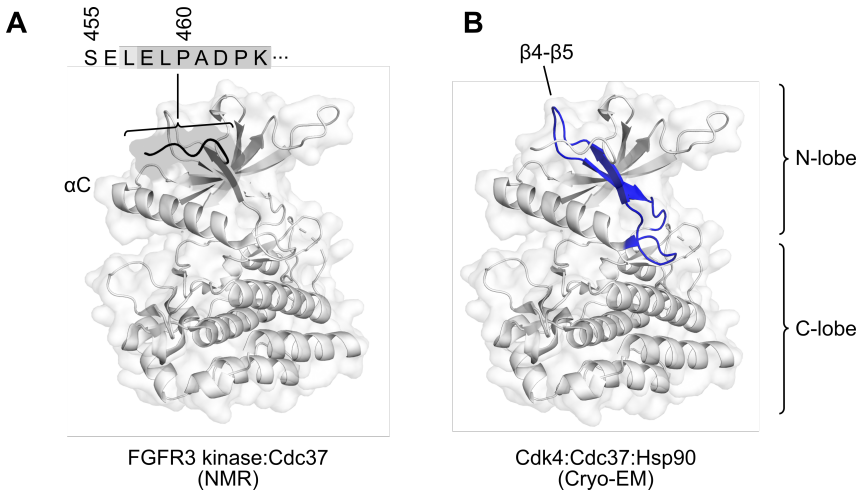


Figure 4.12. Local unfolding of kinases induced by Cdc37 and Hsp90.
A comparison of the locally-unfolded regions of kinases in (co-)chaperone complexes where (A) Cdc37 induces apparent unfolding of the N-terminus of FGFR3 kinase domain through remodelling of the N-lobe and (B) the N-lobe strands β4 and β5 are found to locally unfold in Cdk4 kinase when bound in a closed state ternary complex with Cdc37 and Hsp90. In both cases, the regions found to unfold are highlighted on the structure of FGFR3 kinase domain (PDB entry 4K33).

4.3.3. Summary

Work in this chapter establishes that the co-chaperone Cdc37 induces the local unfolding of kinases of different client strengths at their N-terminus. Though unobserved in the NMR data acquired, this effect was presumed to reflect more global changes occurring elsewhere in the kinase. As features of the N-lobe dictate the strength of client kinases, illustrated in this work to be underlain by a network of residues in the N-lobe which can be perturbed both by kinase-destabilising mutations and kinase inhibitor binding, it was probable that these more extensive structural changes took place there. By integration of NMR data from this chapter with complementary structural data from HDX-MS and SAXS (Roger Williams and Dmitri Svergun laboratories), collaborators were able to define a low-resolution model of kinase recognition and remodelling, suggesting that Cdc37 approaches kinases at their C-lobes and induces remodelling of the kinase N-lobe, partly inactivating the kinase. Given that kinases appear to be remodelled differently when bound to Hsp90, this remodelling by Cdc37 likely makes the N-lobes of kinases more malleable for further structural rearrangements by Hsp90. This work resolves previous uncertainties regarding this client kinase recognition complex, yet raises new questions regarding the mechanistic details of kinase chaperoning. Namely, how do kinases in complex with Cdc37 come to interact with Hsp90? And how does the chaperone Hsp90 further remodel kinases to induce the dramatic unfolding of kinases to separate their two lobes as observed in the closed state of such a complex? This data thus serves as a foundation for additional studies investigating the entry of kinases into the chaperone cycle of Hsp90, as will be explored in the next chapter.

Chapter 5. The entry of kinases into ternary complexes with Hsp90

5.1. Aims

Building upon findings in the previous chapter that Cdc37 remodels the N-lobe of kinases and appears to induce a local unfolding at the kinase N-terminus, but that this remodelling event is not consistent with that observed for Cdk4 bound to Hsp90 in a closed ternary complex state (148), I next looked to study how kinases interact with Hsp90 with a focus on how they are delivered to the chaperone. I would do this through reconstitution of ternary complexes containing FGFR3 kinase domain, Cdc37 and Hsp90. To focus on these early stages of Hsp90's cycle, complexes would initially be reconstituted in the absence of ligands such as ATP, followed by those in their presence to explore their effect on ternary complexes. To understand how or if these complexes differ, and yield mechanistic information regarding the chaperoning of kinases by Hsp90, I would aim to biochemically and structurally characterise these reconstituted complexes using complementary techniques in an integrated structural biology approach.

5.2. Results

5.2.1. Expression of FGFR3 kinase, Cdc37 and Hsp90

To solve the Cdk4-bound ternary complex of Hsp90, following co-expression in insect cells, Verba *et al.* purified the complex directly from cells and in the presence of molybdate, likely therefore capturing the most stable and highly-occupied state of the ternary complex, thought to be the post-hydrolysis state (148). As the cytoplasm is a nucleotide-rich environment (350), purification of the complex in this way is unlikely to capture kinase-bound ternary complexes earlier in the Hsp90 cycle that may be more transient and lowly-populated (i.e. the nucleotide-free state, or those prior to ATP hydrolysis). However, the reconstitution of complexes may provide access to these earlier states. Notably, while it has been reported that reconstitution of kinase-containing complexes is not possible (suggesting that additional cellular components are necessary for their assembly) (148,191,351), each of these previous studies attempted ternary complex reconstitution using Cdk4. In contrast, complexes that were indistinguishable from those assembled in insect cells have been successfully reconstituted with the kinase bRaf (175). This lent promise that ternary complexes comprising FGFR3 kinase may also be possible to be reconstituted, and later shown to be the case (322).

To reconstitute an FGFR3 kinase-bound Hsp90 ternary complex, it was first necessary to express and purify FGFR3 kinase domain, Cdc37 and Hsp90. Given that it is a stronger client than wild-type FGFR3 kinase domain, the I538F variant of FGFR3 would be used for this purpose. Each individual component was recombinantly expressed and purified from *E. coli* using materials and protocols from Dr Tom Bunney. While using

these protocols as guidance, a number of issues were faced during these processes, requiring their troubleshooting and protocol optimisation.

Of the three components, expression and purification of the co-chaperone Cdc37 was the least problematic. Approximately 3 mg of purified protein per litre of culture was obtained following IMAC and ion exchange (IEX) chromatography (Figure 5.1). As phosphorylation of Cdc37 at Ser13 has been described to stabilise kinase-bound ternary complexes (189,190), Cdc37 was co-expressed with casein kinase II to introduce this post-translational modification (352). To confirm that Cdc37 was phosphorylated at Ser13, the purified sample was analysed by mass spectrometry and western blotting using a commercially available antibody that detects phosphorylation at this site (Table 2.2). Intact mass measurements indicated that the purified sample contained four species (Figure 5.1B and Table 5.1). These could be divided into two groups of Cdc37: one corresponding to the expected mass containing two subpopulations of monophosphorylated and unphosphorylated forms, and a second with an ~42 Da mass adduct, each also with monophosphorylated and unphosphorylated forms (Table 5.1). These suggested that approximately 70% of the purified sample was phosphorylated at Ser13. Western blotting confirmed that phosphorylation was at this site (Figure 5.1C). The ~42 Da mass adduct suggested that Cdc37 was also partially amino-acetylated (353); however, efforts to identify the site of this modification by mass spectrometry were unsuccessful. The phosphorylated and unphosphorylated states of Cdc37 were not separated any further.

Expression of Hsp90α in *E. coli* was inconsistent and challenging, with issues relating to slow cell growth and low expression yields encountered. Overcoming these problems required considerable troubleshooting including exploring the use of an alternative *E. coli* expression strain and adapting the procedure of culture inoculation. Notably, it was found that *E. coli* cultures grew very poorly if inoculated using overnight cultures that had reached the stationary phase (showing a growth lag time of up to 4 hours), and also consequently showed poor expression. This was resolved by inoculating the expression media with freshly transformed colonies each time rather than using starter cultures. Purification of Hsp90 by IMAC and IEX chromatography following protocols used for Cdc37 with minor adaptations typically yielded approximately 7 mg per litre of culture (Figure 5.2A).

The N-terminal purification tag of Hsp90 was initially retained in case it would be required in later experiments. However, mass spectrometry experiments suggested that for this tagged Hsp90 sample, the intact protein was ~11 Da heavier than expected and was potentially labile to degradation (Supplementary Figure 4A and B, and Supplementary Table 1), later confirmed by anti-His western blotting (Supplementary Figure 4C).

Consequently, in subsequent purifications of Hsp90, the purification tag of Hsp90 was removed; this yielded intact mass measurements closer to the expected mass for this protein (Table 5.1), and a protein sample that did not suffer so acutely from degradation.

The expression and purification of FGFR3-KD^{I538F} was also problematic. Similar to Hsp90, very slowly growing cultures with low recombinant protein expression were encountered; these could be largely resolved through inoculation of expression media with freshly transformed colonies as with Hsp90. However, difficulties were also met in the lysis of cells in the first step of protein purification. Though enzymatic cell lysis with lysozyme followed by treatment with Triton X-100 successfully lysed *E. coli* cell pellets in the first few batches of kinase expression and purification, this method was consistently ineffective in later attempts. Cell lysis by sonication was also found to be inappropriate; under optimal conditions, approximately 25% of the total *E. coli*-expressed kinase is soluble (Dr Tom Bunney, personal communication), but almost no kinase was found to be soluble following cell lysis by sonication. Following troubleshooting, partial cell lysis was achieved if rapid freeze-thaw cycles were introduced during enzymatic cell lysis. FGFR3-KD^{I538F} was purified by IMAC and IEX chromatography as for Cdc37 and Hsp90 except that its His6-SUMO tag was cleaved and removed by reverse-IMAC prior to IEX (Figure 5.3A). Yields of purified kinase ranged from between 60 to 250 µg per litre of culture.

To obtain a non-phosphorylated kinase, FGFR3-KD^{I538F} was co-expressed with lambda phosphatase. Suppressing activity of the kinase domain in this way was found to improve kinase expression levels previously (Dr Tom Bunney, personal communication), but also ensures that the sample is homogeneous for reconstitution experiments. Intact mass measurements confirmed that the purified kinase was uniformly non-phosphorylated (Figure 5.3B and Table 5.1). Incubation of the kinase with ATP and western blotting with an anti-phosphoFGFR antibody detected autophosphorylation of the kinase activation loop, confirming that the kinase was qualitatively active (Figure 5.3C).

As each component (kinase, Cdc37 and Hsp90) could be purified to greater than 90% purity with IMAC and IEX chromatography alone (estimated from Coomassie-stained SDS-PAGE), to minimise protein losses, these components were not purified any further by size exclusion chromatography (SEC). This was considered unnecessary as this would be performed following complex reconstitution.

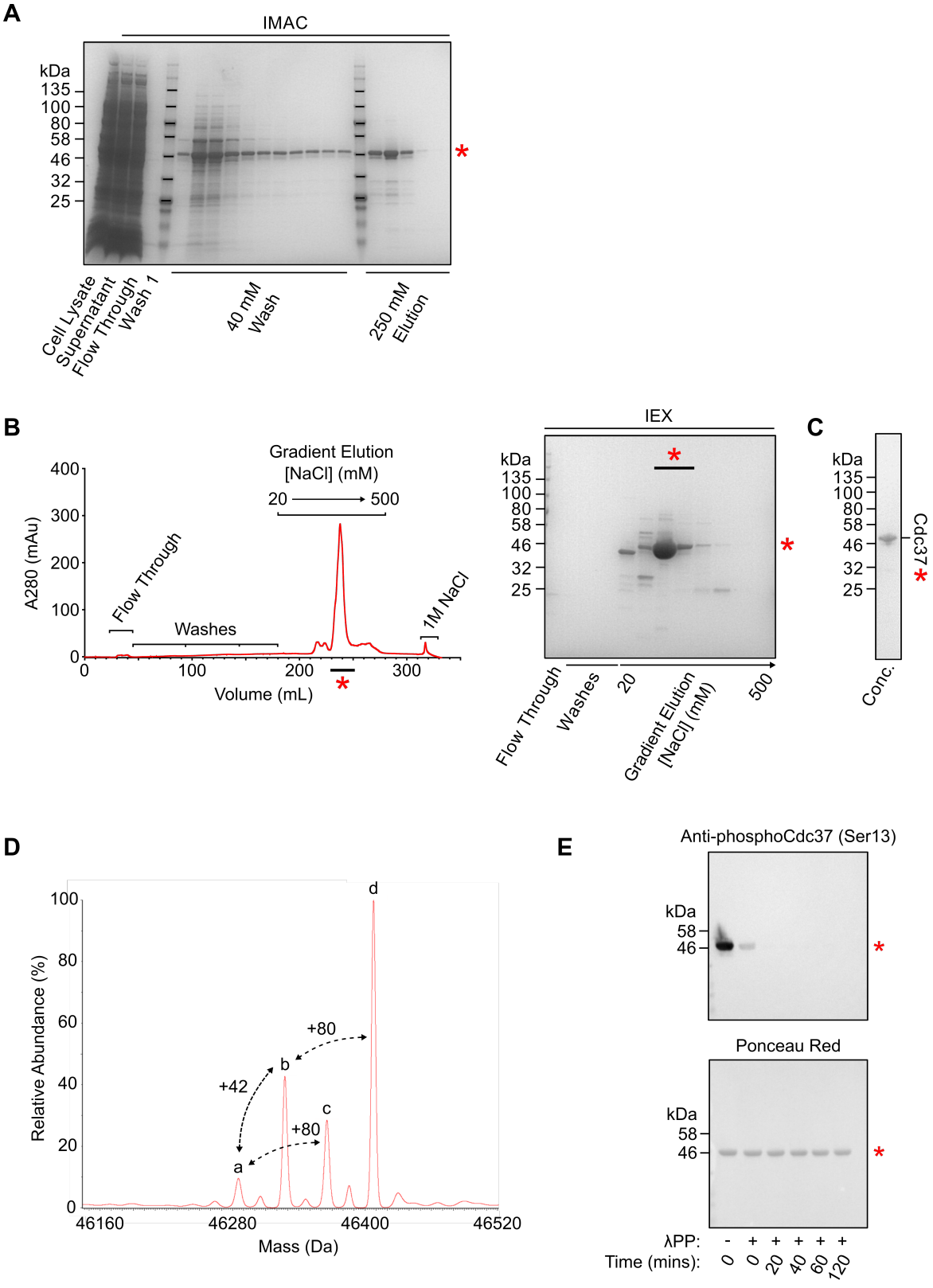


Figure 5.1. Purification of Cdc37.
The caption for this figure follows on the next page.

Following its co-expression in *E. coli* with casein kinase 2, Cdc37 was purified by IMAC using Nickel Sepharose Fast Flow 6 resin (A) then IEX chromatography using a HiTrapQ column (B). Briefly, as described in 2.6.4, following cell lysis, the cell lysate was incubated with nickel resin to capture His6-tagged Cdc37, then unbound and contaminant proteins washed away using His Buffer A containing 40 mM imidazole. Remaining bound proteins were then eluted using 250 mM imidazole. After dialysis into a buffer lacking salt, the eluted proteins were applied to a HiTrapQ column, unbound proteins washed away, then bound proteins eluted with a 20 to 500 mM NaCl gradient. At all stages, samples were taken for analysis by SDS-PAGE with Coomassie staining. Panel A shows this analysis for the IMAC step, and panel B the same analysis and the chromatogram for the IEX step. In both, the red asterisk indicates the location of Cdc37 in each gel. In panel B, the IEX fractions combined to obtain the final purified Cdc37 sample (panel C) are highlighted by a black band and asterisk in both the chromatogram and Coomassie-stained gel. (D) An intact mass spectrum was acquired of the final Cdc37 sample to confirm identity and probe the phosphorylation status of the protein (given its co-expression with casein kinase 2 which phosphorylated Cdc37 at Ser13). This analysis detected four species with some containing mass adducts of +80 Da likely indicating phosphorylation and an approximate +42 Da adduct indicating probable acetylation. The masses and assigned identities of the labelled peaks are given in Table 5.1. (E) To verify the location of Cdc37 phosphorylation and verify that this modification was also removable, Cdc37 was incubated with lambda phosphatase (λPP) and samples taken at various time points. As detailed in 2.8, these samples were analysed by SDS-PAGE followed by western blotting with an antibody that detects phosphorylation of this residue (see Table 2.2). The blot was then stained with Ponceau Red stain to verify equal protein loading in each lane.

Table 5.1. Intact mass measurements of individual components.

Sample ***Expected Mass (Da)***	**Species**	**Observed Mass (Da)**	**Mass Difference* (Da)**	**Annotation**
Cdc37 *46417.47* *46286.47 (−Met)*	a	46284.67 ± 1.54	− 1.8 (−Met)	−Met
	b	46325.75 ± 2.56	+ 39.3 (a)	‘a’ + Ac?
	c	46364.07 ± 3.59	+ 77.6 (a)	‘a’ + Phos.
	d	46405.81 ± 2.75	+ 80.06 (b) + 41.74 (c)	‘b’ + Phos.
Hsp90 *84813.88*	a	84810.87 ± 1.56	− 3.01	-
FGFR3-KD^{I538F} *35478.85*	a	35479.25 ± 0.09	+ 0.4	-

Note: acetylation (Ac); phosphorylation (Phos.); mass difference is with respect to the mean value of the species given in brackets (*); species are labelled as in the appropriate figures (Figure 5.1, Figure 5.2 and Figure 5.3).

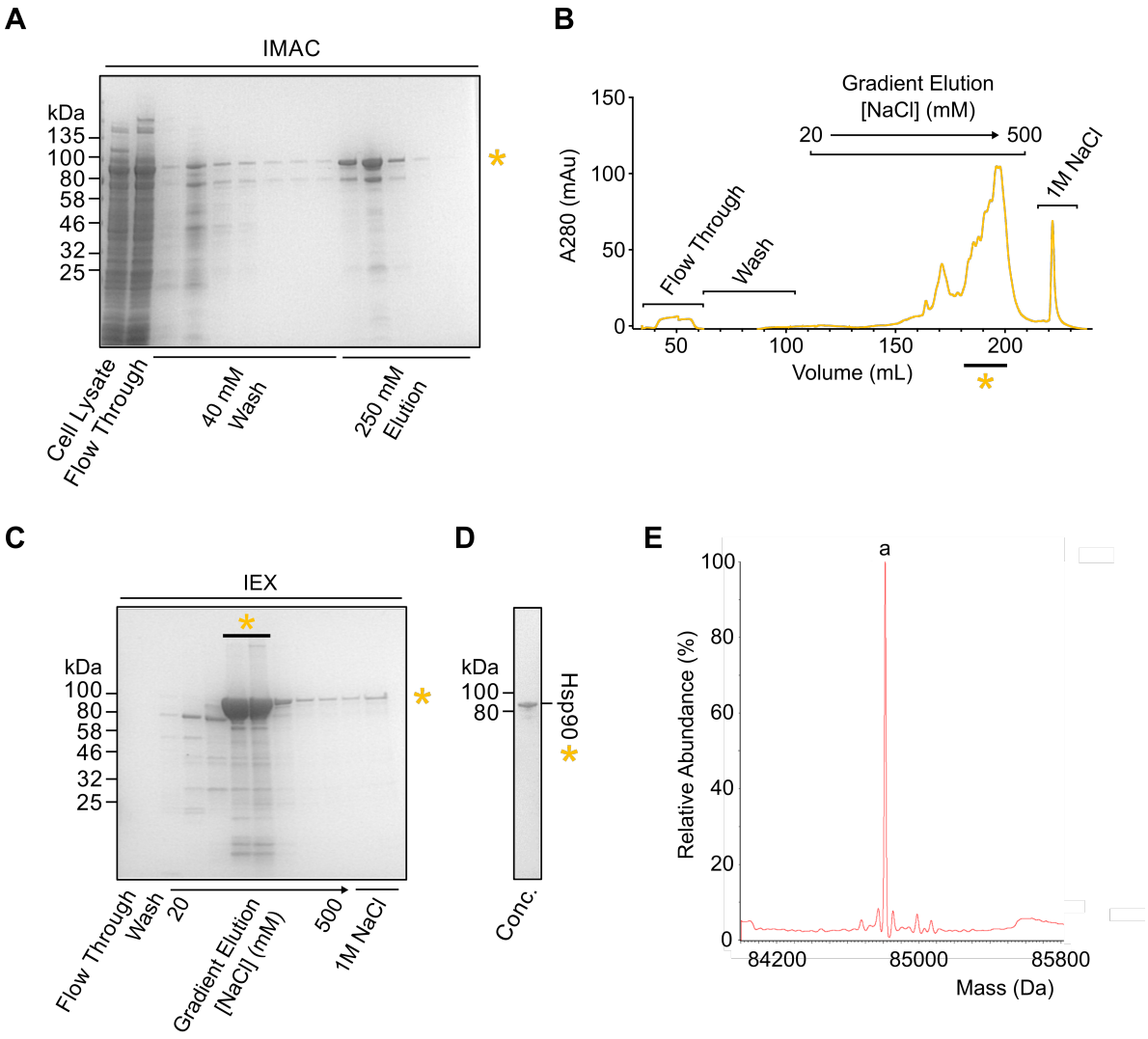


Figure 5.2. Purification of Hsp90.
Following its expression in *E. coli*, Hsp90 was purified by IMAC using Nickel Sepharose Fast Flow 6 resin (A) then IEX chromatography using a HiTrapQ column (B and C). Generally, this purification followed that as detailed in 2.6.4 and Figure 5.1, and samples were taken at each stage for SDS-PAGE analysis with Coomassie staining. Panel A shows this analysis for the IMAC step, and panel C the same analysis for IEX. An example IEX chromatogram is provided in panel B. In each, the orange asterisk indicates the location of Hsp90. In panel B and C, the IEX fractions combined to obtain the final purified Hsp90 sample (panel D) are highlighted by a black band and asterisk. (E) An intact mass spectrum was acquired of the final Hsp90 sample to confirm identity. The mass and assigned identity of the labelled peak is given in Table 5.1. This mass spectrum is of Hsp90 that had had its His6-tag cleaved during dialysis with HRV 3C protease.

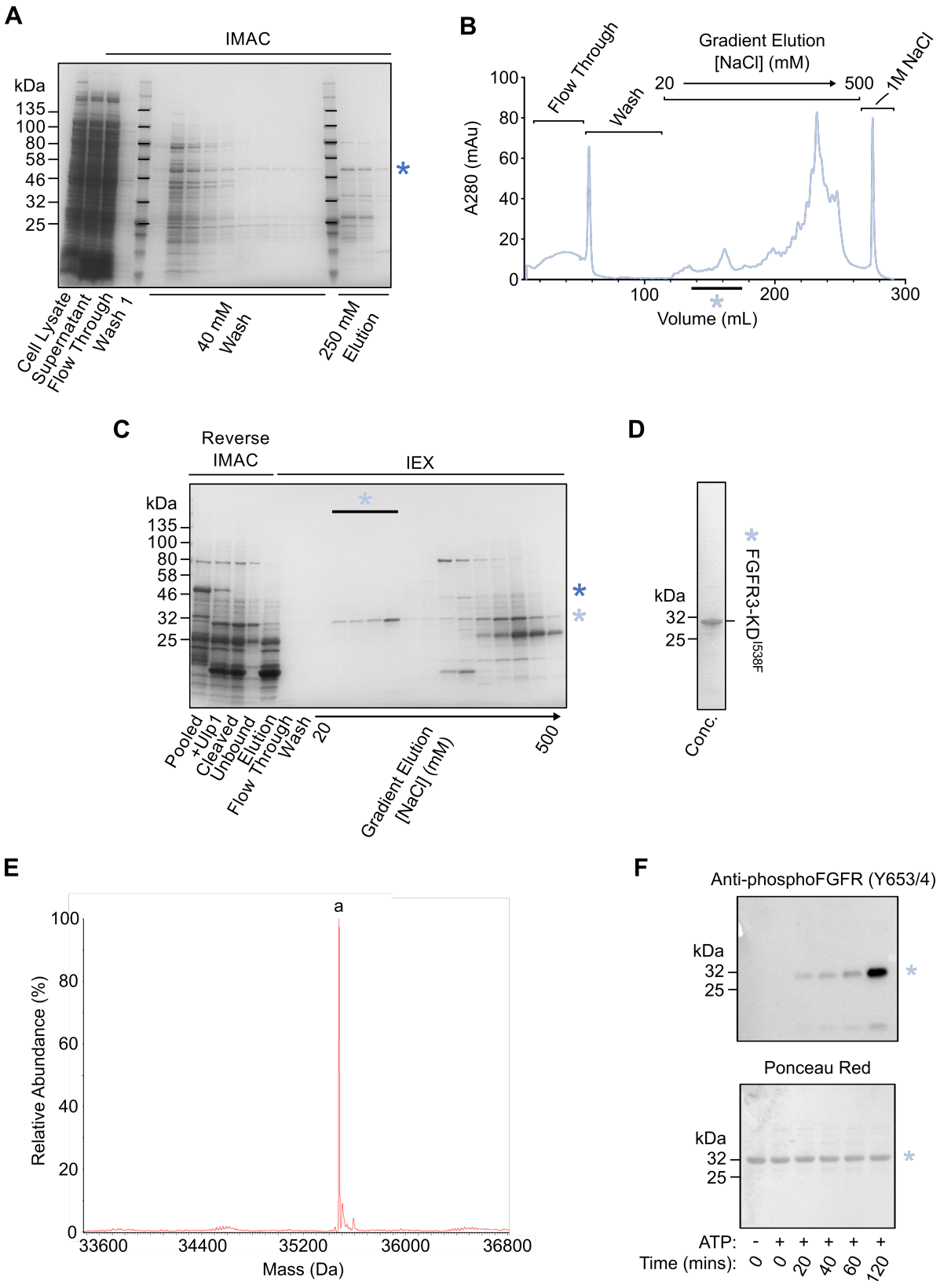


Figure 5.3. Purification of FGFR3-KD^{I538F}.
The caption for this figure follows on the next page.

Following its co-expression in *E. coli* with lambda phosphatase, FGFR3-KD^{I538F} was purified by IMAC using Nickel Sepharose Fast Flow 6 resin (A) then IEX chromatography using a HiTrapQ column (B and C). Briefly, as described in 2.6.4, following cell lysis, the cell lysate was incubated with nickel resin to capture His6-SUMO-tagged kinase, then unbound and contaminant proteins washed away using His Buffer A containing 40 mM imidazole. Remaining bound proteins were then eluted using 250 mM imidazole. After dialysis into a buffer lacking salt, and in the presence of Ulp1 protease to remove the His6-SUMO tag, the dialysed proteins were applied to a HiTrapQ column, unbound proteins washed away, then bound proteins eluted with a 20 to 500 mM NaCl gradient. At all stages, samples were taken for analysis by SDS-PAGE with Coomassie staining. Panel A shows this analysis for the IMAC step, panel B an example IEX chromatogram, and panel C the same SDS-PAGE analysis for the IEX step. In each, the dark blue asterisk indicates the location of tagged FGFR3-KD^{I538F} and the light blue asterisk indicates untagged FGFR3-KD^{I538F}. In panels B and C, the IEX fractions combined to obtain the final purified untagged FGFR3-KD^{I538F} sample (panel D) are highlighted by a black band and asterisk. (D) An intact mass spectrum was acquired of the final FGFR3-KD^{I538F} sample to confirm protein identity and probe the phosphorylation status of the protein (given that it should be non-phosphorylated after its co-expression with lambda phosphatase). This analysis detected a single non-phosphorylated species whose mass and assigned identity is given in Table 5.1. (E) To verify that FGFR3-KD^{I538F} was qualitatively active, as detailed in 2.7, the final purified sample was incubated with ATP and samples taken at various time points for analysis by SDS-PAGE followed by western blotting with an antibody detecting phosphorylation of the kinase activation loop (see Table 2.2). The blot was then stained with Ponceau Red stain to verify equal protein loading in each lane.

5.2.2. Reconstitution of FGFR3 kinase-bound ternary Hsp90 complexes

SEC analysis of the individual components FGFR3-KD^{I538F}, Cdc37 and Hsp90 indicated that each protein was generally pure and monodisperse (some higher mass species were observed with Hsp90, Figure 5.4A). A binary kinase-Cdc37 complex could also be detected following its reconstitution at a 1:1 molar ratio (Figure 5.4B).

As published studies have shown that kinases bind to Cdc37 and Hsp90 in ternary complexes with a stoichiometry of 1:1:2 (kinase:Cdc37:Hsp90) (148,191), this ratio was employed for initial reconstitutions of FGFR3 kinase-bound ternary complexes. In these reconstitutions, no additional nucleotides such as ATP were introduced. As each component was also purified in the absence of nucleotides, these complexes were therefore expected to be nucleotide-free. SEC analysis confirmed that FGFR3-KD^{I538F} could enter a ternary complex with Cdc37 and Hsp90 following their incubation with one another, indicated by the consumption of kinase-Cdc37 binary complexes and the shifting of the Hsp90 elution peak to an earlier elution volume (Figure 5.4C). SDS-PAGE analysis of this peak also confirmed the presence of all three components of the complex (Figure 5.4C). As discussed previously, this is consistent with the published ability to reconstitute a kinase-bound Hsp90 ternary complex with bRaf (175), despite prior reports that such complexes cannot be reconstituted *in vitro* (148,191). As this FGFR3 kinase-

containing reconstituted complex was expected to be nucleotide-free (i.e. the *apo* state), it likely corresponded to a state at or near the start of the chaperone cycle.

While these data showed that nucleotide-free reconstitutions of FGFR3-bound ternary complexes were possible, incomplete consumption of the binary kinase-Cdc37 complex was also noted (Figure 5.4C). This could be problematic for the isolation of pure ternary complexes for structural characterisation as unbound Hsp90 elutes at a similar volume and overlaps with the ternary complex (and incomplete consumption of the binary complex would imply that unbound Hsp90 is present). As free Cdc37 was also observed in SEC analysis of these reconstituted ternary complexes (Figure 5.4), it was hypothesised that there was an insufficient number of kinase-Cdc37 complexes to saturate Hsp90, but that saturation might be attainable if the population of this binary complex was improved. To explore this possibility, reconstitutions were next repeated using either 2-fold or 4-fold molar excesses of the kinase with respect to Cdc37; while these resulted in improved depletion of free Cdc37, unbound binary complexes were still observed, implying that some free, unbound Hsp90 may also still be present (Supplementary Figure 5). It is worth noting here that the binding affinity between Cdc37 and Hsp90 (without kinase) is expected to be in the low to mid-micromolar range (177), and that these reconstitutions used components with concentrations of between 20 and 40 μM. While binary kinase-Cdc37 complexes may have a different affinity to Cdc37 alone, it is possible that complete saturation is therefore unattainable under these conditions.

At the same time as these SEC analyses, ternary complexes were also being prepared and analysed by GraFix (261), a preparatory technique that introduces covalent crosslinking between components of macromolecular complexes to stabilise them. These analyses by GraFix indicated that Hsp90 was indeed unsaturated in ternary complexes reconstituted at a 1:1:2 (kinase:Cdc37:Hsp90) molar ratio, but that this could be improved by reconstituting complexes with excesses of both Cdc37 *and* FGFR3 (discussed in section 5.2.5). SEC analysis of a complex reconstituted with equimolar concentrations of all three components (i.e. 1:1:1 kinase:Cdc37:Hsp90, equating to 2-fold excess of kinase and Cdc37 with respect to dimeric Hsp90) was consistent with this observation by GraFix, showing a ternary complex peak with a more monodisperse elution profile compared to that observed after a reconstitution preformed at 1:1:2 (Figure 5.5). All future reconstitutions of the ternary complex were therefore performed at an equimolar 1:1:1 ratio.

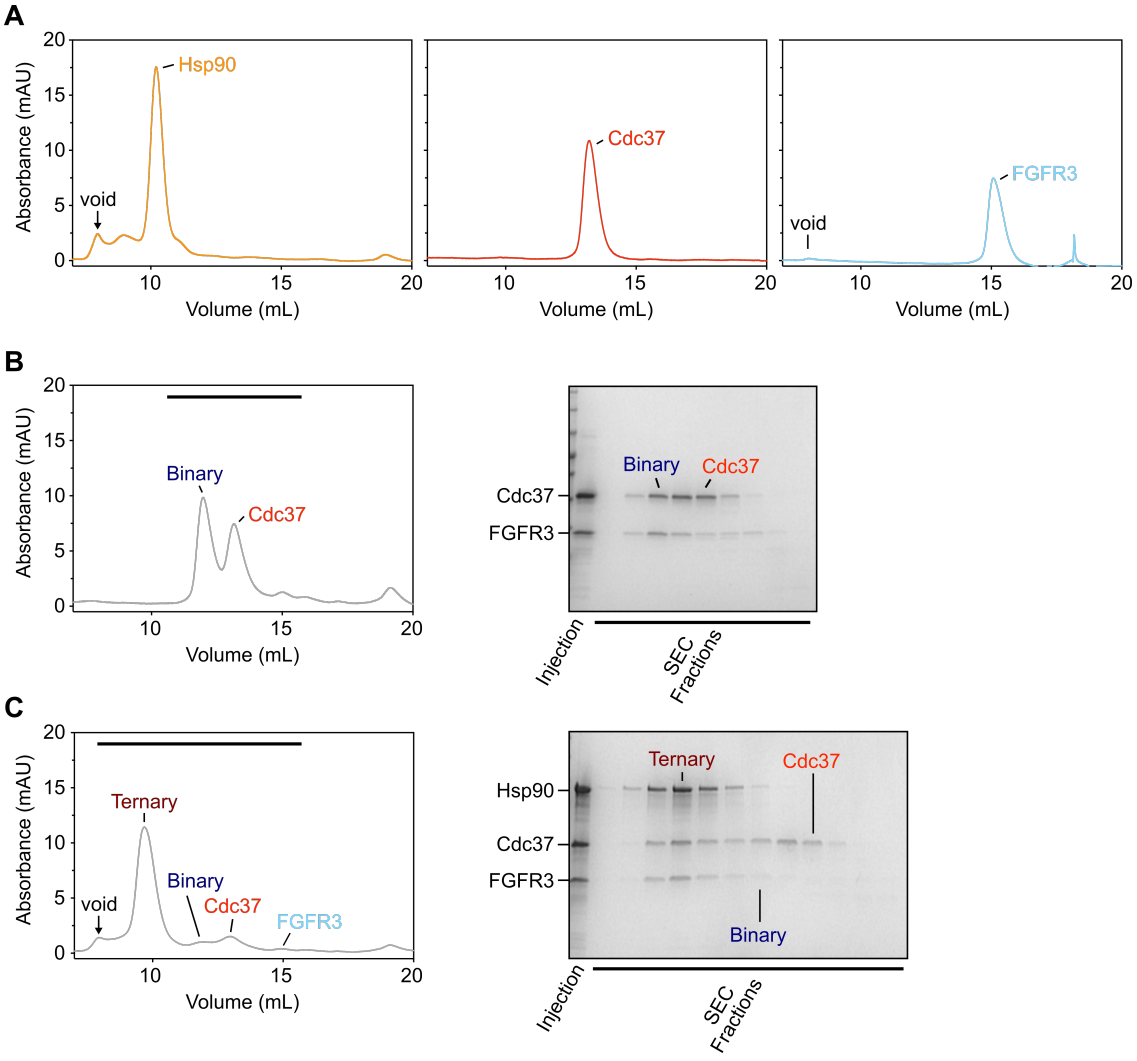


Figure 5.4. Reconstitution of binary and ternary complexes.
(A) To assess the purity of each component of Hsp90, Cdc37 and FGFR3-KD^{I538F} used in complex reconstitution, each component was injected on to a Superdex 200 Increase 10/300 size exclusion chromatography (SEC) column. (B) To verify that the purified Cdc37 and FGFR3-KD^{I538F} were able to form a binary complex, a 1:1 molar mixture of each protein was assessed through SEC as in A, then fractions analysed by SDS-PAGE with Coomassie staining (right panel). The fractions assessed by SDS-PAGE are highlighted by a black bar in the chromatogram. In each (the chromatogram and gel), the location of a binary Cdc37-kinase complex and free Cdc37 are highlighted. (C) Following reconstitution of a ternary complex at a 1:1:2 molar ratio (kinase:Cdc37:Hsp90), the reconstitution sample was analysed by SEC as in A and B, then fractions analysed by SDS-PAGE with Coomassie staining (right panel). The fractions assessed by SDS-PAGE are highlighted by a black bar in the chromatogram. In each (the chromatogram and the gel), the location of ternary and binary complexes, and free Cdc37 and kinase are highlighted (see SEC chromatograms in panel A for a comparison where each individual component was found to elute using the same SEC column).

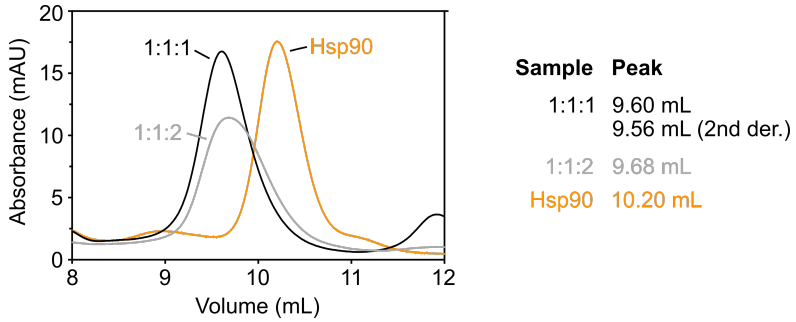


Figure 5.5. Optimisation of ternary complex reconstitution.
SEC traces of complexes reconstituted using a 1:1:1 or 1:1:2 molar ratio of kinase:Cdc37:Hsp90, and of Hsp90 alone after injection on to a Superdex 200 Increase 10/300 column as detailed in 2.12. These traces illustrate the movement of the Hsp90 peak to an earlier elution volume upon forming a ternary complex, and that optimisation of complex reconstitution (i.e. reconstituting at a 1:1:1 ratio rather than 1:1:2 ratio) led to a more uniform and monodisperse SEC profile for the complex suggesting improved Hsp90 consumption. To illustrate this peak movement, the elution volume of the peak is given in each case, determined generally by calculation of the first derivative of the curve. For the 1:1:1 injection, the peak position estimated from the second derivative is also given (relevant for later comparisons to ligand-bound complexes (Figure 5.7)).

5.2.3. Confirmation of ternary complex stoichiometry

Though previous experiments indicate that a kinase-bound Hsp90 ternary complex is of a 1:1:2 stoichiometry (kinase:Cdc37:Hsp90), as both of these studies were completed using the kinase Cdk4 (148,191), to confirm that this stoichiometry is shared among kinases of different types, it was important to demonstrate that this was also the case for reconstituted ternary complexes containing FGFR3 kinase domain. This was essential to determine whether findings made here and from previous studies regarding kinase chaperoning are indeed relevant to the kinase family as a whole.

In the first instance, this was approached using SEC-MALS. Mass measurements by SEC-MALS were tried using Superdex 200 Increase columns of two different dimensions (5/150 and 10/300). While the former had higher throughput, this column with a smaller resin bed volume gave considerably worse resolution than the latter, resulting in traces inappropriate for complex mass estimation. Analyses therefore required the use of a Superdex 200 Increase 10/300 column, as had been used for SEC analyses and sample optimisation. These estimated the mass of the ternary complex to be between ~249 and 259 kDa, masses most consistent with a complex with a 1:1:2 stoichiometry given the mass of each component (an expected mass of ~256 kDa with tagged Hsp90) (Table 5.2). Due to technical problems in the collection of refractive index data for absolute mass determination in these experiments, these molecular mass estimations were obtained using the ultraviolet traces and an estimated extinction coefficient of the complex; this is a valid but non-ideal approach, as assumptions are made when estimating extinction coefficients that can introduce errors in mass estimation.

Table 5.2. SEC-MALS analysis of the *apo* ternary complex.

Sample	Molar Ratio (kinase:Cdc37:Hsp90)	Concentration (μM)	Estimated Mass (kDa)
1	1:1:1	25	249 ± 0.024%
2	1:1:1	25	259 ± 0.247%

Note: masses were estimated using the UV absorbance of traces and a calculated extinction coefficient for the complex; the reported molar ratio and concentrations (with respect to Hsp90) are those of the injected sample.

As an orthogonal approach, the reconstituted complex was also submitted for analysis by native mass spectrometry (MS) to obtain absolute estimations of the ternary complex mass and stoichiometry. These experiments were performed by Dr Rachel George (Faculty of Biological Sciences Mass Spectrometry Facility). Analysis of the mass of the ternary complex by this method presented its own challenges, most notably that the complex was found to be very sensitive to ionisation-induced disassembly. Nevertheless, analysis of a complex reconstituted with equimolar components successfully detected species at ~250 kDa and ~169 kDa (Figure 5.6A and Table 5.3); these were most consistent with the expected 1:1:2 stoichiometry ternary complex (an expected mass of ~251 kDa, on this occasion with untagged Hsp90) and of dimeric Hsp90 (an expected mass of ~169 kDa). Native MS analysis of Hsp90 alone confirmed that the latter mass did indeed correspond to dimeric Hsp90 (Figure 5.6A and Table 5.3). This submission also detected a major population of monomeric Hsp90, once more highlighting sensitivity of the sample to ionisation-induced disassembly.

While the observed ~250 kDa species by native MS was most consistent with the expected 1:1:2 stoichiometry complex, it was approximately 1 kDa lighter than expected from the individual component masses. To troubleshoot this discrepancy, given concerns that this might indicate a (possible) time-dependent degradation of a complex component, the complex was resubmitted for native MS analysis while ensuring a minimal delay between sample submission and data collection. On this occasion, multiple complex species were detected (Figure 5.6B and Table 5.3). Though a complex approximately 1 kDa lighter was detected once more, the most abundant species in this data had an observed mass that exactly matched the expected mass of a 1:1:2 complex containing Cdc37 that was Ser13 phosphorylated and (likely) acetylated (see Table 5.1). This finding suggested that the reconstituted ternary complex was indeed nucleotide-free, and confirmed unequivocally that FGFR3 binds to Hsp90 with the same stoichiometry as Cdk4 (148,191). It also suggested that Ser13-phosphorylated Cdc37 was self-selected for during ternary complex formation.

Curiously, in addition to the two species corresponding to 1:1:2 stoichiometry ternary complexes (species ‘a’ and ‘b’, Table 5.3), two well-populated species approximately 82 kDa heavier than these were also identified (Figure 5.6B and Table 5.3). As the binary kinase-Cdc37 complex has a molecular mass of 81,885 Da, it is probable that these complexes correspond to an Hsp90 dimer bound to two copies of binary complex. This is the first time that such a complex has been detected, though it should be theoretically possible given that there are two possible NTD interaction sites in an Hsp90 dimer, the domain of Hsp90 to which Cdc37 interacts with at a 1:1 stoichiometry (149,177).

Due to limited resolution in size exclusion chromatography between the void volume of the column and the presumed 1:1:2 ternary complex peak, it is currently unclear whether this 2:2:2 complex is also present during SEC analysis or whether it is uniquely detected by the more sensitive native MS. As these two 2:2:2 complex species in native mass spectrometry differ by 1,241 Da (species ‘c’ and ‘e’, Table 5.3), and those corresponding to a 1:1:2 complex differ by 1,216 Da (species ‘a’ and ‘b’, Table 5.3), it is likely that Hsp90 is the truncated component in these in complexes which are ~1 kDa light, rather than the kinase or Cdc37. This suggestion is consistent with earlier observations that Hsp90 was labile to degradation (see results 5.2.1). Curiously, intact mass measurements of the same complex sample that was submitted for native MS analysis detected all three complex components within 7 Da of their expected masses. This made it difficult to confirm the location of this apparent truncation, and suggested that it might arise uniquely during analyses by native MS.

In the same native MS submission, a final protein species approximately 35 kDa lighter than a 1:1:2 complex was also detected. This most likely corresponds to a complex of dimeric Hsp90 bound to Cdc37 but that is missing the kinase domain (Figure 5.6B and Table 5.3). It is unclear at this stage whether this complex was formed through the binding of free Cdc37 to Hsp90, or whether it arises through dissociation of the kinase from a complete ternary complex.

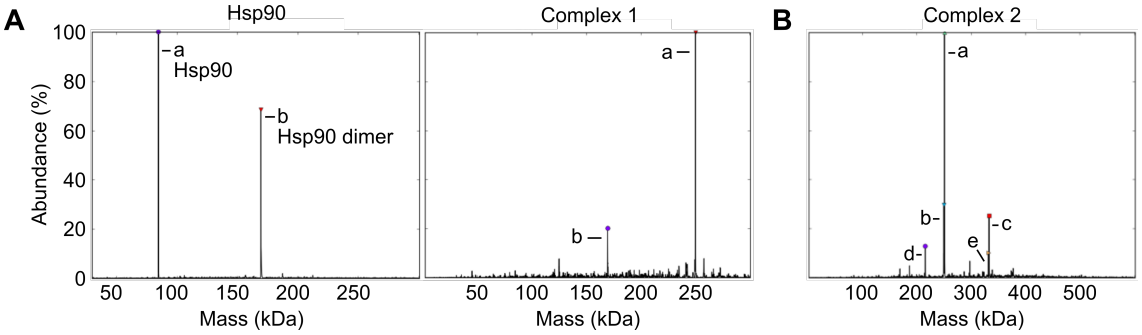


Figure 5.6. Native mass spectrometry of the *apo* ternary complex.
(A) Native mass measurements of Hsp90 (left) and a reconstituted *apo* ternary complex (right) as detailed in 2.15. (B) Native mass spectrum of a second reconstituted *apo* ternary complex, on this occasion detecting multiple species, including both 1:1:2 and 2:2:2 stoichiometry complexes (kinase:Cdc37:Hsp90). In each mass spectrum, the detected species are labelled 'a' to 'e' corresponding to those species listed in Table 5.3.

Table 5.3. Complex species observed by native mass spectrometry.

Species	Relative Abundance (%)	Observed Mass (Da)	Mass Difference (Da)	Annotation
		Hsp90		
a	100	84,814	0	Monomer
b	69	169,624	+ 84,810	Dimer
		Complex 1		
a	100	250,325	− 1,188	1:1:2*
b	20	169,640	− 81,873	Hsp90 Dimer
		Complex 2		
a	100	251,513	0	1:1:2
b	29.64	250,297	− 1,216	1:1:2*
c	25.23	333,419	+ 81,906	2:2:2
d	12.79	216,025	− 35,488	− kinase
e	10.26	332,178	+ 80,665	2:2:2*

Note: the expected mass monomeric untagged Hsp90 is 84,814 Da; of dimeric untagged Hsp90 is 169,628 Da; of a 1:1:2 (kinase:Cdc37:Hsp90) complex is 251,513 Da; of a complex containing dimeric Hsp90 and Cdc37 only is 216,035 Da (i.e. '-kinase'); and of a 2:2:2 complex is 333,398 Da; complex missing ~1 kDa through truncation (*); species are labelled as in Figure 5.6.

Table 5.4. SEC-MALS analysis of ligand-bound complexes.

Sample	Molar Ratio (kinase:Cdc37:Hsp90:ligand)	Concentration (μM)	Estimated Mass (kDa)
+ ATP	1:1:1:5	20	210 ± 0.086%
+ PU-H71	1:1:1:5	20	223 ± 0.191%

Note: masses were estimated using the change in refractive index (dRI); the molar ratios and concentrations (with respect to Hsp90) given are of the injected samples.

5.2.4. The influence of ligands on ternary complex formation

The ternary complexes reconstituted thus far have been nucleotide free. To explore whether ATP or other ligands binding at the nucleotide binding pocket alter complex formation or composition, complexes were reconstituted in the presence of a 5-molar excess of either ATP or the Hsp90 inhibitor PU-H71. In both cases, SEC analysis of these samples saw an altered elution volume of the peak corresponding to the ternary complex (Figure 5.7A). Moreover, two to three complex populations were observed in both cases, highlighted by peak fitting of the SEC traces (Figure 5.7A). In each case, a population eluting at the same volume as the *apo* complex was detected in addition to a second population with retarded elution.

The altered migration of the ligand-bound complexes, located between the elution volumes observed for the ternary complex and Hsp90 alone, could be explained by one of two scenarios: ATP or PU-H71 generate complexes that are missing a protein component (i.e. are physically lighter), or, given that SEC separates proteins both by their size and mass, that they induce ternary complexes containing all three components to obtain a more compact, globular state. While SDS-PAGE analysis of fractions from these SEC experiments suggest that all three components are present over this elution region (Figure 5.7B), as these different complex populations are not sufficiently resolved, it is not possible to confidently conclude that both overlapping populations contain all three complex components. It is also possible that the SEC traces observed in the presence of each of these ligands might reflect fast exchange between different complex conformations or compositions. It is noted that ATP is a substrate for both Hsp90 and FGFR3 kinase, and thus ternary complexes prepared its presence may be additionally or universally ADP-bound given that ATP may be hydrolysed by either component.

As before, the stoichiometry of these complexes formed in the presence of either ATP or PU-H71 was probed using SEC-MALS. A complex reconstituted in the presence of PU-H71 was also submitted for mass analysis by native MS. These data reported conflicting and therefore inconclusive mass estimations. While SEC-MALS suggested that complexes reconstituted with either ligand were lighter than the *apo* complex (Table 5.4), preliminary native MS analysis of the complex reconstituted with PU-H71 detected species with masses that were more consistent with the retention of all three protein components (Figure 5.7C and Table 5.4). However, these masses detected by mass spectrometry were each only ~400 Da heavier than the ~250 and ~251 kDa complexes detected for the *apo* complex, lighter than a single unit of the inhibitor which has a mass of 512.37 g/mol. Native MS data have not yet been acquired for complexes reconstituted in the presence of ATP for comparison. These analyses should be repeated in the future to confirm their findings and to try and resolve these observed discrepancies.

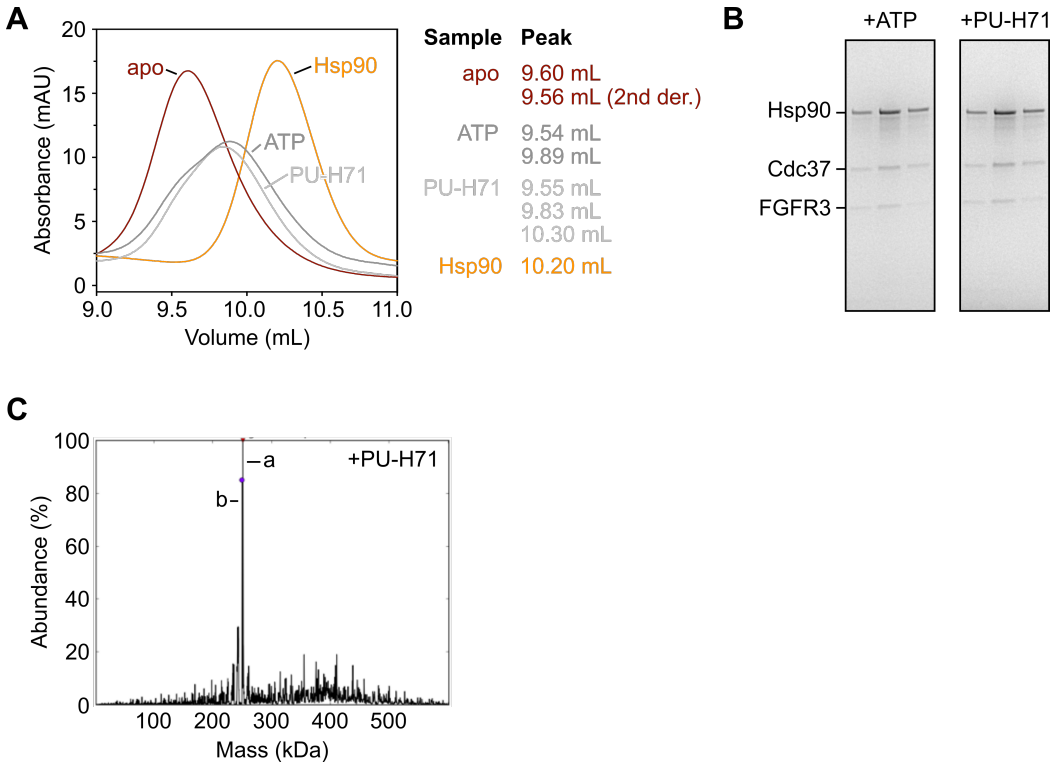


Figure 5.7. Analysis of ligand influence on ternary complex formation.
(A) To analyse the effect of nucleotides on ternary complex formation, complexes were reconstituted at a 1:1:1:5 ratio (kinase:Cdc37:Hsp90:ligand) with either ATP or the Hsp90 inhibitor PU-H71 which binds at the nucleotide binding pocket, then these samples analysed through injection on to a Superdex 200 Increase 10/300 column. These SEC traces were compared to injections of Hsp90 alone or a ternary complex reconstituted in the absence of nucleotides (apo). Complex reconstitution in the presence of ATP or PU-H71 resulted in a retarded peak elution profile with multiple species observed between the elution volume of the apo, nucleotide-free complex and Hsp90, respectively. For each sample, the elution volumes of peaks identified by peak fitting are provided for each curve, calculated from either the first or second derivative of the curves. (B) SEC fractions covering the elution region illustrated in panel A were analysed by SDS-PAGE with Coomassie staining. This illustrated the possible presence of all three protein components in these retarded elution peaks, though it is noted that the multiple complex species are insufficiently resolved in these SEC analyses. (C) To assess whether all three protein components were indeed present in a 'nucleotide-bound' complex, a ternary complex prepared in the presence of a 5-fold molar excess PU-H71 was analysed by native mass spectrometry. Reconstitution with PU-H71 was chosen in this instance to avoid complications of potential ATP turnover by either Hsp90 or kinase. The detected species are labelled as in Table 5.5.

Table 5.5. Species observed by native mass spectrometry with PU-H71.

Species	Relative Abundance (%)	Observed Mass (Da)	Mass Difference** (Da)	Annotation
a	100	251,893	+ 380	1:1:2 + 380 Da adduct
b	84.77	250,688	− 825	1:1:2* + 391 Da adduct

Note: mass differences calculated with respect to a 1:1:2 complex with an expected mass of 251513 Da (**); complex missing ~1 kDa through truncation (*); species labelled as in Figure 5.7C.

5.2.5. Sample preparation for electron microscopy

Having found that FGFR3 kinase-bound ternary complexes could be reconstituted in a nucleotide-free state, thus likely capturing a state located early in the chaperone cycle of Hsp90, I next aimed to obtain structural information about this complex using EM. While SEC elution fractions were of a sufficient concentration to prepare negative-stain EM grids to assess the quality of the sample and obtain low-resolution insights regarding complex structure, the preparation of cryo-EM grids required concentration of the isolated ternary complex. Substantial losses were made during this procedure using ultrafiltration units, indicating that the complex was unstable and/or was involved in non-favourable interactions with the polyethylene sulfate (PES) membrane of these units.

In effort to stabilise the complex for structural studies, GraFix (Gradient Fixation) was tested to reconstitute and prepare ternary complexes for EM. As introduced previously, this preparatory technique introduces covalent crosslinking between components of macromolecular complexes to stabilise them, enabling the study of those that are labile and prone to denaturation and/or disassembly. Using guidelines in the GraFix manual (261), a protocol was devised to enable crosslinking and purification of the FGFR3 kinase-bound Hsp90 complex. Initially, as for SEC analyses, complexes were reconstituted at a 1:1:2 (kinase:Cdc37:Hsp90) molar ratio; while SDS-PAGE analysis of crosslinked and non-crosslinked control fractions of the GraFix gradient indicated that the ternary complex could be reconstituted and isolated by this method, it was clear that Hsp90 was not saturated with kinase and Cdc37 when reconstituted under these conditions, as dimeric Hsp90 was observed to co-exist in fractions with the ternary complex (Figure 5.8A). As discussed previously (see 5.2.2), through altering the molar excesses of kinase and Cdc37 that ternary complexes were reconstituted with, GraFix enabled the optimisation of ternary complex reconstitution, finding that those prepared with equimolar concentrations of each component (corresponding to a 2-fold excess of kinase and Cdc37 with respect an Hsp90 dimer) led to improved depletion of free Hsp90. This was evident in SDS-PAGE analyses both by the reduction of Hsp90 dimers in the complex fractions of crosslinked samples, and by more equal amounts of each component being present in these same fractions of the control, non-crosslinked sample (Figure 5.8B and Figure 5.9A).

Optimisations to the GraFix protocol also needed to be made elsewhere: notably, though recommended by the manual to quench excess of the crosslinker glutaraldehyde with glycine, doing so induced a colour change in the fractions of the gradient (Figure 5.8C). During initial GraFix preparations, this was helpful to demonstrate that linear glycerol and glutaraldehyde gradients had been successfully prepared; however, it was found to be detractive to the monitoring of protein concentration using absorbance at 280 nm, even

after performing buffer exchange. In later preparations, the colour change introduced by quenching of excess glutaraldehyde was minimised by quenching with Tris rather than glycine (Figure 5.8D). Although ternary complexes could be successfully isolated by GraFix, crosslinking of the complex made only minor improvements to its recovery during concentration. GraFix-purified complexes also suffered from a unique quirk where their maximum absorbance occurred at ~265 nm rather than 280 nm as would be expected of pure proteins (Figure 5.8E). This was attributed to a chemical adduct introduced during crosslinking and quenching, but meant that the concentration of GraFix ternary complexes could not be accurately estimated using absorbance at 280 nm following their concentration. Additionally, due to the low recovery achieved during concentration of the complex when using PES ultrafiltration units, a BCA assay was also unable to determine the sample's protein concentration. While it was found that a suitable (though limiting) concentration was achieved that allowed for the preparation of cryo-EM grids, being unable to accurately estimate the concentrations used in the preparation of these grids made efforts to optimise their preparation difficult.

A number of avenues were explored in attempt to troubleshoot the poor yields obtained during protein concentration, using both complexes prepared by SEC and GraFix, and of nucleotide free or PU-H71-bound complexes. Through testing the recovery of proteins following concentration using Vivaspin PES ultrafiltration units of different molecular weight cut offs (30K and 100K), it was clear that these losses were not a consequence of complex disassembly and loss of lower molecular weight components through the 100K filter as used in initial concentration efforts, as all three protein components were retained in the concentrate when using either unit (Figure 5.9B). Moreover, as no proteins were present in the filtrate, this implied that the complex was being lost on the membrane itself, later shown to be the case when testing sample concentration using inverted Proteus X-spinner ultrafiltration units, also with PES membranes (Figure 5.9C). While these inverted concentration units did improve sample recovery following concentration, approximately 85% of the sample was still being lost here, as estimated from quantification of Coomassie-stained SDS-PAGE gels using densitometry. On one occasion, sample concentration was tested using an ultrafiltration device with a regenerated cellulose membrane rather than a PES membrane; this appeared to improve sample recovery, but at least 60-80% of the sample was still unaccounted for (Figure 5.9D). As an alternative to the use of ultrafiltration units, which were all universally poor at concentration of the complex, concentration through immobilisation of the ternary complex onto nickel resin using the C-terminal His6 purification tag retained on Cdc37 was also explored. Though found to be technically possible, concentration of the complex in this manner was practically challenging and would require further optimisation in the future (Figure 5.9E).

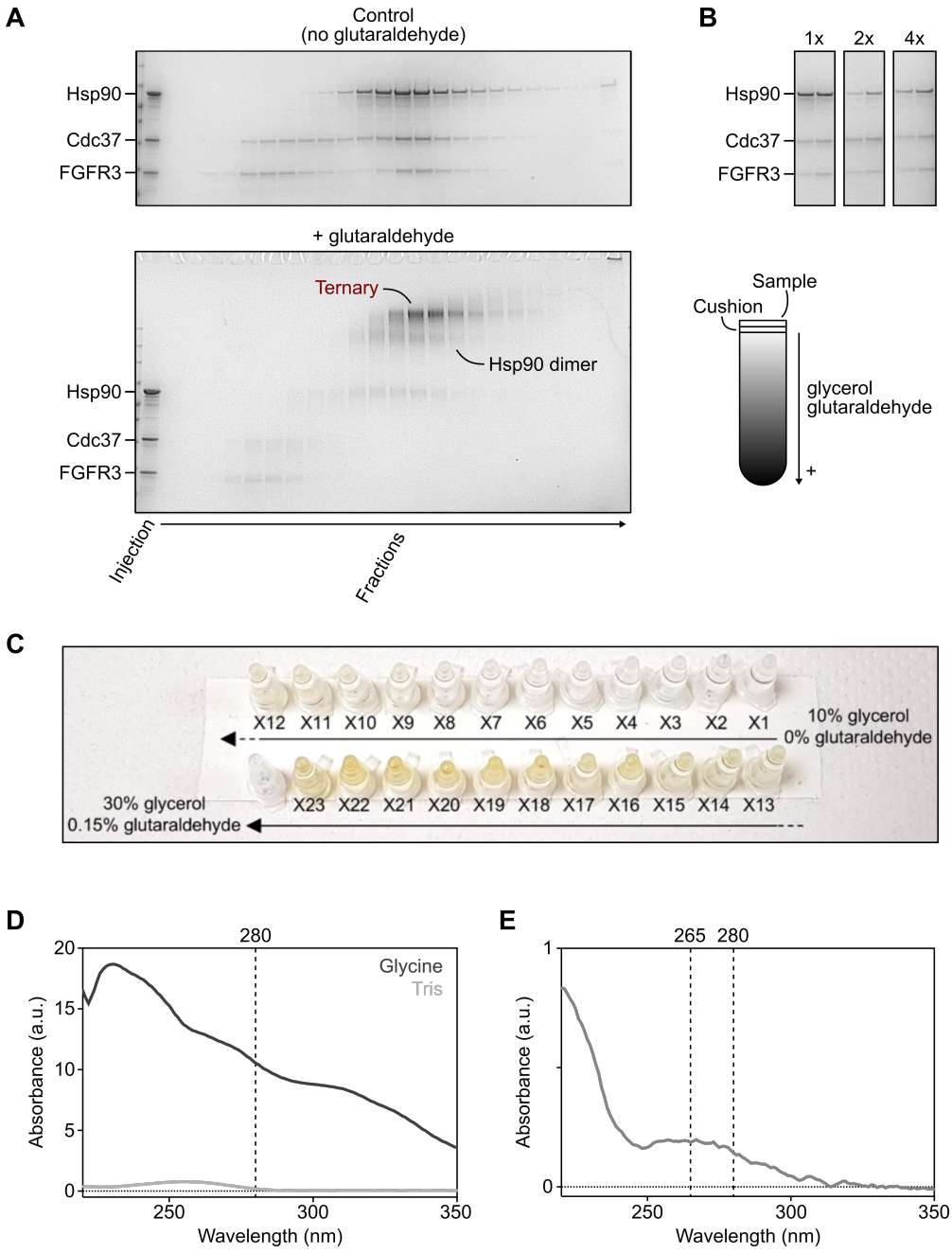


Figure 5.8. Preparation of ternary complexes using GraFix.
(A) Coomassie-stained SDS-PAGE gels of GraFix fractions from gradients with and without the crosslinker glutaraldehyde after loading a ternary complex sample reconstituted at a 1:1:2 molar ratio (kinase:Cdc37:Hsp90) as detailed in 2.12.3. Species speculated to belong to the ternary complex and Hsp90 dimers are highlighted. In the bottom right corner, an illustration of sample preparation by this technique is given. (B) Coomassie-stained SDS-PAGE gels showing optimisation of complex reconstitution. The complex fractions from control gradients lacking crosslinker are shown for samples reconstituted at 1:1:2 (1x), 1:1:1 (2x) and 2:2:1 (4x) molar ratios. (C) A photograph of glycine-quenched GraFix fractions illustrating the quenching-induced colour change. The fractions are labelled from 'X1' to 'X23' corresponding to fractionation from the top to bottom of the gradient. (D) Absorbance spectra of GraFix fractions quenched with either glycine or Tris illustrating a reduction in the background absorbance signal generated. (E) The absorbance spectrum of a GraFix-prepared ternary complex following sample concentration, illustrating maximum absorbance at ~265 nm rather than 280 nm.

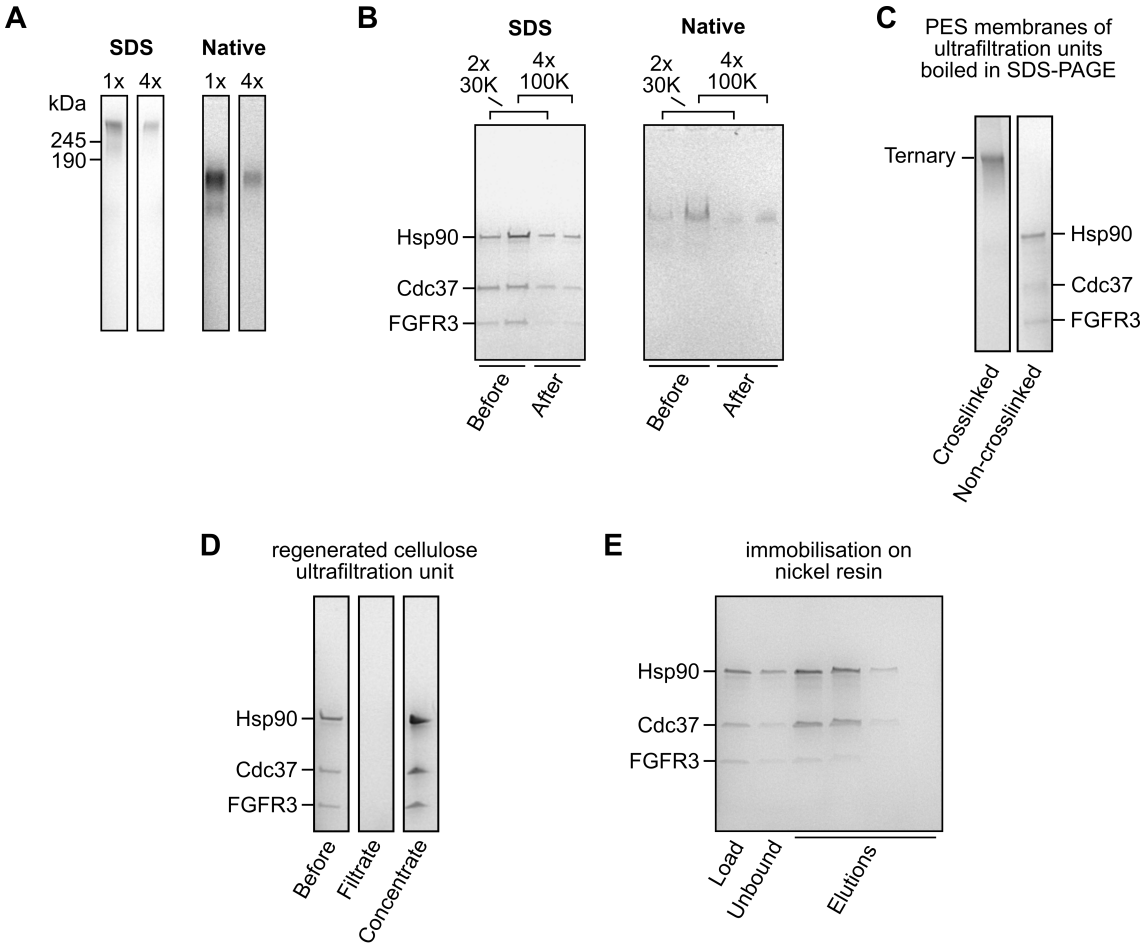


Figure 5.9. Troubleshooting the concentration of ternary complexes.
Reconstituted ternary complexes were found to be difficult to concentrate, and thus various approaches were used in attempt to optimise concentration. (A) Complexes were first concentrated using PES Vivaspin ultrafiltration units. While the complex could be concentrated using these units, as demonstrated with Coomassie-stained SDS-PAGE and Native-PAGE gels of crosslinked complexes prepared at either 1:1:2 (1x) or 2:2:1 (4x) molar ratios of kinase:Cdc37:Hsp90, the majority of the complex was lost during this process. (B) To investigate further, non-crosslinked complexes prepared at either 1:1:1 (2x) or 2:2:1 (4x) molar ratios were concentrated with either 30K or 100K MWCO PES Vivaspin ultrafiltration units. Analysis of the sample before and after concentration through SDS-PAGE and Native-PAGE with Coomassie staining as in panel A, suggested that all three protein components were retained during concentration with these filters and that these proteins remained in complex with one another, indicated by migration as a single band in Native-PAGE. (C) Following attempts to concentrate the complex using alternative ultrafiltration units (PES Proteus X-spinner devices) and the observation of similar protein losses, the membranes from these ultrafiltration devices were isolated and boiled in SDS-PAGE loading buffer. SDS-PAGE analysis of these samples following the concentration of both crosslinked and non-crosslinked complexes indicated that the missing fraction of the concentrated complex sample appeared to be bound to the membrane filter.

The caption for this figure continues on the following page.

Continued caption for Figure 5.9:

(D) Concentration of a non-crosslinked complex was also assessed when using an ultrafiltration device with a regenerated cellulose membrane. Coomassie-stained SDS-PAGE analysis of fractions before and after concentration indicated that these devices could also concentrate the complex, though at least 60% of the sample was still lost during this process. As no proteins were detected in the filtrate, this suggested that the complex was being lost on the device membrane filter once more. Each panel is from the same gel. (E) As an alternative to complex concentration using ultrafiltration devices, given that Cdc37 (but no other proteins in the complex) retains a His6-tag, concentration of the complex through its immobilisation onto Nickel Sepharose Fast Flow 6 resin following SEC purification was tested. In this trial, as detailed in 2.12.4, the SEC-purified non-crosslinked complex was incubated with nickel resin then unbound proteins washed away. Bound proteins were then eluted with 250 mM imidazole. Samples were taken at each stage and analysed by SDS-PAGE with Coomassie staining.

5.2.6. Negative-stain EM analysis of the *apo* ternary complex

To gain initial low-resolution information regarding the FGFR3-bound ternary complexes and to assess sample quality for higher resolution studies by cryo-EM, the *apo*, nucleotide-free complex was analysed by negative-stain EM. Micrographs were prepared for the complex both without crosslinking and following purification by GraFix (Figure 5.10A); as the latter showed better particle concentration and distribution, this sample was used for structural analysis.

Two-dimensional classification of 8,470 manually-picked particles of the GraFix-purified complex generated 2D classes that were consistent in size and shape with distinct features (Figure 5.10B). Following reduction of the dataset by discarding particles that contributed to poorly-defined classes, 2,208 particles corresponding to ~25% of the starting dataset were used for *ab initio* generation of an initial 3D model and then classified into two 3D classes; occupancy between these classes was fairly even, with class 1 containing 58% and class 2 containing 42% of these particles. The two classes were independently refined, yielding two EM density maps, further refined by applying soft masks during refinement to improve the map features. The final maps were estimated to be of 21 Å and 23 Å resolution using the gold standard 0.143 FSC criterion (classes 1 and 2, respectively) (Figure 5.10C). Though showing minor differences, these two maps are similar in both size and shape, and have a reported correlation coefficient of 0.83 using the map comparison tool in Chimera (UCSF). Notably, by comparing these maps with that of Cdk4-bound Hsp90 low-pass filtered to 20 Å (PDB entry 5FWL), it is immediately apparent that these *apo* kinase-bound Hsp90 complex maps are different to the Cdk4-bound structure, suggesting that a novel state of the kinase-Hsp90 chaperone cycle had been captured (Figure 5.10C).

At this low resolution, it is very challenging to confidently fit known protein structures into map densities. However, in each of these maps, a continuous density spanning their length has approximately the correct shape and dimensions to permit rigid docking of one protomer of Hsp90 (chain A, PDB entry 5FWL) (Figure 5.10D). In both cases, rigid docking is unable to satisfactorily place a second protomer of Hsp90 into the remaining density, indicating that conformational rearrangements of this promoter would be required. Notably, the most distinct difference between these maps of the *apo* complex and the published Cdk4-bound complex is the presence of a large density located to one of their sides. This is speculatively assigned to (at least partly) belong to the binary kinase-Cdc37 complex binding to the Hsp90 dimer (Figure 5.10D). While there is insufficient evidence at this resolution to confidently define the orientation of the complex (the single protomer of Hsp90 fits in either direction), as Cdc37 has been mapped to interact with the NTD of Hsp90 (177), it is likely that the binary complex density is found

at this location in the complex. This allows the orientation of Hsp90 to be speculatively defined (as in Figure 5.10D), and suggests that conformational rearrangements of the second Hsp90 protomer of its dimer may take place at the NTD-MD interface to enable interactions with the kinase and Cdc37. The two refined maps of this *apo* ternary complex differ most at the size and placement of the side density, possibly capturing subtly different binding modes between Hsp90 and the binary complex (Figure 5.10C). However, both suggest that the binary complex makes contacts with both the NTD and the MD of Hsp90. At this resolution, it is unclear whether these interactions are made by the kinase, Cdc37 or both components.

These interpretations of the low-resolution maps and the fact that they are nucleotide-free suggest that these capture the initial binding interaction of a binary kinase-Cdc37 complex to Hsp90. Higher resolution information, however, is necessary to confirm this suggestion and the placement of Hsp90, Cdc37 and kinase in the complex maps.

5.2.7. Negative-stain analysis of the PU-H71 bound complex

Building on the finding that the binding of ligands to the nucleotide binding pocket of Hsp90's NTD affects either the composition or shape of the kinase-bound Hsp90 complex (see 5.2.4), negative-stain EM was next used to analyse a GraFix-purified PU-H71-bound complex to explore whether these changes could be observed, and a distinction made between these two possibilities.

In this case, 2D classification of 6,915 manually-picked particles also generated classes with similar sizes and shapes, but with features which were less distinct than those observed with the *apo* complex (Figure 5.11A and B). Following dataset reduction as with the *apo* complex dataset, three-dimensional classification of 3,306 particles into two 3D classes, as before, yielded classes that each contained an approximately equal number of particles (52% and 48%, respectively). Refinement of each 3D class individually or with all 3,306 particles combined produced very similar maps, all with estimated resolutions of ~30 Å; only the latter is shown (Figure 5.11C). Compared to the *apo* complex maps, the PU-H71-bound complex map is more globular, smaller and has fewer discrete features. Notably, the side density of the *apo* map is missing (Figure 5.11C). This could be consistent with either scenario suggested to be induced by PU-H71 presence (i.e. loss of a component or a more compact complex due to ligand-induced structural rearrangements). As with the *apo* complex, and even more so here due to the maps refining to only 30 Å, higher resolution or more structural and/or biochemical restraints would be required to confidently describe the complex's composition and structure; at 30 Å resolution, no structure of Hsp90, Cdc37 nor FGFR3 kinase could be convincingly placed in the map's density.

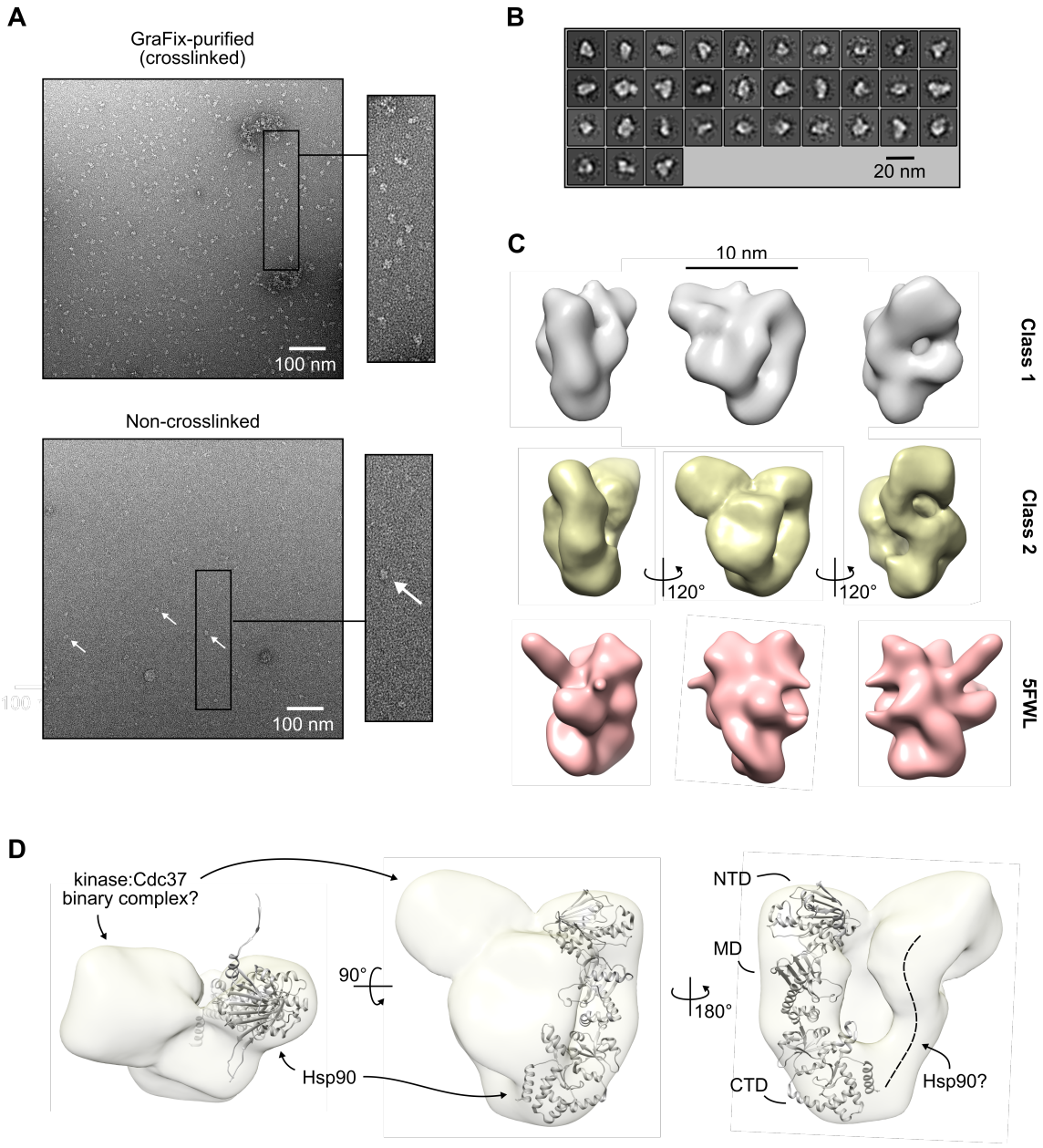


Figure 5.10. Negative-stain EM analysis of the *apo* ternary complex.
(A) Example micrographs of negatively-stained *apo* ternary complex prepared with GraFix in a crosslinked form or in non-crosslinked form as detailed in 2.13. Example particles in the micrograph of the non-crosslinked sample are highlighted by arrows. (B) As detailed in 2.13.2.2, following the manual picking of 8,470 particles of the GraFix-prepared complex, these particles were subject to rounds of 2D classification. A final dataset of 2,208 particles yielded the 2D classes shown. (C) Following *ab initio* model generation, 3D classification and refinement with soft masking, three-dimensional maps of the *apo* complex were obtained from two separate classes. Both 3D classes refined to similar resolutions: 21 and 23 Å resolution for classes 1 and 2, respectively. The map of the Cdk4-bound closed state of the ternary complex (PDB entry 5FWL) low-pass filtered to 20 Å is shown for comparison, illustrating that a different complex conformational state appears to have been obtained. (D) Speculative rigid body fitting of one Hsp90 protomer (PDB entry 5FWL) into one of the negative-stain maps of the *apo* ternary complex (class 2), with the possible locations of the second Hsp90 protomer and the kinase:Cdc37 binary complex highlighted.

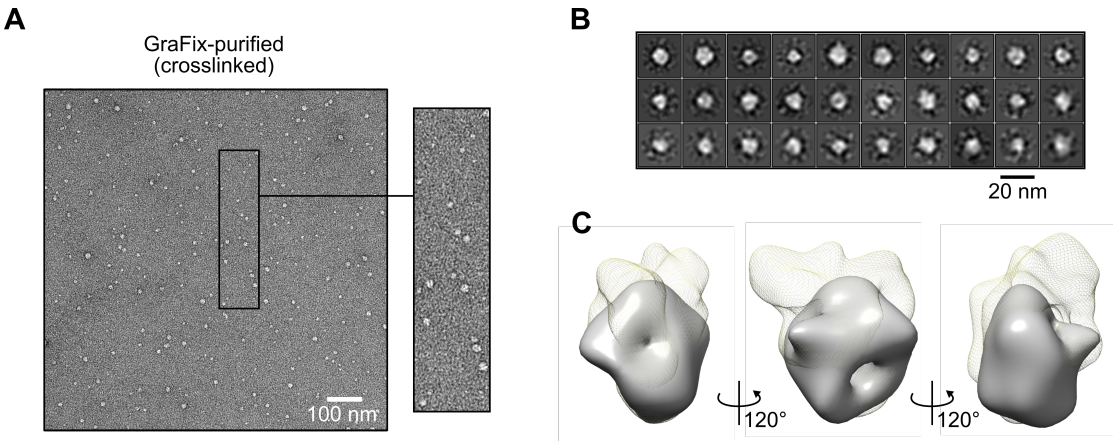


Figure 5.11. Negative-stain EM analysis of the ternary complex in the presence of FGFR inhibitor PU-H71.
(A) An example negative-stain EM micrograph of GraFix-purified ternary complex prepared in the presence of PU-H71, acquired as detailed in 2.13. (B) Two-dimensional class averages from manually picked particles of the complex. (C) Following dataset reduction as detailed in 2.13.2.2, 3,306 particles were subject to 3D classification and refinement yielding a three-dimensional map of the complex with an estimated resolution of ~30 Å. For comparison, this map (grey) is overlaid upon the refined map corresponding to 'Class 1' of the *apo* complex (Figure 5.10), shown as transparent mesh.

5.2.8. Efforts to obtain higher resolution information by cryo-EM

To improve upon the low-resolution maps obtained for the reconstituted FGFR3-bound Hsp90 complexes, samples were prepared for high-resolution structure determination by cryo-EM. Efforts were first focused on the *apo* complex state, using complexes prepared by GraFix. As buffer components such as glycerol reduce the contrast that can be obtained in cryo-EM micrographs, samples purified by GraFix required exchange into a buffer lacking glycerol prior to the preparation of cryo-EM grids. This was tried in a few ways such as through sample dialysis or through using a desalting column; however, as each of these took place before concentration of the complex using ultrafiltration units, which as detailed previously resulted in losses of at least 60% of the complex sample (see 5.2.5), it is currently unclear whether one of these methods is preferable over the other with respect to sample retention. In the future, buffer exchange could be performed concurrently with sample concentration if immobilisation onto nickel resin to perform this were optimised (Figure 5.9E).

As the concentration of *apo* ternary complex samples prepared by GraFix was unknown (due to maximal absorbance at ~265 nm rather than 280 nm, Figure 5.8D), cryo-EM grids of these samples were prepared at several dilutions. The screening of these grids found that while protein was very clearly present in the ice of the grid's holes, there did not appear to be distinct homogenous particles as was observed of the same sample by

negative-stain EM (Figure 5.12A and Figure 5.10A). The collection and processing of an overnight dataset of this GraFix-prepared sample generated 2D classes that indicated the presence of something with fairly consistent size among the classes, but that were generally poor with clear signs of particle overfitting, indicated by radial spikes in their densities (Figure 5.12B). These indicated that the data, grid and likely the sample were of poor quality. In attempt to improve and optimise grid preparation of this GraFix-purified *apo* complex for data collection, grids were next prepared using grids with different hole sizes: namely, Quantifoil® R1.2/1.3 300 mesh (as used previously), Quantifoil® R2/2 200 mesh, and Quantifoil® R2/4 400 mesh copper grids. While samples prepared with different grids can have quite different particle distributions due to differences in ice and grid properties (252), in this case, micrographs of all grids looked similar to those observed previously with no improvements in quality. As before, while protein was clearly present in the grid holes, the sample had no discrete homogenously-sized and well-distributed particles.

At this stage, given the problems that were being faced in concentration of the complex (even when it was crosslinked), and despite the observation that concentrated complexes migrated as a single species in Native-PAGE (Figure 5.9), it was possible that this concentration process using ultrafiltration units may have been spoiling the integrity of the complex, resulting in the poor quality particles observed in cryo-EM. Alternatively, or additionally, poor sample quality in cryo-EM compared to that observed by negative-stain EM of the same sample could instead arise due to challenges uniquely encountered in cryo-grid preparation such as protein adsorption and denaturation at the air-water interface. As described previously, this interface provides a very hydrophobic interaction surface for proteins in solution, and has recently gained prominence as a significant antagonist to high-resolution structural determination by cryo-EM (267). To explore these two possibilities while also preparing grids with an improved estimation of protein concentration and thereby allowing sample optimisation, following the reconstitution of an *apo* ternary complex through incubation of the individual components with one another, this reconstitution mixture was used directly in cryo-EM grid preparation without further purification. While this sample therefore contained mixed proteins and protein complexes (as observed in SEC analyses, see 5.2.2), this would allow an assessment of sample behaviour during cryo-EM grid preparation that was unaffected by possible complications related to crosslinking by GraFix or by complex sample concentration. Additionally, to investigate whether sample denaturation at the air-water interface was an issue for the *apo* ternary complex, cryo-EM grids were prepared of this sample both alone and in the presence of DDM below its CMC to block interactions of the protein complex with this hydrophobic interface (264,265).

The screening of these cryo-EM grids revealed that those prepared of the protein complex alone looked similar to those observed previously for GraFix-prepared samples (Figure 5.12C). However, when DDM was present below its CMC, discrete particles were observed in cryo-EM micrographs for the first time (Figure 5.12C). Though there was also considerable particle heterogeneity in these cryo-EM micrographs, this could be a consequence of ternary complexes not being isolated following their reconstitution and thus grid preparation using an inherently compositionally heterogeneous sample. Regardless of the origin of this heterogeneity, the observation of improved, discrete particles when DDM was present suggested that the *apo* ternary complex was indeed sensitive to the air-water interface as suspected. An overnight dataset was acquired for this sample. Though processing of this dataset generated many 2D classes showing particles of similar size and with distinct shape features (Figure 5.12D), many reminiscent of classes from negative-stain EM (Figure 5.10) and even those of the Cdk4-bound ternary complex (Figure 5.12E), these 2D classes were still of rather poor quality; it was clear that high-resolution 3D reconstructions would not be possible. Curiously, 3D classification of the dataset using an *ab initio* model as a reference and with an imposed 20 Å restriction on resolution, yielded classes that were similar to those obtained from negative-stain EM, except that density attributed to the binary complex was missing (Figure 5.12F). Currently, neither the significance nor the relevance of this is clear.

Building on the finding that the *apo* complex was sensitive to the air-water interface, efforts were focused to obtain more stable protein samples and cryo-grids better suited for high-resolution data collection using a variety of approaches. Working with the hypothesis that GraFix stabilises the complex but does not protect against interactions at the air-water interface, and drawing from improvements in grid quality when using the surfactant DDM to block this interface, grids were prepared of a GraFix-purified sample in the presence of detergents DDM or CHAPSO (264,265). Unfortunately, protein concentration was too low to observe any particles in these instances. Alternatively, grids were prepared for both *apo* and PU-H71-bound complexes prepared by GraFix using carbon-backed lacey grids, where complexes were concentrated directly on the grid through multiple applications of the sample (see 2.14.1 for details). This avoided concentration of the complex by ultrafiltration (suspected to deteriorate the quality of the sample) and provided a carbon surface for the absorption of the complex, minimising interactions with the air-water interface. Though particles were potentially visible in micrographs of a grid prepared with PU-H71 in this manner, the contrast was very poor and insufficient for particle picking. Consequently, despite efforts to optimise samples, high-resolution structural information regarding neither *apo* nor inhibitor-bound reconstituted FGFR3-bound ternary complexes has yet been obtained. A summary of grids and conditions tested and their outcomes are provided in Table 5.6.

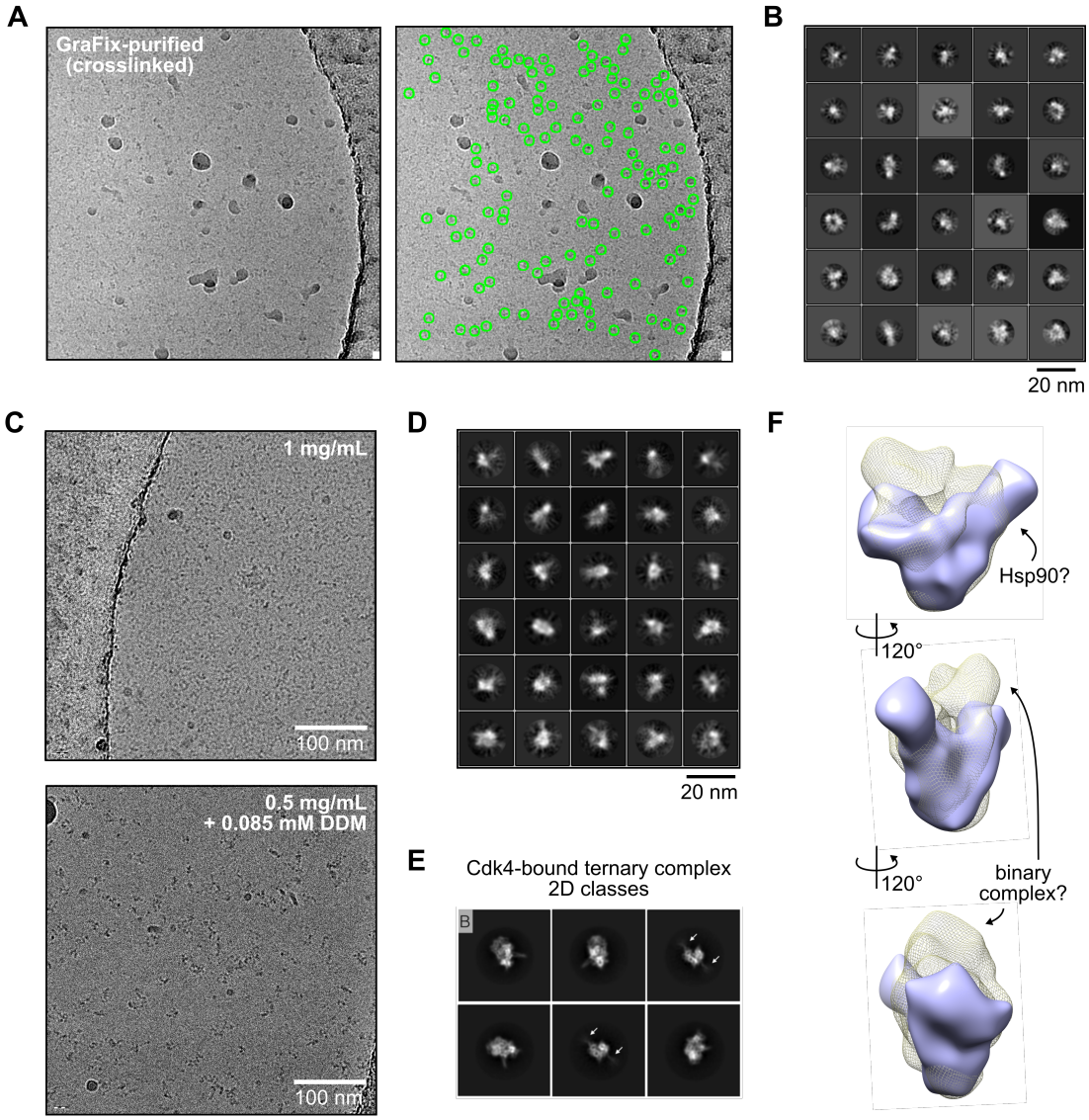


Figure 5.12. Cryo-EM analysis of the *apo* ternary complex.
(A) Example cryo-EM micrograph of the GraFix-purified *apo* complex (left) and picked particles (right) during data processing. (B) After processing as detailed in 2.14.2, 2D classification of the GraFix-purified complex yielded 2D classes showing the presence of a species with consistent size among the classes with clear shape features. However, these classes also showed indications of overfitting. (C) To assess the effect of the air-water interface on the quality of prepared cryo-grids of the complex, as detailed in the main text, grids were prepared of a reconstituted but not purified ternary complex, either alone, or in the presence of DDM below its critical micelle concentration (CMC). Example cryo-EM micrographs of the reconstituted complex alone at 1 mg/mL and at 0.5 mg/mL with 0.085 mM DDM are shown, illustrating the positive effect of blocking the air-water interface on grid quality. (D) Following collection of a cryo-EM dataset of the complex in the presence of DDM (panel C), particle picking and image processing (see 2.14.2) yielded 2D class averages with improved shape features, similar to those observed of the Cdk4-bound ternary complex (panel E, reproduced from Verba *et al*. (148)). (F) A three-dimensional map of the complex in the presence of DDM following 3D classification of particles from panel D, limiting 3D reconstruction to 20 Å resolution. The map of the *apo* complex from negative-stain EM is overlaid as a transparent mesh for comparison, illustrating apparent differences between the two maps.

Table 5.6: A summary of prepared cryo-EM grids of kinase-bound ternary complexes.

Sample	Concentration	Grid	Comment
Apo, GraFix-purified	-	Quantifoil® R1.2/1.3 Cu, 300 mesh	Proteins visible in ice but poor-quality particles See Figure 5.12A and B
Apo, GraFix-purified	-	Quantifoil® R2/2 Cu, 200 mesh	Same grid quality as with R1.2/1.3 grids
Apo, GraFix-purified	-	Quantifoil® R2/4 Cu, 400 mesh	Same grid quality as with R1.2/R1.3 grids
Apo, 1:1:1 reconstitution (no complex isolation)	1.0 mg/mL	Quantifoil® R1.2/1.3 Cu, 300 mesh	Proteins visible in ice but no distinct particles See Figure 5.12C
Apo, 1:1:1 reconstitution (no complex isolation)	0.5 mg/mL	Quantifoil® R1.2/1.3 Cu, 300 mesh	Proteins visible in ice but no distinct particles
Apo, 1:1:1 reconstitution (no complex isolation) +0.085 mM DDM	0.5 mg/mL	Quantifoil® R1.2/1.3 Cu, 300 mesh	Improved sample distribution and discrete particles observed Image processing yielded low-quality classes similar to those of the published Cdk4-bound ternary complex (148) See Figure 5.12D and F
Apo, GraFix-purified +0.085 mM DDM	-	Quantifoil® R1.2/1.3 Cu, 300 mesh	Protein concentration too low
Apo, GraFix-purified +8 mM CHAPSO	-	Quantifoil® R1.2/1.3 Cu, 300 mesh	Protein concentration too low
Apo, GraFix-purified	-	Carbon-backed Lacy Cu, 400 mesh	Contrast very low due to carbon support layer No particles observed
+PU-H71, GraFix-purified	-	Carbon-backed Lacy Cu, 400 mesh	Contrast very low due to carbon support layer A few particles potentially observed

Note that due to the fact that the crosslinked GraFix-purified samples had an absorbance maximum at 265 nm rather than 280 nm (as discussed in 5.2.5 and illustrated Figure 5.9E), the concentration of these samples used for grid preparation is unknown.

5.3. Discussion

In this chapter, to expand our understanding regarding how kinases are chaperoned by the cellular chaperone Hsp90, the processes of kinase delivery to Hsp90 and the structure and composition of kinase-bound ternary Hsp90 complexes at the early stage of its ATPase cycle were explored. This built upon prior observations that the co-chaperone Cdc37 responsible for recognition and recruitment of client kinases to Hsp90 also remodels their N-lobes, resulting in an apparent local unfolding of the kinase N-terminus (Chapter 4), but that more extensive kinase remodelling is observed in the cryo-EM structure of a Cdk4-bound ternary complex in the closed state of Hsp90 where the two kinase lobes are separated, connected by the locally-unfolded strands β4 and β5 that thread through the core of the Hsp90 complex (148). As these two cases of kinase local unfolding differ, work in this chapter aimed to improve our description of structural changes that occur to the kinase between these two snapshots of kinases in the chaperone cycle of Hsp90.

To access ternary complex states at the start of the Hsp90 chaperone cycle that describe the process of client kinase transfer to Hsp90 but which may be short-lived in the cellular environment due to high nucleotide concentrations (estimated to be in the low millimolar range for ATP in mammalian cells (350)) or the action of additional co-chaperones encouraging Hsp90 cycle progression, in this chapter, the reconstitution of complexes was employed. In these experiments, the kinase domain of the receptor tyrosine kinase FGFR3 containing an I538F substitution that improves client strength was used (322). While it is assumed that all eukaryotic kinases are processed by Hsp90 similarly given that they share the same protein fold, this is yet to be explicitly demonstrated at the structural level. Reconstitution experiments of ternary complexes using the tyrosine kinase FGFR3 (a non-model client of Hsp90) would therefore complement the existing structure of Cdk4 (a serine/threonine kinase) bound to Hsp90 (148). To reconstitute complexes, it was first necessary to express and purify the individual components of the complex. This was accomplished through recombinant expression in *E. coli*, with each protein being obtained to purities of greater than 90% following purification using IMAC and IEX chromatography (Figure 5.1, Figure 5.2, and Figure 5.3). Ternary complexes could be successfully reconstituted both in the absence of any additional ligands, and also when incubated in the presence of ATP or the Hsp90 inhibitor PU-H71, both which bind at the nucleotide binding pocket of the NTD of Hsp90 (Figure 5.4 and Figure 5.7). While it has previously been reported from studies using Cdk4 that kinase-bound Hsp90 ternary complexes cannot be reconstituted *in vitro* (148,191), complexes reconstituted in this manner have been reported for the kinase bRaf (175). Given the demonstrated ability

to reconstitute FGFR3 kinase-bound ternary complexes of Hsp90 as shown here, speculatively, the ability (or inability) to reconstitute ternary complexes *in vitro* could be kinase-specific.

5.3.1. Insights into the mechanisms of kinase chaperoning

To characterise complexes obtained from reconstitution experiments, SEC-MALS and native mass spectrometry were used to define their stoichiometry. In the case of the ternary complex reconstituted in the absence of any ligands, and therefore presumed to be in a nucleotide-free, *apo* state, these analyses showed that ternary complexes containing FGFR3 kinase were of a 1:1:2 (kinase:Cdc37:Hsp90) stoichiometry (Table 5.2, Figure 5.6 and Table 5.3), consistent with earlier findings of Cdk4 bound to Hsp90 (148,191). This suggests that tyrosine and serine/threonine kinases are indeed processed similarly by Hsp90, and that observations in studies using either of these kinase subfamilies are likely applicable to the kinase family as a whole. Interestingly, complexes with a 2:2:2 stoichiometry were also detected following reconstitution (Figure 5.6B and Table 5.3). Though previously suggested to be possible (177), these complexes had not been observed experimentally before, likely a consequence of purifying intact kinase-bound complexes following the co-expression of their components (191). While detection of this higher order complex was unexpected, it is perhaps unsurprising given that an Hsp90 dimer contains two NTDs and that Cdc37 and Hsp90 interact with one another at this location with a 1:1 stoichiometry (149,177), implying therefore that there are two possible interaction sites on a Hsp90 dimer for kinase-Cdc37 binary complexes. Whether these 2:2:2 complexes also occur in the cell or are an artefact of *in vitro* studies is unknown, but they are likely ultimately unproductive given that the closed Cdk4-bound state of Hsp90 contains only one copy each of Cdc37 and kinase (148). However, it does indicate that both NTDs of Hsp90 are able to interact with kinase-Cdc37 binary complexes, and suggests that asymmetry of Hsp90 as seen in the Cdk4-bound state (148) (and found in other Hsp90 complexes (152,157,354,355)) is not established until after binary complex binding.

To better characterise the reconstituted *apo* ternary complex, low-resolution structural information was obtained from negative-stain EM. This analysis obtained two subtly different maps of this complex at 21 to 23 Å resolution, most likely corresponding to the 1:1:2 stoichiometry complex (Figure 5.10). While the resolution obtained for each of these maps was insufficient to build a model of all three components in the complex, it undoubtedly shows a novel and previously unreported state of a kinase-bound ternary Hsp90 complex. The most striking difference between the maps of this nucleotide-free, *apo* state and the previously published Cdk4-bound structure corresponding to the closed state of the cycle (148), is the presence of a large density on one side of the

complex, speculatively attributed to the kinase-Cdc37 binary complex binding to the NTDs of Hsp90 (Figure 5.10).

Curiously, if these binary complexes can bind to either NTD in the dimer (suggested from detection of 2:2:2 complexes), as this density is observed uniquely on one side of the Hsp90 dimer, it implies that asymmetry is established in a kinase-bound ternary complex soon after binding of kinase and Cdc37, and without requiring nucleotides or other additional factors. Due to the limited resolution of these maps, the origin of this asymmetry is currently unclear, but it may reflect a specific binding interface between the binary complex and Hsp90. The low-resolution map indicates that Hsp90 is in a considerably more 'open' state than that of the Cdk4-bound structure (as the Hsp90 dimer from this structure does not fit rigidly in the low-resolution *apo* maps), and also suggests that the two protomers in the Hsp90 dimer have different conformational states, with the possible 'bending' of one protomer of Hsp90 at the NTD-MD interface (Figure 5.10D). This would likely be enabled by the charged linker connecting these two domains (Figure 5.13). Though it had not been demonstrated experimentally, a conformational state like this has previously been speculated by Verba and Agard as a means to interpret how their closed Cdk4-bound ternary complex state is established (170), obtained through the modelling of Cdc37 binding simultaneously to both kinase and Hsp90 while also preserving the Cdc37-Hsp90 interactions observed in the crystal structure of Cdc37 binding to the NTD of Hsp90 (149).

SEC analyses indicated that ligands of Hsp90 such as ATP and the inhibitor PU-H71, which both bind at the nucleotide binding pocket of the NTD, can alter Hsp90 complexes (Figure 5.7). However, data acquired in this project were unable to conclusively confirm whether complexes reconstituted in the presence of each of these ligands contains all three components of the ternary complex, or if they perhaps drive the formation of complexes containing only some (likely two) of the components. Functionally, of these two possibilities, given that ATP is required to bind to Hsp90 in order to achieve ATP hydrolysis, and that all three proteins are present in the probable post-hydrolysis state of the Cdk4-bound structure (148), it is tempting to speculate that ATP presence induces conformational rearrangements of Hsp90 and the kinase-bound complex rather than the loss of a complex component. Though it has previously been suggested from pull-down assays that Cdc37 and Hsp90 do not interact with one another in the ATP-bound state of Hsp90 (specifically when in the presence of ATP and molybdate, or the non-hydrolysable ATP analogue AMP-PNP) (356), these assays did not assess the contribution of kinase to these interactions. Similar studies using analytical ultracentrifugation (AUC) and protein homologues from *C. elegans* suggest not only that Cdc37 and Hsp90 interact with one another in the presence of AMP-PNP or the slowly-

hydrolysable analogue ATPγS, but that these ligands also affect the sedimentation of the complex compared to its nucleotide-free state, implying that they induce structural changes in the Cdc37-Hsp90 complex (357). This was interpreted as a transition between the open and closed states of Hsp90 (357), and may be analogous to changes observed in this work.

If speculating that ATP presence induces similar effects to the *apo* FGFR3-bound ternary complex, speculatively, these structural rearrangements may comprise reorganisation of the binary complex towards the MD of Hsp90 and the rearrangement of Hsp90 towards a closed, nucleotide-bound state from the 'open' state that is thought to be observed in a nucleotide-free state by negative-stain EM (Figure 5.10 and Figure 5.13). It is noted that ATP is a substrate of both Hsp90 and kinases, and thus it is very possible that at least some hydrolysis of ATP may have occurred during complex reconstitution in the experiments described here, whether through action of Hsp90 or FGFR3 kinase. The currently acquired data are unable to conclude whether ADP- or ATP-bound complexes (or a mixture of both) were obtained during these reconstitutions with ATP, and thus the interpretation of potential structural and/or compositional changes of the complex in the presence of ATP are more complicated than they appear at face value. In the future, reconstitutions in the presence of non-hydrolysable ATP analogues and with ADP itself should be completed to separate any potential different effects of each nucleotide on kinase-bound ternary complexes.

To circumvent these issues of possible ATP hydrolysis, complex reconstitutions were also completed in the presence of PU-H71, a non-hydrolysable inhibitor of Hsp90 that also binds at the nucleotide pocket of the NTD. As SEC analyses indicate that the Hsp90 inhibitor PU-H71 induces similar perturbations in the complex's elution profile as to when in the presence of ATP, it is probable that this inhibitor induces similar structural changes in the ternary complex as does nucleotides. Notably, immobilised PU-H71 has previously been shown to isolate kinase-bound complexes of Hsp90 from cells (197), implying that it stabilises ternary complexes containing kinases rather than inducing the dissociation of a component. Similarities between complexes prepared in the presence of PU-H71 or ATP may therefore suggest that nucleotides induce conformational changes in kinase-bound Hsp90 ternary complexes rather than altering their composition and inducing dissociation of a protein component. This suggestion is somewhat supported by the preliminary native MS data of the ternary complex reconstituted in the presence of PU-H71 which suggests that all three protein components are present (Figure 5.7B). However, this analysis should be repeated in the future to rationalise discrepancies in the masses of the detected species found by SEC-MALS, which did suggest the possible loss of a complex component. Though this complex was also analysed by negative-stain

EM, the low 30 Å resolution map of a GraFix-prepared complex in the presence of PU-H71 could be consistent with either the speculated structural changes or component loss, and therefore was not informative in distinguishing between these two possible outcomes (Figure 5.11).

Given that Cdc37 also inhibits the ATPase activity of Hsp90 (177), proposed from the crystal structure of M-Cdc37 bound to the NTD of Hsp90 to be mediated through preventing NTD lid closure to enclose ATP and enable NTD dimerisation (149,178), how does the Hsp90 cycle progress to a closed state? In the cell, one could imagine that contributions are provided here through the action of other (general) co-chaperones such as Aha1 which promotes ATP hydrolysis (358). However, studies are conflicting as to whether both Cdc37 and Aha1 can bind simultaneously to Hsp90. For example, while yeast Aha1, human Cdc37 and yeast Hsp90 have been found to coexist in a complex by circular dichroism (359), analyses of the same proteins using SEC suggest that Aha1 and Cdc37 compete for binding to Hsp90 (150). In the case of homologues from *C. elegans*, AUC analyses suggest that Aha1 and Cdc37 can bind simultaneously to nucleotide-free or ADP-bound Hsp90, but that Aha1 displaces Cdc37 in the presence of AMP-PNP or ATPγS (357). It is possible that these discrepancies are a consequence of studying proteins originating from different species; nonetheless, if SEC and negative-stain analyses of ligand-bound complexes in this thesis do indeed indicate that nucleotide binding at the NTD of Hsp90 induces structural rearrangements of ternary complexes, likely towards a closed state, this suggests that these changes are intrinsic to the binding of nucleotide and binary kinase-Cdc37 complexes to Hsp90 alone and does not require the action of additional co-chaperones. It should be noted that while the Cdk4-bound ternary complex is proposed to correspond to a post-hydrolysis state, given the resolution of this structure, it is also possible that the complex is in a closed ATP-bound state (148). Perhaps, therefore, these data in this thesis and the published cryo-EM ternary complex structure together suggest that progression of a kinase-bound ternary complex to an ATP-bound closed state can occur independently of other factors, with the presence of the kinase perhaps altering Cdc37-Hsp90 interactions to facilitate Hsp90 NTD dimerisation, but that progression past this closed state requires additional input, likely from Aha1, that sees Cdc37 dissociation and ATP hydrolysis (Figure 5.13). In the cell, these processes are likely organised; in fact, in the human system, a series of ordered phosphorylation events on both Cdc37 and Hsp90 have previously been found to orchestrate Cdc37 dissociation, Aha1 association and client dissociation (187,188).

5.3.2. Future work necessary to understand kinase chaperoning by Hsp90

Further structural and biochemical evidence will be required to confirm the speculations made here. In addition to obtaining higher resolution structural information for the *apo* and PU-H71-bound ternary complexes, analyses could also be performed with mutants of Hsp90 that are either unable to bind nucleotides (D93N) or that can bind ATP but cannot hydrolyse it, trapping a closed Hsp90 state (E47A) (360). Reconstitution of complexes with each of these mutants, or with wild-type Hsp90 and alternative nucleotide analogues (e.g. AMP-PNP), could allow interrogation of the nucleotide-dependency of kinase-bound Hsp90 complex composition and structure. The effects of additional Hsp90 co-chaperones such as Aha1 could also be explored *in vitro*.

While efforts were made during this project to obtain higher resolution structural information using cryo-EM, challenges faced during sample and cryo-EM grid preparation meant that this could not be achieved (see sections 5.2.5 and 5.2.8). The most successful attempt at cryo-EM analysis came from that of a non-crosslinked *apo* ternary complex. Processing of this dataset generated 2D classes, which though containing features similar to those observed by negative-stain EM, did not contain high-resolution information (Figure 5.12D). Three-dimensional classification of this dataset while imposing a 20 Å resolution limit generated reconstructions that were generally similar in size to those obtained by negative-stain EM, but which were different in shape and were missing density attributed to the binary complex (Figure 5.12F). At this stage, the source of these differences is unclear, but it could originate from the use of covalent crosslinking in preparation of the complex for negative-stain EM but not for cryo-EM (at least in this instance). This possibility will need to be explored in the future, and the relevance of this apparently missing density resolved. Higher-resolution information will be necessary to accurately describe how kinases are remodelled by Hsp90 during their chaperoning.

It is worth highlighting that though complexes containing FGFR3 kinase domain could be prepared, and that these complexes were monodisperse by SEC and Native-PAGE analyses and were tractable to structural analyses by EM and native mass spectrometry, these Hsp90 complexes have not been explicitly demonstrated to be functionally active. Although the ability of ATP and PU-H71 to induce similar changes in complex reconstitutions is indicative that Hsp90 is able to bind nucleotides, turnover of ATP has not yet been demonstrated. Given that it has previously been suggested that kinase-bound ternary Hsp90 complexes cannot be reconstituted *in vitro* (at least in the case of Cdk4 (148,191)), it is crucial that this is demonstrated for complexes reconstituted with FGFR3 to validate their functional relevance.

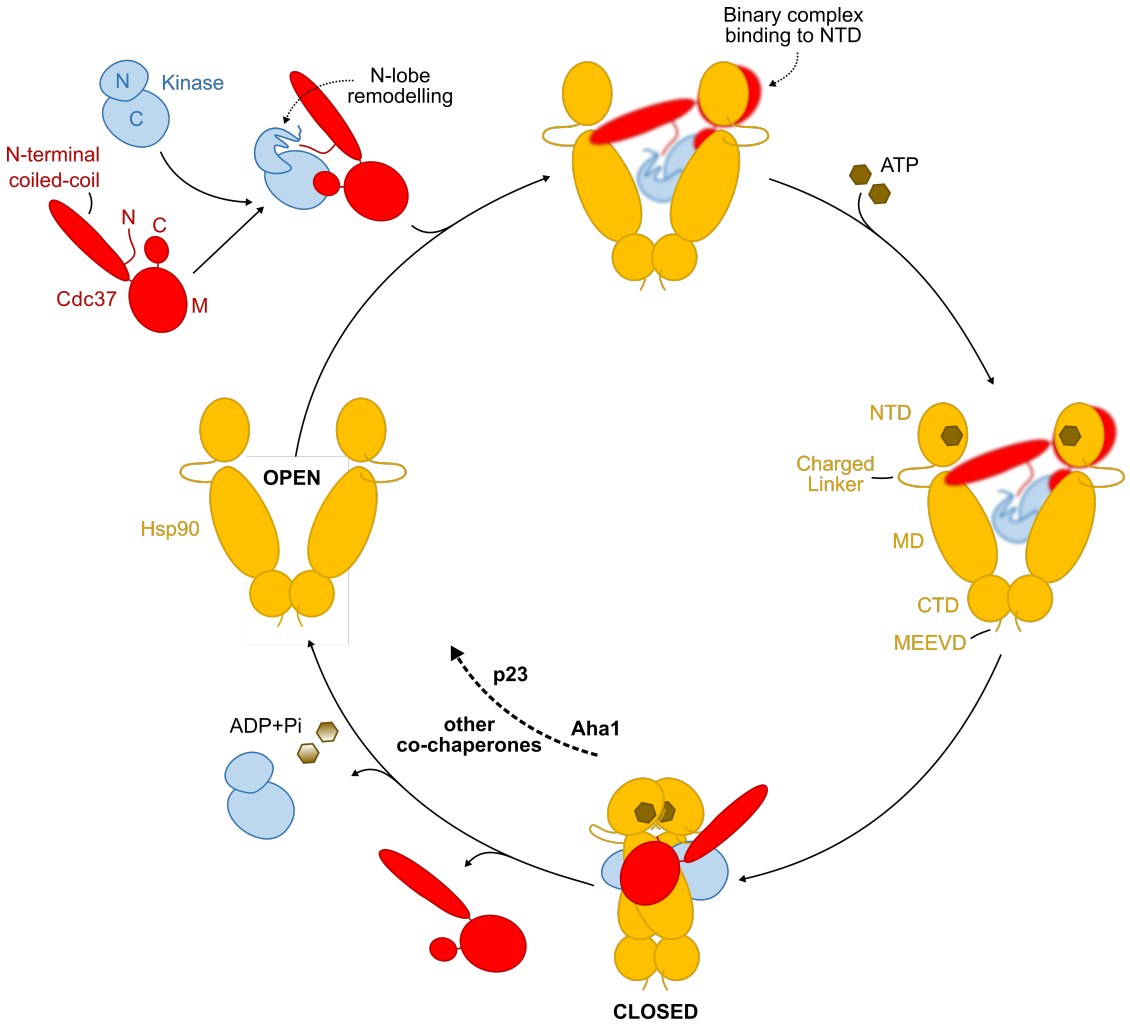


Figure 5.13. Speculative model of kinase chaperoning by Hsp90.
Following the recognition of client kinases, the co-chaperone Cdc37 remodels the N-lobe of kinases, presumably increasing its malleability to be further remodelled by Hsp90. Binary complexes of kinase-bound Cdc37 bind to the NTD of Hsp90 in the open state; this appears to occur when Hsp90 is nucleotide-free. Following ATP binding, conformational rearrangements, possibly partly mediated through co-chaperone-driven ordered processes, obtain the closed state of the kinase-bound ternary complex where the two kinase lobes are separated with local unfolding of beta strands β4 and β5 threading through the core of the complex. Further processes, orchestrated by co-chaperones, lead to Cdc37 and then kinase release. A series of ordered post-translational modifications on Hsp90 and co-chaperones likely provide the driving force for directionality in the cycle. It is possible that kinases are not released following one ATPase cycle but remain bound to Hsp90 until they can dissociate in a stable form, or are signalled for proteosomal degradation.

5.3.3. Opportunities to optimise samples for high-resolution structural studies

Through attempts to optimise grid preparation for cryo-EM, the *apo* ternary complex was found to be sensitive to the air-water interface (Figure 5.12C), implying that it presents hydrophobic features (259). When cryo-EM grids were prepared in the presence of DDM to block access to this interface, while the integrity of the sample during grid preparation was improved, data collected of this sample was still of insufficient quality for high-resolution reconstructions. It is noted that the Cdk4-bound ternary complex was also prepared in the presence of DDM, suggesting that this closed state of the kinase-bound complex also suffered from sensitivity to the air-water interface, potentially suggesting that this is a common feature among Hsp90 complexes (148). This is also consistent with the observation of poor sample recovery following concentration of both *apo* and PU-H71 bound reconstituted ternary complexes (see 5.2.5). The chemical crosslinking of complexes by GraFix, while being used successfully for the analysis of these complexes by negative-stain EM, did not resolve the difficulties faced during concentration of the complex using ultrafiltration units, nor problems faced in the preparation of high-quality cryo-EM grids.

In the future, to try and optimise the preparation of grids for high-resolution cryo-EM analyses, alternative methods developed for samples that are sensitive to the air-water interface should be explored. This could include the use of graphene-backed grids that provide a single layer of carbon for particles to adhere to (259); as possible particles were observed but with very poor contrast when using carbon-backed lacey grids which have thicker carbon layers, this approach could be worthwhile. Alternatively, additional options such as the ‘writing’ or spraying of samples on grids could be explored (267,268).

It may also be necessary to optimise the sample. In addition to probing complex formation with different ligands in attempt to stabilise the reconstituted complex, it may be worthwhile to screen the stability of the complex in a variety of buffers. Here, considerations will also have to be made to ensure the buffer is also appropriate for cryo-EM, given that additives such as glycerol and salt, which are often used to stabilise proteins, also harm image contrast (252). This effect would be particularly acute in this instance given the relatively small size of the complex at ~250 kDa. Access to many of these optimisation procedures is reliant on having a sufficient amount of protein sample at an appropriate concentration. It is thus also essential that problems surrounding the concentration of the complex is also addressed. As demonstrated (Figure 5.9E), this may be possible through immobilisation of the sample onto nickel resin using the His6 tag that is retained on Cdc37. As only the co-chaperone retains a tag, the binding of complex fractions from either SEC or GraFix to nickel resin in this manner should also isolate ternary complexes from any remaining free Hsp90 following reconstitution.

5.3.4. Alternative methods for complex characterisation

Should high-resolution information remain inaccessible, an integrated structural biology approach could be used to interpret the low-resolution negative-stain EM maps of the *apo* and ligand-bound complexes, and help build mechanistic insights regarding the chaperone cycle. This approach would also be appropriate in cases where only intermediate resolution (5 to 10 Å) information can be obtained by cryo-EM.

Additional structural information detailing these complexes could be obtained from a number of sources including by mass spectrometry approaches such as hydrogen-deuterium exchange (HDX-MS) or crosslinking mass spectrometry (361,362). These could provide restraints describing which proteins see one another in these complexes, defining protein interfaces, and help describe structural conformational changes that may occur. HDX-MS studies may be of particular interest here as they could be used in combination with the data obtained from earlier studies on the binary kinase-Cdc37 complex (Chapter 4, (322)) to investigate structural changes that occur in the kinase and co-chaperone during kinase transfer to Hsp90. NMR approaches could also be used to provide additional restraints, though due to the size of the complex, these NMR studies would require ^{13}C-methyl labelling (310,363). It is important to stress here that to obtain these additional experimental data presents its own challenges. For example, to interpret methyl-NMR data, the assignments of methyl peaks in NMR spectra are required; these can be difficult to obtain, expensive and time consuming (363). However, if available, the combination of these additional structural and biochemical restraints with known high-resolution structures of the individual components of the complex may permit the building of pseudo-atomic models of these complexes in the absence of high-resolution EM data. This approach has been used successfully in the past to describe structures of the nuclear core complex and others where high-resolution analyses were not possible (https://pdb-dev.wwpdb.org/, (364,365)). However, as components of the kinase-bound ternary complex are expected to undergo structural remodelling compared to their isolated high-resolution structures, an extensive set of restraints would be required to faithfully model these complex structures, possibly limiting the viability of this approach.

Chapter 6. Summary and Outlook

In this project, a variety of complementary biochemical and structural tools were used to investigate the mechanisms by which the receptor tyrosine kinase FGFR3 is activated and regulated. Two aspects were explored: how FGFRs are activated in context of their full-length receptors, and how kinases more generally are remodelled and regulated through their chaperoning by the ubiquitous cellular chaperone Hsp90. In both instances, better understanding of these processes is essential to appreciate their biology and allow for more informed approaches to the development of therapies targeting them.

In chapter 3, to address the current lack of high-resolution full-length structures of any RTK, efforts towards solving a high-resolution structure of full-length FGFR3 were made. Accordingly, expression of FGFR3 and a variety of oncogenic variants was established in Sf9 insect cells, yielding expression levels appropriate for structural studies. Full-length FGFR3 was also successfully partially purified and preliminary negative-stain EM micrographs analyses were completed. A number of challenges were faced during this work limiting progression to high-resolution structural studies, most notably the inability to convincingly demonstrate activity of the recombinantly expressed and purified receptor, and concerns that expression of the receptor in Sf9 insect cells produced poor quality protein. Moreover, where negative-stain EM data were obtained, these were difficult to interpret, partly due to the low molecular weight of the receptor. To address this issue, preliminary experiments were completed to work towards producing functionally-relevant yet larger and less conformationally heterogeneous receptor samples, through demonstrating expression of oncogenic FGFR3 fusions and demonstrating that FRS2α PTB domain interacts with the juxtamembrane domain of FGFR3, providing a possible route to restrict probable conformational heterogeneity of the receptor. Thus, though efforts to obtain a high-resolution structure of FGFR3 were unsuccessful within the time frame of this project, the foundations for future studies were established.

In chapters 4 and 5, the interaction of kinases with Hsp90 and its co-chaperone Cdc37 was explored, using FGFR3 kinase domain as a representative kinase. In chapter 4, NMR studies found that kinases are locally unfolded at their N-terminus upon interaction with Cdc37, and that this is likely underpinned by remodelling elsewhere in the kinase N-lobe given that the FGFR3 oncogenic mutation I538F located in the DFG-latch (which strengthens the client status of FGFR3), induces structural perturbations in the N-lobe, and that perturbations in overlapping regions are induced by PD173074 inhibitor binding (which outcompetes Cdc37). Structural remodelling of the N-lobe was later supported by HDX-MS data acquired by collaborators investigating the interaction between co-

chaperone and kinase, and this data and the NMR data detailed in the chapter could be used to define a structural model describing this client recognition complex.

Comparisons of those regions found to unfold in the binary kinase-Cdc37 client recognition complex and in the previously published ternary complex structure of Cdk4 kinase bound to Hsp90 (148), indicated that Hsp90 induces further remodelling of kinases. Therefore, in chapter 5, efforts were made to reconstitute ternary complexes of FGFR3 kinase bound to Cdc37 and Hsp90, with a focus on capturing complexes early in the Hsp90 cycle to help understand how kinases are transferred to the chaperone. Reconstitution of ternary complexes was successful, and the complex of the nucleotide-free, *apo* state was found to have a 1:1:2 (kinase:Cdc37:Hsp90) stoichiometry by SEC-MALS and native mass spectrometry, consistent with previous findings for the Ser/Thr kinase Cdk4 (148,191). Curiously, 2:2:2 complexes were also detected, capturing two kinase-Cdc37 complexes simultaneously approaching Hsp90, implying that (at least *in vitro*), binary complex binding does not itself immediately lead to kinase transfer and commit Hsp90 to an asymmetric kinase-bound complex. Negative-stain EM analysis of the *apo* complex suggests that binary complexes first approach Hsp90 at the latter's amino-terminal domains, and when forming an asymmetric complex, induces structural changes in one protomer of the Hsp90 dimer. However, the resolution of the obtained map was insufficient to place all three protein components of the complex, preventing the identification of any additional structural and mechanistic details. Reconstitutions of complexes in the presence of substrate ATP and the Hsp90 inhibitor PU-H71 indicated that structural or compositional differences are generated in Hsp90 complexes in their presence. Functionally, it would seem most appropriate that these ligands affect the structure of ternary complexes, and if so, it is speculated that their binding induces rearrangements to a more globular complex through relocation of kinase and Cdc37 components. It was hoped that high-resolution structural determination of the *apo* and ligand-bound complexes could provide supporting evidence for these proposed stages of the chaperoning cycle of kinase by Hsp90, but difficulties met in preparing protein samples and cryo-EM grids suitable for these high-resolution studies prevented this from being realised in the scope of the project.

Despite these set-backs, low-resolution snapshots of these early stages of the kinase-chaperoning cycle of Hsp90 have been captured for the first time, providing the first experimental structural insights into the processes occurring following client kinase recognition by the Hsp90 co-chaperone Cdc37. Though cryo-EM attempts to obtain high-resolution structural information of these complexes were mostly unsuccessful in the time frame of this project, progress has been made in the optimisation of these samples, and opportunities identified that could pave the way to obtaining such structures in the future.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298,** 1912–34 (2002).
2. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–34 (2010).
3. Itoh, N. & Ornitz, D. M. Evolution of the Fgf and Fgfr gene families. *Trends Genet.* **20,** 563–9 (2004).
4. Beenken, A. & Mohammadi, M. The FGF family: biology, pathophysiology and therapy. *Nat. Rev. Drug Discov.* **8,** 235–53 (2009).
5. Dorey, K. & Amaya, E. FGF signalling: diverse roles during early vertebrate embryogenesis. *Development* **137,** 3731–42 (2010).
6. Razzaque, M. S. The FGF23-Klotho axis: endocrine regulation of phosphate homeostasis. *Nat. Rev. Endocrinol.* **5,** 611–9 (2009).
7. Farrell, B. & Breeze, A. L. Structure, activation and dysregulation of fibroblast growth factor receptor kinases: perspectives for clinical targeting. *Biochem. Soc. Trans.* **46,** 1753–1770 (2018).
8. Goetz, R. & Mohammadi, M. Exploring mechanisms of FGF signalling through the lens of structural biology. *Nat. Rev. Mol. Cell Biol.* **14,** 166–80 (2013).
9. Zhang, X. *et al.* Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *J. Biol. Chem.* **281,** 15694–700 (2006).
10. Zhu, X. *et al.* Three-dimensional structures of acidic and basic fibroblast growth factors. *Science* **251,** 90–3 (1991).
11. Chen, G. *et al.* α-Klotho is a non-enzymatic molecular scaffold for FGF23 hormone signalling. *Nature* **553,** 461–466 (2018).
12. Lee, S. *et al.* Structures of β-klotho reveal a 'zip code'-like mechanism for endocrine FGF signalling. *Nature* **553,** 501–505 (2018).
13. Kuzina, E. S. *et al.* Structures of ligand-occupied β-Klotho complexes reveal a molecular mechanism underlying endocrine FGF specificity and activity. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 7819–7824 (2019).
14. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98,** 641–50 (1999).
15. Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell* **101,** 413–24 (2000).
16. Stauber, D. J., DiGabriele, A. D. & Hendrickson, W. A. Structural interactions of fibroblast growth factor receptor with its ligands. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 49–54 (2000).
17. Beenken, A., Eliseenkova, A. V, Ibrahimi, O. A., Olsen, S. K. & Mohammadi, M. Plasticity in interactions of fibroblast growth factor 1 (FGF1) N terminus with FGF receptors underlies promiscuity of FGF1. *J. Biol. Chem.* **287,** 3067–78 (2012).
18. Ibrahimi, O. A. *et al.* Structural basis for fibroblast growth factor receptor 2 activation in Apert syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7182–7 (2001).
19. Olsen, S. K. *et al.* Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain. *Genes Dev.* **20,** 185–98 (2006).
20. Olsen, S. K. *et al.* Insights into the molecular basis for fibroblast growth factor receptor autoinhibition and ligand-binding promiscuity. *Proc. Natl. Acad. Sci. U. S.*

A. **101,** 935–40 (2004).

21. Rapraeger, A. C., Krufka, A. & Olwin, B. B. Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science* **252,** 1705–8 (1991).

22. Yayon, A., Klagsbrun, M., Esko, J. D., Leder, P. & Ornitz, D. M. Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* **64,** 841–8 (1991).

23. Spivak-Kroizman, T. *et al.* Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell* **79,** 1015–24 (1994).

24. Schlessinger, J. *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **6,** 743–50 (2000).

25. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407,** 1029–34 (2000).

26. Wang, F., Kan, M., Xu, J., Yan, G. & McKeehan, W. L. Ligand-specific structural domains in the fibroblast growth factor receptor. *J. Biol. Chem.* **270,** 10222–30 (1995).

27. Chen, F. & Hristova, K. The physical basis of FGFR3 response to fgf1 and fgf2. *Biochemistry* **50,** 8576–82 (2011).

28. Miki, T. *et al.* Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 246–50 (1992).

29. Orr-Urtreger, A. *et al.* Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2). *Dev. Biol.* **158,** 475–86 (1993).

30. Liu, Y. *et al.* Regulation of receptor binding specificity of FGF9 by an autoinhibitory homodimerization. *Structure* **25,** 1325-1336.e3 (2017).

31. Yeh, B. K. *et al.* Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2266–71 (2003).

32. Harmer, N. J. *et al.* Towards a resolution of the stoichiometry of the fibroblast growth factor (FGF)-FGF receptor-heparin complex. *J. Mol. Biol.* **339,** 821–34 (2004).

33. Kiselyov, V. V, Bock, E., Berezin, V. & Poulsen, F. M. NMR structure of the first Ig module of mouse FGFR1. *Protein Sci.* **15,** 1512–5 (2006).

34. Kalinina, J. *et al.* The alternatively spliced acid box region plays a key role in FGF receptor autoinhibition. *Structure* **20,** 77–88 (2012).

35. Wang, F., Kan, M., Yan, G., Xu, J. & McKeehan, W. L. Alternately spliced NH2-terminal immunoglobulin-like Loop I in the ectodomain of the fibroblast growth factor (FGF) receptor 1 lowers affinity for both heparin and FGF-1. *J. Biol. Chem.* **270,** 10231–5 (1995).

36. Shi, E. *et al.* Control of fibroblast growth factor receptor kinase signal transduction by heterodimerization of combinatorial splice variants. *Mol. Cell. Biol.* **13,** 3907–18 (1993).

37. Sarabipour, S. & Hristova, K. Mechanism of FGF receptor dimerization and activation. *Nat. Commun.* **7,** 10262 (2016).

38. Comps-Agrar, L., Dunshee, D. R., Eaton, D. L. & Sonoda, J. Unliganded fibroblast

growth factor receptor 1 forms density-independent dimers. *J. Biol. Chem.* **290,** 24166–77 (2015).

39. Sarabipour, S. & Hristova, K. FGFR3 unliganded dimer stabilization by the juxtamembrane domain. *J. Mol. Biol.* **427,** 1705–14 (2015).

40. Maruyama, I. N. Mechanisms of activation of receptor tyrosine kinases: monomers or dimers. *Cells* **3,** 304–30 (2014).

41. Bocharov, E. V *et al.* Structure of FGFR3 transmembrane domain dimer: implications for signaling and human pathologies. *Structure* **21,** 2087–93 (2013).

42. Bae, J. H. *et al.* Asymmetric receptor contact is required for tyrosine autophosphorylation of fibroblast growth factor receptor in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2866–71 (2010).

43. Cymer, F. & Schneider, D. Transmembrane helix-helix interactions involved in ErbB receptor signaling. *Cell Adh. Migr.* **4,** 299–312 (2010).

44. Sternberg, M. J. E. & Gullick, W. J. A sequence motif in the transmembrane region of growth factor receptors with tyrosine kinase activity mediates dimerization. *Protein Eng.* **3,** 245–8 (1990).

45. Bugge, K., Lindorff-Larsen, K. & Kragelund, B. B. Understanding single-pass transmembrane receptor signaling from a structural viewpoint-what are we missing? *FEBS J.* **283,** 4424–4451 (2016).

46. Hidai, C. *et al.* FGF-1 enhanced cardiogenesis in differentiating embryonal carcinoma cell cultures, which was opposite to the effect of FGF-2. *J. Mol. Cell. Cardiol.* **35,** 421–5 (2003).

47. Mohammadi, M., Schlessinger, J. & Hubbard, S. R. Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell* **86,** 577–87 (1996).

48. Chen, H. *et al.* A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol. Cell* **27,** 717–30 (2007).

49. Huang, Z. *et al.* Structural mimicry of a-loop tyrosine phosphorylation by a pathogenic FGF receptor 3 mutation. *Structure* **21,** 1889–96 (2013).

50. Lesca, E., Lammens, A., Huber, R. & Augustin, M. Structural analysis of the human fibroblast growth factor receptor 4 kinase. *J. Mol. Biol.* **426,** 3744–3756 (2014).

51. Furdui, C. M., Lew, E. D., Schlessinger, J. & Anderson, K. S. Autophosphorylation of FGFR1 kinase is mediated by a sequential and precisely ordered reaction. *Mol. Cell* **21,** 711–7 (2006).

52. Mohammadi, M. *et al.* Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Mol. Cell. Biol.* **16,** 977–89 (1996).

53. Hou, J. *et al.* Identification of tyrosines 154 and 307 in the extracellular domain and 653 and 766 in the intracellular domain as phosphorylation sites in the heparin-binding fibroblast growth factor receptor tyrosine kinase (flg). *Protein Sci.* **2,** 86–92 (1993).

54. Peters, K. G. *et al.* Point mutation of an FGF receptor abolishes phosphatidylinositol turnover and Ca2+ flux but not mitogenesis. *Nature* **358,** 678–81 (1992).

55. Mohammadi, M. *et al.* Point mutation in FGF receptor eliminates phosphatidylinositol hydrolysis without affecting mitogenesis. *Nature* **358,** 681–4 (1992).

56. Hart, K. C., Robertson, S. C. & Donoghue, D. J. Identification of tyrosine residues in constitutively activated fibroblast growth factor receptor 3 involved in mitogenesis, Stat activation, and phosphatidylinositol 3-kinase activation. *Mol. Biol. Cell* **12,** 931–42 (2001).

57. Kong, M., Wang, C. S. & Donoghue, D. J. Interaction of fibroblast growth factor receptor 3 and the adapter protein SH2-B. A role in STAT5 activation. *J. Biol. Chem.* **277,** 15962–70 (2002).

58. Kouhara, H. *et al.* A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell* **89,** 693–702 (1997).

59. Dhalluin, C. *et al.* Structural basis of SNT PTB domain interactions with distinct neurotrophic receptors. *Mol. Cell* **6,** 921–9 (2000).

60. Hadari, Y. R., Gotoh, N., Kouhara, H., Lax, I. & Schlessinger, J. Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8578–83 (2001).

61. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. Ten. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 17783–8 (2006).

62. Hari, S. B., Merritt, E. A. & Maly, D. J. Sequence determinants of a specific inactive protein kinase conformation. *Chem. Biol.* **20,** 806–15 (2013).

63. Kornev, A. P., Taylor, S. S. & Ten Eyck, L. F. A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 14377–82 (2008).

64. Hu, J. *et al.* Kinase regulation by hydrophobic spine assembly in cancer. *Mol. Cell. Biol.* **35,** 264–76 (2015).

65. Chen, H. *et al.* Elucidation of a four-site allosteric network in fibroblast growth factor receptor tyrosine kinases. *Elife* **6,** 299–300 (2017).

66. Ornitz, D. M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip. Rev. Dev. Biol.* **4,** 215–66 (2015).

67. Deng, C. X. *et al.* Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes Dev.* **8,** 3045–57 (1994).

68. Yu, K. *et al.* Conditional inactivation of FGF receptor 2 reveals an essential role for FGF signaling in the regulation of osteoblast function and bone growth. *Development* **130,** 3063–74 (2003).

69. Scotet, E. & Houssaint, E. The choice between alternative IIIb and IIIc exons of the FGFR-3 gene is not strictly tissue-specific. *Biochim. Biophys. Acta* **1264,** 238–42 (1995).

70. Colvin, J. S., Bohne, B. A., Harding, G. W., McEwen, D. G. & Ornitz, D. M. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nat. Genet.* **12,** 390–7 (1996).

71. Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A. & Leder, P. Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell* **84,** 911–21 (1996).

72. Firnberg, N. & Neubüser, A. FGF signaling regulates expression of Tbx2, Erm, Pea3, and Pax3 in the early nasal region. *Dev. Biol.* **247,** 237–50 (2002).

73. Manning, B. D. & Cantley, L. C. AKT/PKB signaling: navigating downstream. *Cell* **129,** 1261–74 (2007).

74. Timsah, Z. *et al.* Competition between Grb2 and Plcγ1 for FGFR2 regulates basal phospholipase activity and invasion. *Nat. Struct. Mol. Biol.* **21,** 180–8 (2014).

75. Ornitz, D. M. & Marie, P. J. Fibroblast growth factor signaling in skeletal development and disease. *Genes Dev.* **29,** 1463–86 (2015).

76. Babina, I. S. & Turner, N. C. Advances and challenges in targeting FGFR signalling in cancer. *Nat. Rev. Cancer* **17,** 318–332 (2017).

77. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–8 (2007).

78. Tavormina, P. L. *et al.* Another mutation that results in the substitution of an unpaired cysteine residue in the extracellular domain of FGFR3 in thanatophoric dysplasia type I. *Hum. Mol. Genet.* **4,** 2175–7 (1995).

79. Logié, A. *et al.* Activating mutations of the tyrosine kinase receptor FGFR3 are associated with benign skin tumors in mice and humans. *Hum. Mol. Genet.* **14,** 1153–60 (2005).

80. Superti-Furga, A. *et al.* A glycine 375-to-cysteine substitution in the transmembrane domain of the fibroblast growth factor receptor-3 in a newborn with achondroplasia. *Eur. J. Pediatr.* **154,** 215–9 (1995).

81. Bellus, G. A. *et al.* Achondroplasia is defined by recurrent G380R mutations of FGFR3. *Am. J. Hum. Genet.* **56,** 368–73 (1995).

82. Bellus, G. A. *et al.* A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nat. Genet.* **10,** 357–9 (1995).

83. Meyers, G. A., Orlow, S. J., Munro, I. R., Przylepa, K. A. & Jabs, E. W. Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation in Crouzon syndrome with acanthosis nigricans. *Nat. Genet.* **11,** 462–4 (1995).

84. Patani, H. *et al.* Landscape of activating cancer mutations in FGFR kinases and their differential responses to inhibitors in clinical use. *Oncotarget* **7,** 24252–68 (2016).

85. C, G. P. D., Rajith, B. & Chakraborty, C. Predicting the impact of deleterious mutations in the protein kinase domain of FGFR2 in the context of function, structure, and pathogenesis - a bioinformatics approach. *Appl. Biochem. Biotechnol.* **170,** 1853–70 (2013).

86. Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337,** 1231–5 (2012).

87. Wu, Y.-M. *et al.* Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov.* **3,** 636–47 (2013).

88. Pal, K. *et al.* Structural Basis of TPR-Mediated Oligomerization and Activation of Oncogenic Fusion Kinases. *Structure* **25,** 867-877.e3 (2017).

89. Williams, S. V, Hurst, C. D. & Knowles, M. A. Oncogenic FGFR3 gene fusions in bladder cancer. *Hum. Mol. Genet.* **22,** 795–803 (2013).

90. Porta, R. *et al.* FGFR a promising druggable target in cancer: Molecular biology and new drugs. *Crit. Rev. Oncol. Hematol.* **113,** 256–267 (2017).

91. Chae, Y. K. *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. *Oncotarget* **8,** 16052–16074 (2017).

92. Touat, M., Ileana, E., Postel-Vinay, S., André, F. & Soria, J.-C. Targeting FGFR Signaling in Cancer. *Clin. Cancer Res.* **21,** 2684–94 (2015).

93. Klein, T. *et al.* Structural and dynamic insights into the energetics of activation loop rearrangement in FGFR1 kinase. *Nat. Commun.* **6,** 7877 (2015).

94. Tucker, J. A. *et al.* Structural insights into FGFR kinase isoform selectivity: diverse binding modes of AZD4547 and ponatinib in complex with FGFR1 and FGFR4. *Structure* **22,** 1764–1774 (2014).

95. Andre, F. *et al.* Abstract LB-145: Results of a phase I study of AZD4547, an inhibitor of fibroblast growth factor receptor (FGFR), in patients with advanced solid tumors. in *Clin. Trials* **73,** LB-145-LB-145 (American Association for Cancer Research, 2013).

96. Xie, L. *et al.* FGFR2 gene amplification in gastric cancer predicts sensitivity to the selective FGFR inhibitor AZD4547. *Clin. Cancer Res.* **19,** 2572–83 (2013).

97. Zhang, J. *et al.* Translating the therapeutic potential of AZD4547 in FGFR1-amplified non-small cell lung cancer through the use of patient-derived tumor xenograft models. *Clin. Cancer Res.* **18,** 6658–67 (2012).

98. Chell, V. *et al.* Tumour cell responses to new fibroblast growth factor receptor tyrosine kinase inhibitors and identification of a gatekeeper mutation in FGFR3 as a mechanism of acquired resistance. *Oncogene* **32,** 3059–70 (2013).

99. Trudel, S. *et al.* CHIR-258, a novel, multitargeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma. *Blood* **105,** 2941–8 (2005).

100. O'Hare, T. *et al.* AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16,** 401–12 (2009).

101. Bello, E. *et al.* E-3810 is a potent dual inhibitor of VEGFR and FGFR that exerts antitumor activity in multiple preclinical models. *Cancer Res.* **71,** 1396–405 (2011).

102. Gavine, P. R. *et al.* AZD4547: an orally bioavailable, potent, and selective inhibitor of the fibroblast growth factor receptor tyrosine kinase family. *Cancer Res.* **72,** 2045–56 (2012).

103. Nakanishi, Y. *et al.* The fibroblast growth factor receptor genetic status as a potential predictor of the sensitivity to CH5183284/Debio 1347, a novel selective FGFR inhibitor. *Mol. Cancer Ther.* **13,** 2547–58 (2014).

104. Guagnano, V. *et al.* Discovery of 3-(2,6-dichloro-3,5-dimethoxy-phenyl)-1-{6-[4-(4-ethyl-piperazin-1-yl)-phenylamino]-pyrimidin-4-yl}-1-methyl-urea (NVP-BGJ398), a potent and selective inhibitor of the fibroblast growth factor receptor family of receptor tyrosine kinase. *J. Med. Chem.* **54,** 7066–83 (2011).

105. Lamont, F. R. *et al.* Small molecule FGF receptor inhibitors block FGFR-dependent urothelial carcinoma growth in vitro and in vivo. *Br. J. Cancer* **104,** 75–82 (2011).

106. Mohammadi, M. *et al.* Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *EMBO J.* **17,** 5896–904 (1998).

107. Brameld, K. A. *et al.* Discovery of the irreversible covalent FGFR inhibitor 8-(3-(4-Acryloylpiperazin-1-yl)propyl)-6-(2,6-dichloro-3,5-dimethoxyphenyl)-2-(methylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (PRN1371) for the treatment of solid tumors. *J. Med. Chem.* **60,** 6516–6527 (2017).

108. Meric-Bernstam, F. *et al.* Efficacy of TAS-120, an irreversible fibroblast growth factor receptor (FGFR) inhibitor, in cholangiocarcinoma patients with FGFR pathway alterations who were previously treated with chemotherapy and other FGFR inhibitors. (2018). doi:10.3252/pso.eu.20wcgic.2018

109. Hoeflich, K. BLU-554, a novel, potent and selective inhibitor of FGFR4 for the treatment of liver cancer. (2015). at <http://media.ilc-congress.eu/wp-content/uploads/2015/04/Dr-Klaus-Hoeflich_Friday-Press-Conference.pdf>

110. Kim, R. *et al.* First-in-human study of BLU-554, a potent, highly selective FGFR4 inhibitor designed for hepatocellular carcinoma (HCC) with FGFR4 pathway activation. (2016). at <http://www.blueprintmedicines.com/wp-

content/uploads/2016/11/BLU-554_Kim_ENA2016_Poster.pdf>

111. Hagel, M. *et al.* First Selective Small Molecule Inhibitor of FGFR4 for the Treatment of Hepatocellular Carcinomas with an Activated FGFR4 Signaling Pathway. *Cancer Discov.* **5,** 424–37 (2015).

112. Joshi, J. J. *et al.* H3B-6527 Is a Potent and Selective Inhibitor of FGFR4 in FGF19-Driven Hepatocellular Carcinoma. *Cancer Res.* **77,** 6999–7013 (2017).

113. Bono, F. *et al.* Inhibition of tumor angiogenesis and growth by a small-molecule multi-FGF receptor blocker with allosteric properties. *Cancer Cell* **23,** 477–88 (2013).

114. Pierce, K. L. *et al.* FPA144, a humanized monoclonal antibody for both FGFR2-amplified and nonamplified, FGFR2b-overexpressing gastric cancer patients. *J. Clin. Oncol.* **32,** e15074–e15074 (2014).

115. Sommer, A. *et al.* Preclinical Efficacy of the Auristatin-Based Antibody-Drug Conjugate BAY 1187982 for the Treatment of FGFR2-Positive Solid Tumors. *Cancer Res.* **76,** 6331–6339 (2016).

116. Schatz, C. A. *et al.* Abstract 4766: Pharmacodynamic and stratification biomarker for the anti-FGFR2 antibody (BAY1179470) and the FGFR2-ADC. *Cancer Res.* **74,** 4766–4766 (2014).

117. Eli Lilly and Company. FGFR-3 Antibody-Drug Conjugate LY3076226. 1–3 (2016). at <http://lillyoncologypipeline.com/assets/pdf/fgfr_3_antibody_drug_conjugate.pdf>

118. Blackwell, C. *et al.* Inhibition of FGF/FGFR autocrine signaling in mesothelioma with the FGF ligand trap, FP-1039/GSK3052230. *Oncotarget* **7,** 39861–39871 (2016).

119. Herbert, C. *et al.* Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. *Cancer Cell* **23,** 489–501 (2013).

120. Trudel, S. *et al.* A phase I study of the safety and pharmacokinetics of escalating doses of MFGR1877S, a fibroblast growth factor receptor 3 (FGFR3) antibody, in patients with relapsed or refractory t(4;14)-positive multiple myeloma. *Blood* **120,** 4029 (2012).

121. White, S. H. Membrane proteins of known 3D structure. (2019). at <https://blanco.biomol.uci.edu/mpstruc/>

122. Gutmann, T., Kim, K. H., Grzybek, M., Walz, T. & Coskun, Ü. Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *J. Cell Biol.* **217,** 1643–1649 (2018).

123. Mi, L.-Z. *et al.* Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. *Nat. Struct. Mol. Biol.* **18,** 984–9 (2011).

124. Mi, L.-Z. *et al.* Functional and structural stability of the epidermal growth factor receptor in detergent micelles and phospholipid nanodiscs. *Biochemistry* **47,** 10314–23 (2008).

125. Chen, P.-H., Unger, V. & He, X. Structure of Full-Length Human PDGFRβ Bound to Its Activating Ligand PDGF-B as Determined by Negative-Stain Electron Microscopy. *J. Mol. Biol.* **427,** 3921–34 (2015).

126. Opatowsky, Y. *et al.* Structure, domain organization, and different conformational states of stem cell factor-induced intact KIT dimers. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 1772–7 (2014).

127. Kavran, J. M. *et al.* How IGF-1 activates its receptor. *Elife* **3,** (2014).

128. Diwanji, D., Thaker, T. & Jura, N. More than the sum of the parts: Toward full-length receptor tyrosine kinase structures. *IUBMB Life* **71,** 706–720 (2019).

129. Scapin, G. *et al.* Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis. *Nature* **556,** 122–125 (2018).

130. Weis, F. *et al.* The signalling conformation of the insulin receptor ectodomain. *Nat. Commun.* **9,** 4420 (2018).

131. Monsonego-Ornan, E., Adar, R., Feferman, T., Segev, O. & Yayon, A. The transmembrane mutation G380R in fibroblast growth factor receptor 3 uncouples ligand-mediated receptor activation from down-regulation. *Mol. Cell. Biol.* **20,** 516–22 (2000).

132. Bae, J. H. *et al.* The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. *Cell* **138,** 514–24 (2009).

133. Uchikawa, E., Choi, E., Shang, G., Yu, H. & Bai, X. Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *Elife* **8,** (2019).

134. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18,** 345–360 (2017).

135. Lindquist, S. The heat-shock response. *Annu. Rev. Biochem.* **55,** 1151–91 (1986).

136. Li, J. & Buchner, J. Structure, function and regulation of the hsp90 machinery. *Biomed. J.* **36,** 106–17 (2012).

137. Chen, B., Piel, W. H., Gui, L., Bruford, E. & Monteiro, A. The HSP90 family of genes in the human genome: insights into their divergence and evolution. *Genomics* **86,** 627–37 (2005).

138. Zhang, S. L. *et al.* Regulation of human hsp90alpha gene expression. *FEBS Lett.* **444,** 130–5 (1999).

139. Richter, K. *et al.* Conserved conformational changes in the ATPase cycle of human Hsp90. *J. Biol. Chem.* **283,** 17757–65 (2008).

140. Hessling, M., Richter, K. & Buchner, J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **16,** 287–93 (2009).

141. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11,** 515–28 (2010).

142. Colombo, G., Morra, G., Meli, M. & Verkhivker, G. Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7976–81 (2008).

143. Ali, M. M. U. *et al.* Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440,** 1013–7 (2006).

144. Prodromou, C. *et al.* The ATPase cycle of Hsp90 drives a molecular 'clamp' via transient dimerization of the N-terminal domains. *EMBO J.* **19,** 4383–92 (2000).

145. Meyer, P. *et al.* Structural and functional analysis of the middle segment of Hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol. Cell* **11,** 647–658 (2003).

146. Li, J., Soroka, J. & Buchner, J. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim. Biophys. Acta* **1823,** 624–35 (2012).

147. Oroz, J., Blair, L. J. & Zweckstetter, M. Dynamic Aha1 co-chaperone binding to

human Hsp90. *Protein Sci.* **28,** 1545–1551 (2019).

148. Verba, K. A. *et al.* Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352,** 1542–7 (2016).

149. Roe, S. M. *et al.* The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* **116,** 87–98 (2004).

150. Harst, A., Lin, H. & Obermann, W. M. J. Aha1 competes with Hop, p50 and p23 for binding to the molecular chaperone Hsp90 and contributes to kinase and hormone receptor activation. *Biochem. J.* **387,** 789–96 (2005).

151. Li, J., Richter, K., Reinstein, J. & Buchner, J. Integration of the accelerator Aha1 in the Hsp90 co-chaperone cycle. *Nat. Struct. Mol. Biol.* **20,** 326–31 (2013).

152. Li, J., Richter, K. & Buchner, J. Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nat. Struct. Mol. Biol.* **18,** 61–6 (2011).

153. Hagn, F. *et al.* Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nat. Struct. Mol. Biol.* **18,** 1086–93 (2011).

154. Echeverría, P. C., Bernthaler, A., Dupuis, P., Mayer, B. & Picard, D. An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine. *PLoS One* **6,** e26044 (2011).

155. Taipale, M. *et al.* Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* **150,** 987–1001 (2012).

156. Karagöz, G. E. *et al.* Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell* **156,** 963–74 (2014).

157. Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157,** 1685–97 (2014).

158. Rudiger, S., Freund, S. M. V, Veprintsev, D. B. & Fersht, A. R. CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11085–90 (2002).

159. Park, S. J., Borin, B. N., Martinez-Yamout, M. A. & Dyson, H. J. The client protein p53 adopts a molten globule-like state in the presence of Hsp90. *Nat. Struct. Mol. Biol.* **18,** 537–41 (2011).

160. Bohen, S. P. & Yamamoto, K. R. Isolation of Hsp90 mutants by screening for decreased steroid receptor function. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 11424–8 (1993).

161. Nathan, D. F. & Lindquist, S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* **15,** 3917–25 (1995).

162. Harris, S. F., Shiau, A. K. & Agard, D. A. The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. *Structure* **12,** 1087–97 (2004).

163. Lorenz, O. R. *et al.* Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Mol. Cell* **53,** 941–953 (2014).

164. Laederich, M. B., Degnin, C. R., Lunstrum, G. P., Holden, P. & Horton, W. A. Fibroblast growth factor receptor 3 (FGFR3) is a strong heat shock protein 90 (Hsp90) client: implications for therapeutic manipulation. *J. Biol. Chem.* **286,** 19597–604 (2011).

165. Boczek, E. E. *et al.* Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E3189-98 (2015).

166. Citri, A. *et al.* Hsp90 recognizes a common surface on client kinases. *J. Biol.*

Chem. **281,** 14361–9 (2006).

167. Prince, T., Sun, L. & Matts, R. L. Cdk2: a genuine protein kinase client of Hsp90 and Cdc37. *Biochemistry* **44,** 15287–95 (2005).

168. Terasawa, K. *et al.* Cdc37 interacts with the glycine-rich loop of Hsp90 client kinases. *Mol. Cell. Biol.* **26,** 3378–89 (2006).

169. Zhao, Q., Boschelli, F., Caplan, A. J. & Arndt, K. T. Identification of a conserved sequence motif that promotes Cdc37 and cyclin D1 binding to Cdk4. *J. Biol. Chem.* **279,** 12560–4 (2004).

170. Verba, K. A. & Agard, D. A. How Hsp90 and Cdc37 Lubricate Kinase Molecular Switches. *Trends Biochem. Sci.* **42,** 799–811 (2017).

171. Luo, Q., Boczek, E. E., Wang, Q., Buchner, J. & Kaila, V. R. I. Hsp90 dependence of a kinase is determined by its conformational landscape. *Sci. Rep.* **7,** 43996 (2017).

172. Perdew, G. H., Wiegand, H., Vanden Heuvel, J. P., Mitchell, C. & Singh, S. S. A 50 kilodalton protein associated with raf and pp60(v-src) protein kinases is a mammalian homolog of the cell cycle control protein cdc37. *Biochemistry* **36,** 3600–7 (1997).

173. Keramisanou, D. *et al.* Molecular Mechanism of Protein Kinase Recognition and Sorting by the Hsp90 Kinome-Specific Cochaperone Cdc37. *Mol. Cell* **62,** 260–271 (2016).

174. Shao, J., Irwin, A., Hartson, S. D. & Matts, R. L. Functional dissection of cdc37: characterization of domain structure and amino acid residues critical for protein kinase binding. *Biochemistry* **42,** 12577–88 (2003).

175. Polier, S. *et al.* ATP-competitive inhibitors block protein kinase recruitment to the Hsp90-Cdc37 system. *Nat. Chem. Biol.* **9,** 307–12 (2013).

176. Eckl, J. M. *et al.* Hsp90·Cdc37 Complexes with Protein Kinases Form Cooperatively with Multiple Distinct Interaction Sites. *J. Biol. Chem.* **290,** 30843–54 (2015).

177. Siligardi, G. *et al.* Regulation of Hsp90 ATPase activity by the co-chaperone Cdc37p/p50cdc37. *J. Biol. Chem.* **277,** 20151–9 (2002).

178. Sreeramulu, S. *et al.* The human Cdc37.Hsp90 complex studied by heteronuclear NMR spectroscopy. *J. Biol. Chem.* **284,** 3885–96 (2009).

179. Eckl, J. M. *et al.* Cdc37 (cell division cycle 37) restricts Hsp90 (heat shock protein 90) motility by interaction with N-terminal and middle domain binding sites. *J. Biol. Chem.* **288,** 16032–42 (2013).

180. Citri, A. *et al.* Hsp90 restrains ErbB-2/HER2 signalling by limiting heterodimer formation. *EMBO Rep.* **5,** 1165–70 (2004).

181. Eachkoti, R., Reddy, M. V. R., Lieu, Y. K., Cosenza, S. C. & Reddy, E. P. Identification and characterisation of a novel heat shock protein 90 inhibitor ONO4140. *Eur. J. Cancer* **50,** 1982–92 (2014).

182. Eckl, J. M., Daake, M., Schwartz, S. & Richter, K. Nucleotide-Free sB-Raf is Preferentially Bound by Hsp90 and Cdc37 In Vitro. *J. Mol. Biol.* **428,** 4185–4196 (2016).

183. Lachowiec, J., Lemus, T., Borenstein, E. & Queitsch, C. Hsp90 promotes kinase evolution. *Mol. Biol. Evol.* **32,** 91–9 (2015).

184. Shan, Y., Arkhipov, A., Kim, E. T., Pan, A. C. & Shaw, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7270–5 (2013).

185. Grammatikakis, N., Lin, J. H., Grammatikakis, A., Tsichlis, P. N. & Cochran, B. H. p50(cdc37) acting in concert with Hsp90 is required for Raf-1 function. *Mol. Cell. Biol.* **19,** 1661–72 (1999).

186. Hallett, S. T. *et al.* Differential Regulation of G1 CDK Complexes by the Hsp90-Cdc37 Chaperone System. *Cell Rep.* **21,** 1386–1398 (2017).

187. Xu, W. *et al.* Dynamic tyrosine phosphorylation modulates cycling of the HSP90-P50(CDC37)-AHA1 chaperone machine. *Mol. Cell* **47,** 434–43 (2012).

188. Bachman, A. B. *et al.* Phosphorylation induced cochaperone unfolding promotes kinase recruitment and client class-specific Hsp90 phosphorylation. *Nat. Commun.* **9,** 265 (2018).

189. Shao, J., Prince, T., Hartson, S. D. & Matts, R. L. Phosphorylation of serine 13 is required for the proper function of the Hsp90 co-chaperone, Cdc37. *J. Biol. Chem.* **278,** 38117–20 (2003).

190. Vaughan, C. K. *et al.* Hsp90-dependent activation of protein kinases is regulated by chaperone-targeted dephosphorylation of Cdc37. *Mol. Cell* **31,** 886–95 (2008).

191. Vaughan, C. K. *et al.* Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol. Cell* **23,** 697–707 (2006).

192. Whitesell, L. & Lindquist, S. L. HSP90 and the chaperoning of cancer. *Nat. Rev. Cancer* **5,** 761–72 (2005).

193. Trepel, J., Mollapour, M., Giaccone, G. & Neckers, L. Targeting the dynamic HSP90 complex in cancer. *Nat. Rev. Cancer* **10,** 537–549 (2010).

194. Miyata, Y., Nakamoto, H. & Neckers, L. The therapeutic target Hsp90 and cancer hallmarks. *Curr. Pharm. Des.* **19,** 347–65 (2013).

195. Trendowski, M. PU-H71: An improvement on nature's solutions to oncogenic Hsp90 addiction. *Pharmacol. Res.* **99,** 202–16 (2015).

196. NIH US National Library of Medicine. ClinicalTrials.gov. at <https://clinicaltrials.gov/>

197. Moulick, K. *et al.* Affinity-based proteomics reveal cancer-specific networks coordinated by Hsp90. *Nat. Chem. Biol.* **7,** 818–26 (2011).

198. Beebe, K. *et al.* Posttranslational modification and conformational state of heat shock protein 90 differentially affect binding of chemically diverse small molecule inhibitors. *Oncotarget* **4,** 1065–74 (2013).

199. Smith, J. R. *et al.* Restricting direct interaction of CDC37 with HSP90 does not compromise chaperoning of client proteins. *Oncogene* **34,** 15–26 (2015).

200. Sreeramulu, S., Gande, S. L., Göbel, M. & Schwalbe, H. Molecular mechanism of inhibition of the human protein complex Hsp90-Cdc37, a kinome chaperone-cochaperone, by triterpene celastrol. *Angew. Chem. Int. Ed. Engl.* **48,** 5853–5 (2009).

201. Roth, J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **102,** 285–303 (2002).

202. Freeze, H. H. & Kranz, C. Endoglycosidase and glycoamidase release of N-linked glycans. *Curr. Protoc. Mol. Biol.* **Chapter 17,** Unit 17.13A (2010).

203. Citri, A., Kochupurakkal, B. S. & Yarden, Y. The achilles heel of ErbB-2/HER2: regulation by the Hsp90 chaperone machine and potential for pharmacological intervention. *Cell Cycle* **3,** 51–60 (2004).

204. Sakagami, M., Morrison, P. & Welch, W. J. Benzoquinoid ansamycins (herbimycin A and geldanamycin) interfere with the maturation of growth factor receptor

tyrosine kinases. *Cell Stress Chaperones* **4,** 19–28 (1999).

205. Citri, A. *et al.* Drug-induced ubiquitylation and degradation of ErbB receptor tyrosine kinases: implications for cancer therapy. *EMBO J.* **21,** 2407–17 (2002).

206. Peiris, M. N., Meyer, A. N., Nelson, K. N., Bisom-Rapp, E. W. & Donoghue, D. J. Oncogenic fusion protein BCR-FGFR1 requires BCR-mediated oligomerization and chaperonin Hsp90 for activation. *Haematologica* (2019). doi:10.3324/haematol.2019.220871

207. Lamberti, D. *et al.* HSP90 inhibition drives degradation of FGFR2 fusion proteins: implications for treatment of cholangiocarcinoma. *Hepatology* **69,** 131–142 (2019).

208. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7,** 1029–38 (1998).

209. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–80 (2001).

210. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–6 (2006).

211. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47,** D520–D528 (2019).

212. Bill, R. M. *et al.* Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* **29,** 335–40 (2011).

213. Pandey, A., Shin, K., Patterson, R. E., Liu, X.-Q. & Rainey, J. K. Current strategies for protein production and purification enabling membrane protein structural biology. *Biochem. Cell Biol.* **94,** 507–527 (2016).

214. Grisshammer, R. & Tate, C. G. Overexpression of integral membrane proteins for structural studies. *Q. Rev. Biophys.* **28,** 315–422 (1995).

215. Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E. & Chisti, Y. Protein production using the baculovirus-insect cell expression system. *Biotechnol. Prog.* **30,** 1–18 (2014).

216. Zhao, L.-H. *et al.* Structure and dynamics of the active human parathyroid hormone receptor-1. *Science* **364,** 148–153 (2019).

217. Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–7 (2007).

218. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289,** 739–45 (2000).

219. Date, S. S., Fiori, M. C., Altenberg, G. A. & Jansen, M. Expression in Sf9 insect cells, purification and functional reconstitution of the human proton-coupled folate transporter (PCFT, SLC46A1). *PLoS One* **12,** e0177572 (2017).

220. Ghiglione, C. *et al.* The transmembrane molecule kekkon 1 acts in a feedback loop to negatively regulate the activity of the Drosophila EGF receptor during oogenesis. *Cell* **96,** 847–56 (1999).

221. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature* **460,** 592–8 (2009).

222. Jasti, J., Furukawa, H., Gonzales, E. B. & Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. *Nature* **449,** 316–23 (2007).

223. Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462,** 745–56 (2009).

224. Wang, F. *et al.* A homeo-interaction sequence in the ectodomain of the fibroblast growth factor receptor. *J. Biol. Chem.* **272,** 23887–95 (1997).

225. Markovic-Mueller, S. *et al.* Structure of the Full-length VEGFR-1 Extracellular Domain in Complex with VEGF-A. *Structure* **25,** 341–352 (2017).

226. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* **1666,** 105–17 (2004).

227. Stetsenko, A., Guskov, A., Stetsenko, A. & Guskov, A. An overview of the top ten detergents used for membrane protein crystallization. *Crystals* **7,** 197 (2017).

228. Smith, B. J. SDS Polyacrylamide Gel Electrophoresis of Proteins. *Methods Mol. Biol.* **1,** 41–55 (1984).

229. Chipot, C. *et al.* Perturbations of native membrane protein structure in alkyl phosphocholine detergents: a critical assessment of NMR and biophysical studies. *Chem. Rev.* **118,** 3559–3607 (2018).

230. Orwick-Rydmark, M., Arnold, T. & Linke, D. The Use of Detergents to Purify Membrane Proteins. *Curr. Protoc. protein Sci.* **84,** 4.8.1-4.8.35 (2016).

231. Anatrace. Anatrace February Newsletter: 2018 Detergent Usage Overview. (2019). at <https://www.anatrace.com/Landing/2019/Feb19-Newsletter>

232. Magnani, F., Shibata, Y., Serrano-Vega, M. J. & Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A2a receptor. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10744–9 (2008).

233. Weiss, H. M. & Grisshammer, R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur. J. Biochem.* **269,** 82–92 (2002).

234. Sonoda, Y. *et al.* Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. *FEBS Lett.* **584,** 2539–47 (2010).

235. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* **1666,** 62–87 (2004).

236. Dürr, U. H. N., Soong, R. & Ramamoorthy, A. When detergent meets bilayer: birth and coming of age of lipid bicelles. *Prog. Nucl. Magn. Reson. Spectrosc.* **69,** 1–22 (2013).

237. Denisov, I. G. & Sligar, S. G. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem. Rev.* **117,** 4669–4713 (2017).

238. Tribet, C., Audebert, R. & Popot, J. L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 15047–50 (1996).

239. Carlson, M. L. *et al.* The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *Elife* **7,** e34085 (2018).

240. Frauenfeld, J. *et al.* A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* **13,** 345–51 (2016).

241. Morrison, K. A. *et al.* Membrane protein extraction and purification using styrene-maleic acid (SMA) copolymer: effect of variations in polymer structure. *Biochem. J.* **473,** 4349–4360 (2016).

242. Postis, V. *et al.* The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim. Biophys. Acta* **1848,** 496–501 (2015).

243. Simon, K. S., Pollock, N. L. & Lee, S. C. Membrane protein nanoparticles: the shape of things to come. *Biochem. Soc. Trans.* **46,** 1495–1504 (2018).

244. Oluwole, A. O. *et al.* Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angew. Chem. Int. Ed. Engl.* **56,** 1919–1924 (2017).

245. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20,** 967–76 (2012).

246. Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem. Sci.* **37,** 343–52 (2012).

247. Heydenreich, F. M., Vuckovic, Z., Matkovic, M. & Veprintsev, D. B. Stabilization of G protein-coupled receptors by point mutations. *Front. Pharmacol.* **6,** 82 (2015).

248. Kühlbrandt, W. Biochemistry. The resolution revolution. *Science* **343,** 1443–4 (2014).

249. Cheng, Y. Membrane protein structural biology in the era of single particle cryo-EM. *Curr. Opin. Struct. Biol.* **52,** 58–63 (2018).

250. Kühlbrandt, W. Cryo-EM enters a new era. *Elife* **3,** e03678 (2014).

251. Doerr, A. Single-particle cryo-electron microscopy. *Nat. Methods* **13,** 23 (2016).

252. Drulyte, I. *et al.* Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr. Sect. D, Struct. Biol.* **74,** 560–571 (2018).

253. Tegunov, D. & Cramer, P. Real-time cryo-EM data pre-processing with Warp. *bioRxiv* 338558 (2018). doi:10.1101/338558

254. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161,** 438–449 (2015).

255. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100,** 3–15 (2016).

256. Scarff, C. A., Fuller, M. J. G., Thompson, R. F. & Iadaza, M. G. Variations on negative stain electron microscopy methods: tools for tackling challenging systems. *J. Vis. Exp.* e57199 (2018). doi:10.3791/57199

257. Dubochet, J., Groom, M. & Mueller-Neuteboom, S. in *Adv. Opt. electron Microsc.* (eds. Barer, R. & Cosslett, V. E.) 107–135 (Academic Press, London, 1982).

258. Wu, S., Armache, J.-P. & Cheng, Y. Single-particle cryo-EM data acquisition by using direct electron detection camera. *Reprod. Syst. Sex. Disord.* **65,** 35–41 (2016).

259. D'Imprima, E. *et al.* Protein denaturation at the air-water interface and how to prevent it. *Elife* **8,** e42747 (2019).

260. Noble, A. J. *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7,** e34257 (2018).

261. Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol.* **481,** 109–26 (2010).

262. Engel, C., Plitzko, J. & Cramer, P. RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat. Commun.* **7,** 12129 (2016).

263. Kang, J. Y. *et al.* Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife* **6,** e25478 (2017).

264. Chen, J., Noble, A. J., Kang, J. Y. & Darst, S. A. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: bacterial RNA polymerase and CHAPSO. *J. Struct. Biol. X* **1,** 100005 (2019).

265. Snijder, J. *et al.* An antibody targeting the fusion machinery neutralizes dual-tropic infection and defines a site of vulnerability on Epstein-Barr virus. *Immunity* **48,** 799-811.e9 (2018).

266. Dandey, V. P. *et al.* Spotiton: New features and applications. *J. Struct. Biol.* **202,** 161–169 (2018).

267. Noble, A. J. *et al.* Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat. Methods* **15,** 793–795 (2018).

268. Kontziampasis, D. *et al.* A Cryo-EM Grid Preparation Device for Time-Resolved Structural Studies. *bioRxiv* 563254 (2019). doi:10.1101/563254

269. Arnold, S. A. *et al.* Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J. Struct. Biol.* **197,** 220–226 (2017).

270. Xu, Y. *et al.* Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat. Commun.* **8,** 15741 (2017).

271. Jang, S. *et al.* Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev.* **33,** 620–625 (2019).

272. Tian, S. *et al.* Pih1p-Tah1p puts a lid on hexameric AAA+ ATPases Rvb1/2p. *Structure* **25,** 1519-1529.e4 (2017).

273. Kyrieleis, O. J. P. *et al.* Three-dimensional architecture of the human BRCA1-A histone deubiquitinase core complex. *Cell Rep.* **17,** 3099–3106 (2016).

274. Ukleja, M. *et al.* The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. *Nat. Commun.* **7,** 10433 (2016).

275. Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173,** 530–40 (2011).

276. Danev, R. & Baumeister, W. Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol.* **46,** 87–94 (2017).

277. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95,** 78–85 (2016).

278. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7,** e42166 (2018).

279. Nguyen, N. P., Gotberg, J., Ersoy, I., Bunyak, F. & White, T. DRPnet - Automated Particle Picking in Cryo-Electron Micrographs using Deep Regression. *bioRxiv* 616169 (2019). doi:10.1101/616169

280. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2,** 218 (2019).

281. Hoang, T. V, Cavin, X., Schultz, P. & Ritchie, D. W. gEMpicker: a highly parallel GPU-accelerated particle picking tool for cryo-electron microscopy. *BMC Struct. Biol.* **13,** 25 (2013).

282. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18037–41 (2013).

283. Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat. Protoc.* **14,** 100–118 (2019).

284. Pichkur, E. *et al.* Towards on-the-fly Cryo-Electron Microscopy Data Processing by High Performance Data Analysis. *J. Phys. Conf. Ser.* **955,** 012005 (2018).

285. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. Sect. D, Struct. Biol.* **73,** 496–502 (2017).

286. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–30 (2012).

287. Scheres, S. H. W. Classification of structural heterogeneity by maximum-likelihood methods. *Methods Enzymol.* **482,** 295–320 (2010).

288. Sigworth, F. J., Doerschuk, P. C., Carazo, J.-M. & Scheres, S. H. W. An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol.* **482,** 263–94 (2010).

289. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **58,** 677–89 (2015).

290. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14,** 290–296 (2017).

291. Challenges for cryo-EM. *Nat. Methods* **15,** 985–985 (2018).

292. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9,** 853–4 (2012).

293. Gauto, D. F. *et al.* Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat. Commun.* **10,** 2697 (2019).

294. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10,** 1864 (2019).

295. Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8,** 16099 (2017).

296. Fan, X. *et al.* Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat. Commun.* **10,** 2386 (2019).

297. EMBL-EBI. EMDB molecular weight distribution statistics. (2019). at <https://www.ebi.ac.uk/pdbe/emdb/statistics_mol_wt.html/>

298. Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol.* **579,** 125–57 (2016).

299. Yan, R., Zhao, X., Lei, J. & Zhou, Q. Structure of the human LAT1-4F2hc heteromeric amino acid transporter complex. *Nature* **568,** 127–130 (2019).

300. Safdari, H. A., Pandey, S., Shukla, A. K. & Dutta, S. Illuminating GPCR signaling by Cryo-EM. *Trends Cell Biol.* **28,** 591–594 (2018).

301. James, T. L. in *Nucl. Magn. Reson.* 1–31 (Bethesda: Biophysical Society, 1998). at <https://qudev.phys.ethz.ch/phys4/studentspresentations/nmr/James_Fundamentals_of_NMR.pdf>

302. Marion, D. An introduction to biological NMR spectroscopy. *Mol. Cell. Proteomics* **12,** 3006–25 (2013).

303. Hornak, J. P. The Basics of NMR. (2017). at <http://www.cis.rit.edu/htbooks/nmr/inside.htm>

304. Keeler, J. *Understanding NMR spectroscopy*. (John Wiley and Sons, 2010).

305. Lee, J. H., Okuno, Y. & Cavagnero, S. Sensitivity enhancement in solution NMR: emerging ideas and new frontiers. *J. Magn. Reson.* **241,** 18–31 (2014).

306. Takeuchi, K., Baskaran, K. & Arthanari, H. Structure determination using solution

NMR: Is it worth the effort? *J. Magn. Reson.* **306,** 195–201 (2019).

307. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73,** 1–16 (2013).

308. Markus, M. A., Dayie, K. T., Matsudaira, P. & Wagner, G. Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin 14T. *J. Magn. Reson. B* **105,** 192–5 (1994).

309. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12366–71 (1997).

310. Boswell, Z. K. & Latham, M. P. Methyl-Based NMR Spectroscopy Methods for Uncovering Structural Dynamics in Large Proteins and Protein Complexes. *Biochemistry* **58,** 144–155 (2019).

311. Jaipuria, G., Krishnarjuna, B., Mondal, S., Dubey, A. & Atreya, H. S. Amino acid selective labeling and unlabeling for protein resonance assignments. *Adv. Exp. Med. Biol.* **992,** 95–118 (2012).

312. Rasia, R. M., Brutscher, B. & Plevin, M. J. Selective isotopic unlabeling of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. *ChemBioChem* **13,** 732–739 (2012).

313. Krishnarjuna, B., Jaipuria, G., Thakur, A., D'Silva, P. & Atreya, H. S. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. *J. Biomol. NMR* **49,** 39–51 (2011).

314. Bersch, B., Groves, M. R., Klare, J. P., Torda, A. E. & Ortiz de Orué Lucana, D. in *Syst. Biol. Free Radicals Antioxidants* 505–534 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-30018-9_196

315. Alderson, R. Protein NMR spectroscopy: folded vs. unfolded protein. *Figshare* (2017). doi:10.6084/m9.figshare.5394388.v2

316. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814,** 942–68 (2011).

317. di Martino, E., L'Hôte, C. G., Kennedy, W., Tomlinson, D. C. & Knowles, M. A. Mutant fibroblast growth factor receptor 3 induces intracellular signaling and cellular transformation in a cell type- and mutation-specific manner. *Oncogene* **28,** 4306–16 (2009).

318. von Stetten, D., Noirclerc-Savoye, M., Goedhart, J., Gadella, T. W. J. & Royant, A. Structure of a fluorescent protein from *Aequorea victoria* bearing the obligate-monomer mutation A206K. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68,** 878–882 (2012).

319. Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9,** 2574–85 (2014).

320. Garzoni, F., Bieniossek, C., Berger, I., Altenfeld, A. & Musacchio, A. *The MultiBac BEVS for producing proteins and their complexes (Prot54)*. 1–20 (2012). at <https://www.epigenesys.eu/images/stories/protocols/pdf/20120313121202_p54.pdf>

321. Zeng, L., Kuti, M., Mujtaba, S. & Zhou, M.-M. Structural insights into FRS2α PTB domain recognition by neurotrophin receptor TrkB. *Proteins* **82,** 1534–41 (2014).

322. Bunney, T. D. *et al.* Disease variants of FGFR3 reveal molecular basis for the recognition and additional roles for Cdc37 in Hsp90 chaperone system. *Structure* **26,** 446-458.e8 (2018).

323. Pervushin, K. V, Wider, G. & Wüthrich, K. Single transition-to-single transition

polarization transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR* **12,** 345–8 (1998).

324. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** (1995).

325. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–96 (2005).

326. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–12 (2004).

327. Mureddu, L. & Vuister, G. W. Simple high-resolution NMR spectroscopy as a tool in molecular biology. *FEBS J.* **286,** 2035–2042 (2019).

328. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50,** 43–57 (2011).

329. Stolt-Bergner, P. *et al.* Baculovirus-driven protein expression in insect cells: A benchmarking study. *J. Struct. Biol.* 0–1 (2018). doi:10.1016/j.jsb.2018.03.004

330. Lambert, T. J. FPbase: a community-editable fluorescent protein database. *Nat. Methods* **16,** 277–278 (2019).

331. Hagihara, Y. & Saerens, D. Engineering disulfide bonds within an antibody. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 2016–2023 (2014).

332. Sarabipour, S. & Hristova, K. Pathogenic cysteine removal mutations in FGFR extracellular domains stabilize receptor dimers and perturb the TM dimer structure. *J. Mol. Biol.* **428,** 3903–3910 (2016).

333. Hauer, F. *et al.* GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. *Structure* **23,** 1769–75 (2015).

334. Gewering, T., Januliene, D., Ries, A. B. & Moeller, A. Know your detergents: A case study on detergent background in negative stain electron microscopy. *J. Struct. Biol.* **203,** 242–246 (2018).

335. Zhang, Y., McKeehan, K., Lin, Y., Zhang, J. & Wang, F. Fibroblast Growth Factor Receptor 1 (FGFR1) Tyrosine Phosphorylation Regulates Binding of FGFR Substrate 2α (FRS2α) But Not FRS2β to the Receptor. *Mol. Endocrinol.* **22,** 167–175 (2008).

336. Melillo, R. M. *et al.* Docking protein FRS2 links the protein tyrosine kinase RET and its oncogenic forms with the mitogen-activated protein kinase signaling cascade. *Mol. Cell. Biol.* **21,** 4177–87 (2001).

337. Burgar, H. R., Burns, H. D., Elsden, J. L., Lalioti, M. D. & Heath, J. K. Association of the signaling adaptor FRS2 with fibroblast growth factor receptor 1 (FGFR1) is mediated by alternative splicing of the juxtamembrane domain. *J. Biol. Chem.* **277,** 4018–4023 (2002).

338. Lievens, P. M.-J., Roncador, A. & Liboi, E. K644E/M FGFR3 mutants activate Erk1/2 from the endoplasmic reticulum through FRS2α and PLCγ-independent pathways. *J. Mol. Biol.* **357,** 783–792 (2006).

339. Yan, K. S. *et al.* FRS2 PTB Domain Conformation Regulates Interactions with Divergent Neurotrophic Receptors. *J. Biol. Chem.* **277,** 17088–17094 (2002).

340. Pelosse, M. *et al.* MultiBac: from protein complex structures to synthetic viral nanosystems. *BMC Biol.* **15,** 99 (2017).

341. He, W. *et al.* Cell-free expression of functional receptor tyrosine kinases. *Sci. Rep.* **5,** 12896 (2015).

342. Sanfelice, D. *et al.* NMR backbone assignments of the tyrosine kinase domain of human fibroblast growth factor receptor 3 in apo state and in complex with inhibitor

PD173074. *Biomol. NMR Assign.* **12,** 231–235 (2018).

343. Zhao, Q., Boschelli, F., Caplan, A. J. & Arndt, K. T. Identification of a conserved sequence motif that promotes Cdc37 and cyclin D1 binding to Cdk4. *J. Biol. Chem.* **279,** 12560–4 (2004).

344. Prince, T. & Matts, R. L. Definition of protein kinase sequence motifs that trigger high affinity binding of Hsp90 and Cdc37. *J. Biol. Chem.* **279,** 39975–81 (2004).

345. Prince, T. & Matts, R. L. Exposure of protein kinase motifs that trigger binding of Hsp90 and Cdc37. *Biochem. Biophys. Res. Commun.* **338,** 1447–54 (2005).

346. Fedorov, O. *et al.* A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 20523–8 (2007).

347. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467,** 596–9 (2010).

348. Sekhar, A. *et al.* Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. *Elife* **7,** (2018).

349. Jiang, Y. & Kalodimos, C. G. Confirmation for conformational selection. *Elife* **7,** (2018).

350. Park, J. O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* **12,** 482–9 (2016).

351. Verba, K. A. & Agard, D. A. Protein Expression and Purification of the Hsp90-Cdc37-Cdk4 Kinase Complex from Saccharomyces cerevisiae. *Bio-protocol* **7,** (2017).

352. Miyata, Y. & Nishida, E. CK2 controls multiple protein kinases by phosphorylating a kinase-targeting molecular chaperone, Cdc37. *Mol. Cell. Biol.* **24,** 4065–74 (2004).

353. Parker, C. E., Mocanu, V., Mocanu, M., Dicheva, N. & Warren, M. R. *Mass Spectrometry for Post-Translational Modifications*. *Neuroproteomics* (CRC Press/Taylor & Francis, 2010).

354. Southworth, D. R. & Agard, D. A. Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol. Cell* **42,** 771–81 (2011).

355. Retzlaff, M. *et al.* Asymmetric activation of the Hsp90 dimer by its cochaperone Aha1. *Mol. Cell* **37,** 344–54 (2010).

356. Zhang, T. *et al.* Characterization of celastrol to inhibit Hsp90 and Cdc37 interaction. *J. Biol. Chem.* **284,** 35381–9 (2009).

357. Gaiser, A. M., Kretzschmar, A. & Richter, K. Cdc37-Hsp90 complexes are responsive to nucleotide-induced conformational changes and binding of further cofactors. *J. Biol. Chem.* **285,** 40921–32 (2010).

358. Wolmarans, A., Lee, B., Spyracopoulos, L. & LaPointe, P. The mechanism of Hsp90 ATPase stimulation by Aha1. *Sci. Rep.* **6,** 33179 (2016).

359. Siligardi, G. *et al.* Co-chaperone regulation of conformational switching in the Hsp90 ATPase cycle. *J. Biol. Chem.* **279,** 51989–98 (2004).

360. Obermann, W. M. J., Sondermann, H., Russo, A. A., Pavletich, N. P. & Hartl, F. U. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. *J. Cell Biol.* **143,** 901–10 (1998).

361. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25,** 1000–1008 (2018).

362. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40,** 1224–34 (2011).

363. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.* **35,** 60–7 (2015).

364. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555,** 475–482 (2018).

365. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177,** 1384–1403 (2019).

Appendix 1: Primers

Generation of monomeric GFP by site-directed mutagenesis

Primer	Sequence
pFB_GFP-A206K_F	GACACAATCTaagCTTTCGAAAGATCC
pFB_GFP-A206K_R	GACAGGTAATGGTTGTCTG

Amplification and assembly of FGFR3-5F1Y fragments into pFastBac

Primer	Sequence
FGFR3_Frag1_X-F	GGAAACAGCGATGCCTAGG
FGFR3_Frag1_X-R	CCTCAGCCATGACCACCTGG
FGFR3_Frag2_X-F	CCAGGTGGTCATGGCTGAGG
FGFR3_Frag2_X-R	AGTACTTCTAGTCCTGCAGGAGTTC
pFB_CGVH-F	CCTGCAGGACTAGAAGTACT
pFB_CGVH-R	CCTAGGCATCGCTGTTTCC

Generation of FGFR3-5F1Y variants by site-directed mutagenesis

Primer	Sequence
FGFR3_C375Y-F	TGGTTCCGTGtacGCTGGTATCC
FGFR3_C375Y-R	GCCTCGTCAGCCTCGACC
FGFR3_E652K-F	CTACTTCAAGaaaACCACCAACGG
FGFR3_E652K-R	TCCAGGTTGTGCACGTCA
FGFR3_S249C-F	TCTCGAGCGCtgcCCTCACAGAC
FGFR3_S249C-R	ACGTCCAGAGTGTAGGTCTGAC
FGFR3_C228R-F	CAACTACACTcgcGTGGTCGAGAAC
FGFR3_C228R-R	CCACGGTCGGAAGGAACC
FGFR3_K510A-F	CGTGGCTGTGgctATGCTGAAGGAC
FGFR3_K510A-R	GTCACAGGCTTAGCAGCG
pFB_CVGH_delGFP-F	ATGCGTGGCAGCCACC
pFB_CVGH_delGFP-R	ATGTGGTCCTTGAAATAGTACTTCTAGTC

Generation of fusion constructs using ligation-independent cloning

Primer	Sequence
FGFR3_del19-F	CTACCGACCCTGCAGGACTAGAAGTACTATTTC
FGFR3_del19-R	CTGCAGGGTCGGTAGAAGTCACGGTCAG
FGFR3_Fusion-F	CCTGCAGGACTAGAAGTACTATTTC
FGFR3_Fusion-R	GTCGGTAGAAGTCACGGTCAG
FGFR3_deldel-F	GTGACTTCTACCGACGTGAGTGCTGGCTCTGGC
FGFR3_deldel-R	TTCTAGTCCTGCAGGCTCACACCTGCTCCTCAG
FGFR3_RTT112-F	GTGACTTCTACCGACGTAAAGGCGACACAGGAGG
FGFR3_RTT112-R	TTCTAGTCCTGCAGGGATCTTCTCCATCTTGGAGATGAGG
FGFR3_SW780-F	GTGACTTCTACCGACAATGTTATGGAACAGTTCAATCC
FGFR3_SW780-R	TTCTAGTCCTGCAGGTCGAATGATGGGTGCTGAGC

Correction of erroneous mutation in Cdc37 and deletion of Cdc37 from pCDF

Primer	Sequence
Cdc37_G376V-F	GAGAAGGATGtCAGTGTGCTGG
Cdc37_G376V-R	ATCGCCCGTCTTGGGAAC
Cdc37_G376V_pCDF-F	GAGAAGGATGtCAGTGTGTTCG
pCDF-Duet_delCdc37-F	TTCGAGTCTGGTAAAGAAAC
pCDF-Duet_delCdc37-R	CTACATTATGTATATCTCCTTCTTATAC

Generation of Hsp90 variants by site-directed mutagenesis

Primer	Sequence
Hsp90-E47A_F	CTTTCTGAGAgcgCTCATTTCAAATTC
Hsp90-E47A_R	ATCTCTTTGTTCGAGTAGAAAG
Hsp90-D93N_F	CACTATTGTGaacACTGGAATTGGAATG
Hsp90-D93N_R	AGAGTTCGATCTTGTTTG

Cloning of FRS2α PTB domain

Primer	Sequence
pRSET_linearise-F	TAATAAGACGTCTAAAAGGG
pRSET_linearise-R	GGGCCCCTGGAACAGAAC
FRS2a_PTB_pRSET-F	AAGTTCTGTTCCAGGGGCCCCCAGATAACCATCGGAAC
FRS2a_PTB_pRSET-R	CCCTTTTAGACGTCTTATTAATTATTTCTTTCTACAACTGGC
FRS2a_PTB_correction-F	GACACTGTCCCAGATAACCATCGGAACAAGTTTAAG
FRS2a_PTB_correction-R	TTTATCTGGGGGCCCCTGGAACAGAAC
FRS2a_PTB_NHV-F	GGCGGTAGCGACACTGTCCCAGATAACCATCG
pET28_linearise-F	CATGGTATATCTCCTTCTTAAAG
pET28_linearise-R	GGCGGTAGCCTGGAAGTTCTGTTCCAG
FRS2_PTB_CVH-F	TTCCAGGCTACCGCCATTATTTCTTTCTACAACTGGCTC
FRS2_PTB_CVH-R	AGGAGATATACCATGGACACTGTCCCAGATAACCA

Appendix 2: FASTA sequences of all constructs

Full-length FGFR3

>FGFR3-5F1Y-CVmGH

MPRGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVF
GSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHED
SGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTR
PERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQ
QWSLVMESVVPSDRGNYT**C**VVENKFGSIRQTYTLDVLER**S**PHRPILQAGLPAN
QTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKSWISE
SVEADVRLRLANVSERDGGEYLCRATNFIGVAEKAFWLSVHGPRAAEEELVEA
DEAGSV**Y**AGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKR
QVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLG
KPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAV**K**MLKDDATDKDLSDLVSEMEM
MKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDFSFDTC
KPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADF
GLARDVHNLDYFK**K**TTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFT
LGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLFMIMRECWHAAPSQRPTFK
QLVEDLDRVLTVTSTD<///>EFLDLSAPFEQFSPGGQDTPSSSSSGDDSVFAHDL
LPPAPPSSGGSRT<///>PAGLEVLFQGPHMSKGEELFTGVVPILVELDGDVNGH
KFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMK
RHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
GNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPI
GDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKTAG
SAAGSASIMRGSHHHHHHHH

Key: HRV-3C site, mGFP, His8

The sites of C228R, S249C, Y375C, K510A and K652E substitutions are bolded and underlined, and exon 19 which is replaced in fusion constructs bounded by <///>.

Full-length fusion constructs

All fusion constructs are like FGFR3-5F1Y-CVmGH except exon 19 is replaced. Only the replacement residues are given below, bounded by <///>.

>FGFR3-5F1Y-deldel-CVmGH

<///>VSAGSGLVPPAYAPPPAVPGHPAPQSAEVWGGPSGPGALADLDCSSSSQSPGS
SENQMVSPGKVSGSPEQAVEENLSSYSLDRRVTPASETLEDPCRTESQHKAETPHG
AEEECKAETPHGAEEECRHGGVCAPAAVATSPPGAIPKEACGGAPLQGLPGEALGC
PAGVGTPVPADGTQTLTCAHTSAPESTAPTNHLVAGRAMTLSPQEEVAAGQMASSS
RSGPVKLEFDVSDGATSKRAPPPRRLGERSGLKPPLRKAAVRQQKAPQEVEEDDGR
SGAGEDPPMPASRGSYHLDWDKMDDPNFIPFGGDTKSGCSEAQPPESPETRLGQP
AAEQLHAGPATEEPGPCLSQQLHSASAEDTPVVQLAAETPTAESKERALNSASTSLP
TSCPGSEPVPTHQQGQPALELKEESFRDPAEVLGTGAEVDYLEQFGTSSFKESALRK
QSLYLKFDPLLRDSPGRPVPVATETSSMHGANETPSGRPREAKLVEFDFLGALDIPVP
GPPPGVPAPGGPPLSTGPIVDLLQYSQKDLDAVVKATQEENRELRSRCE<///>

Key: FGFR3 (+33), TACC3 (144-663)

>FGFR3-5F1Y-RTT112-CVmGH

</>VKATQEENRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEVQKQKELSKAEI
QKVLKEKDQLTTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYLARITQ
EGQRYQALKAHAEEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQ
KTKENEELTRICDDLISKMEKI</>

Key: TACC3 (648-838)

>FGFR3-5F1Y-RT4-CVmGH

</>VSAGSGLVPPAYAPPPAVPGHPAPQSAEVWGGPSGPGALADLDCSSSSQSPGS
SENQMVSPGKVSGSPEQAVEENLSSYSLDRRVTPASETLEDPCRTESQHKAETPHG
AEEECKAETPHGAEEECRHGGVCAPAAVATSPPGAIPKEACGGAPLQGLPGEALGC
PAGVGTPVPADGTQTLTCAHTSAPESTAPTNHLVAGRAMTLSPQEEVAAGQMASSS
RSGPVKLEFDVSDGATSKRAPPPRRLGERSGLKPPLRKAAVRQQKAPQEVEEDDGR
SGAGEDPPMPASRGSYHLDWDKMDDPNFIPFGGDTKSGCSEAQPPESPETRLGQP
AAEQLHAGPATEEPGPCLSQQLHSASAEDTPVVQLAAETPTAESKERALNSASTSLP
TSCPGSEPVPTHQQGQPALELKEESFRDPAEVLGTGAEVDYLEQFGTSSFKESALRK
QSLYLKFDPLLRDSPGRPVPVATETSSMHGANETPSGRPREAKLVEFDFLGALDIPVP
GPPPGVPAPGGPPLSTGPIVDLLQYSQKDLDAVVKATQEENRELRSRCEELHGKNLE
LGKIMDRFEEVVYQAMEEVQKQKELSKAEIQKVLKEKDQLTTDLNSMEKSFSDLFKRF
EKQKEVIEGYRKNEESLKKCVEDYLARITQEGQRYQALKAHAEEKLQLANEEIAQVRS
KAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDDLISKMEKI</>

Key: FGFR3 (+33), TACC3 (144-838)

>FGFR3-5F1Y-SW780-CVmGH

</>NVMEQFNPGLRNLINLGKNYEKAVNAMILAGKAYYDGVAKIGEIATGSPVSTELG
HVLIEISSTHKKLNESLDENFKKFHKEIIHELEKKIELDVKYMNATLKRYQTEHKNKLESL
EKSQAELKKIRRKSQGSRNALKYEHKEIEYVETVTSRQSEIQKFIADGCKEALLEEKRR
FCFLVDKHCGFANHIHYYHLQSAELLNSKLPRWQETCVDAIKVPEKIMNMIEEIKTPAS
TPVSGTPQASPMIERSNVVRKDYDTLSKCSPKMPPAPSGRAYTSPLIDMFNNPATAA
PNSQRVNNSTGTSEDPSLQRSVSVATGLNMMKKQKVKTIFPHTAGSNKTLLSFAQGD
VITLLIPEEKDGWLYGEHDVSKARGWFPSSYTKLLEENETEAVTVPTPSPTPVRSISTV
NLSENSSVVIPPPDYLECLSMGAAADRRADSARTTSTFKAPASKPETAAPNDANGTA
KPPFLSGENPFATVKLRPTVTNDRSAPIIR</>

Key: BAIAP2L1 (18-511)

>FGFR3-5F1Y-del19-CVmGH

No additional residues

FRS2α PTB domain

>FRS2α^{N8-122}

MRGSHHHHHHGMASLEVLFQGPPDKDTVPDNHRNKFKVINVDDDGNELGSGIMELT
DTELILYTRKRDSVKWHYLCLRRYGYDSNLFSFESGRRCQTGQGIFAFKCARAEELFN
MLQEIMQNNSINVVEEPVVERNN

>FRS2α^{N9-122}

MRGSHHHHHHGMASLEVLFQGPGGSDTVPDNHRNKFKVINVDDDGNELGSGIMELT
DTELILYTRKRDSVKWHYLCLRRYGYDSNLFSFESGRRCQTGQGIFAFKCARAEELFN
MLQEIMQNNSINVVEEPVVERNN

>FRS2α^{C9-122}

MDTVPDNHRNKFKVINVDDDGNELGSGIMELTDTELILYTRKRDSVKWHYLCLRRYG
YDSNLFSFESGRRCQTGQGIFAFKCARAEELFNMLQEIMQNNSINVVEEPVVERNNG
GSLEVLFQGPLEHHHHHH

FGFR3 kinase domain

>FGFR3-KD^{I538F}

MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL
MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSELELPAD
PKWELSRARLTLGKPLGEG**A**FGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDL
SDLVSEMEMMKMIGKHKN**F**INLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLD
YSFDT**S**KPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIA
DFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLG
GSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDL
DRVLTVTSTDEYLDLSAPFE
Key: SUMO, His6

The sites of C482A, I538F and C582S substitutions present in this construct but not the wild-type kinase domain are bolded and underlined.

Human Cdc37

>Cdc37

MVDYSVWDHIEVSDDEDETHPNIDTASLFRWRHQARVERMEQFQKEKEELDRGCRE
CKRKVAECQRKLKELEVAEGGKAELERLQAEAQQLRKEERSWEQKLEEMRKKEKSM
PWNVDTLSKDGFSKSMVNTKPEKTEEDSEEVREQKHKTFVEKYEKQIKHFGMLRRW
DDSQKYLSDNVHLVCEETANYLVIWCIDLEVEEKCALMEQVAHQTIVMQFILELAKSLK
VDPRACFRQFFTKIKTADRQYMEGFNDELEAFKERVRGRAKLRIEKAMKEYEEEERK
KRLGPGGLDPVEVYESLPEELQKCFDVKDVQMLQDAISKMDPTDAKYHMQRCIDSGL
WVPNSKASEAKEGEEAGPGDPLLEAVPKTGDEKDVSVLEVLFQGPLEHHHHHH

Key: HRV 3C, His6

Human Hsp90α

>Hsp90α

MRGSHHHHHHGMASLEVLFQGPMPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIIN
TFYSNKEIFLR**E**LISNSSDALDKIRYESLTDPSKLDSGKELHINLIPNKQDRTLTIV**D**TGIG
MTKADLINNLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHND
DEQYAWESSAGGSFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFI
GYPITLFVEKERDKEVSDDEAEEKEDKEEEKEKEEKESEDKPEIEDVGSDEEEEKKDG
DKKKKKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEFYKSLTNDWEDHLAVKHFSV
EGQLEFRALLFVPRRAPFDLFENRKKKNNIKLYVRRVFIMDNCEELIPEYLNFIRGVVD
SEDLPLNISREMLQQSKILKVIRKNLVKKCLELFTELAEDKENYKKFYEQFSKNIKLGIH
EDSQNRKKLSELLRYYTSASGDEMVSLKDYCTRMKENQKHIYYITGETKDQVANSAF
VERLRKHGLEVIYMIEPIDEYCVQQLKEFEGKTLVSVTKEGLELPEDEEEKKKQEEKKT
KFENLCKIMKDILEKKVEKVVVSNRLVTSPCCIVTSTYGWTANMERIMKAQALRDNST
MGYMAAKKHLEINPDHSIIETLRQKAEADKNDKSVKDLVILLYETALLSSGFSLEDPQT
HANRIYRMIKLGLGIDEDDPTADDTSAAVTEEMPPLEGDDDTSRMEEVD

Key: HRV 3C, His6

The sites of E47A and D93N substitutions are bolded and underlined.

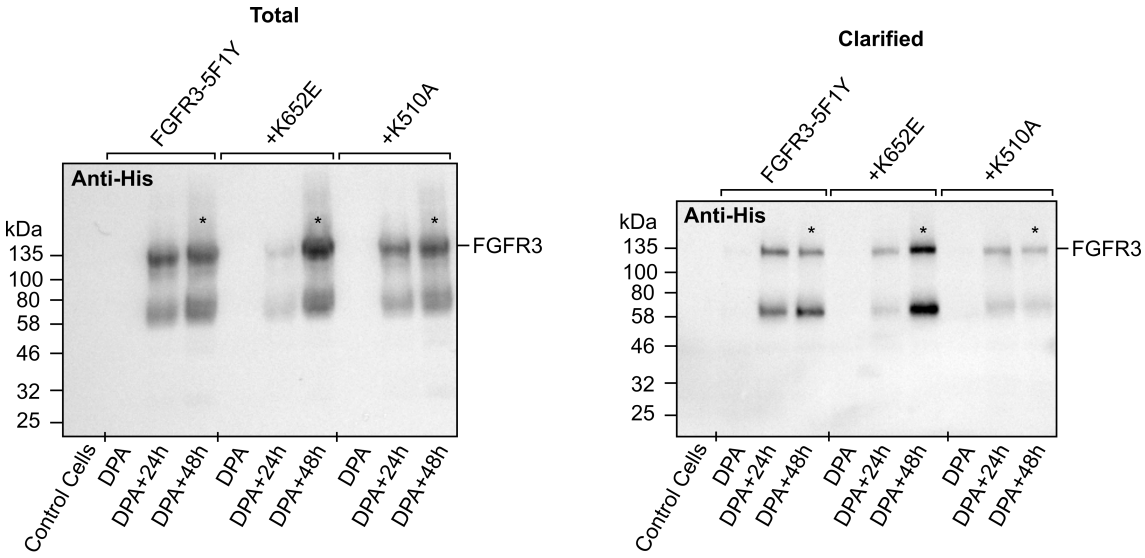
Ulp1 (SUMO protease)

>pFGET19_Ulp1

MRGSHHHHHHGLVPRGSLVPELNEKDDDQVQKALASRENTQLMNRDNIEITVRDFKT
LAPRRWLNDTIIEFFMKYIEKSTPNTVAFNSFFYTNLSERGYQGVRRWMKRKKTQIDK
LDKIFTPINLNQSHWALGIIDLKKKTIGYVDSLSNGPNAMSFAILTDLQKYVMEESKHTI
GEDFDLIHLDCPQQPNGYDCGIYVCMNTLYGSADAPLDFDYKDAIRMRRFIAHLILTDA
LK

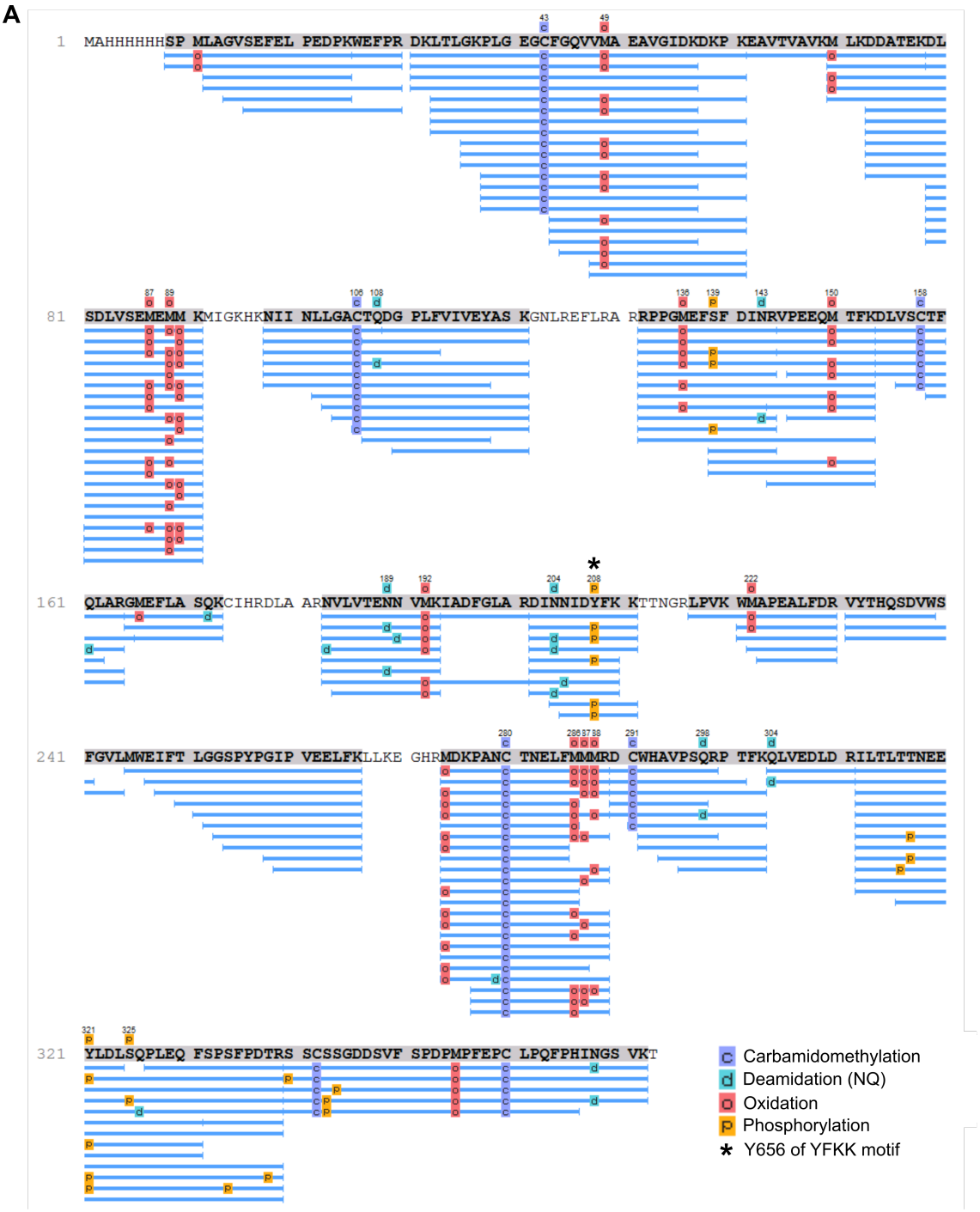
Key: His6

Appendix 3: Raw data for the calculation of relative expression levels of full-length FGFR3 receptors



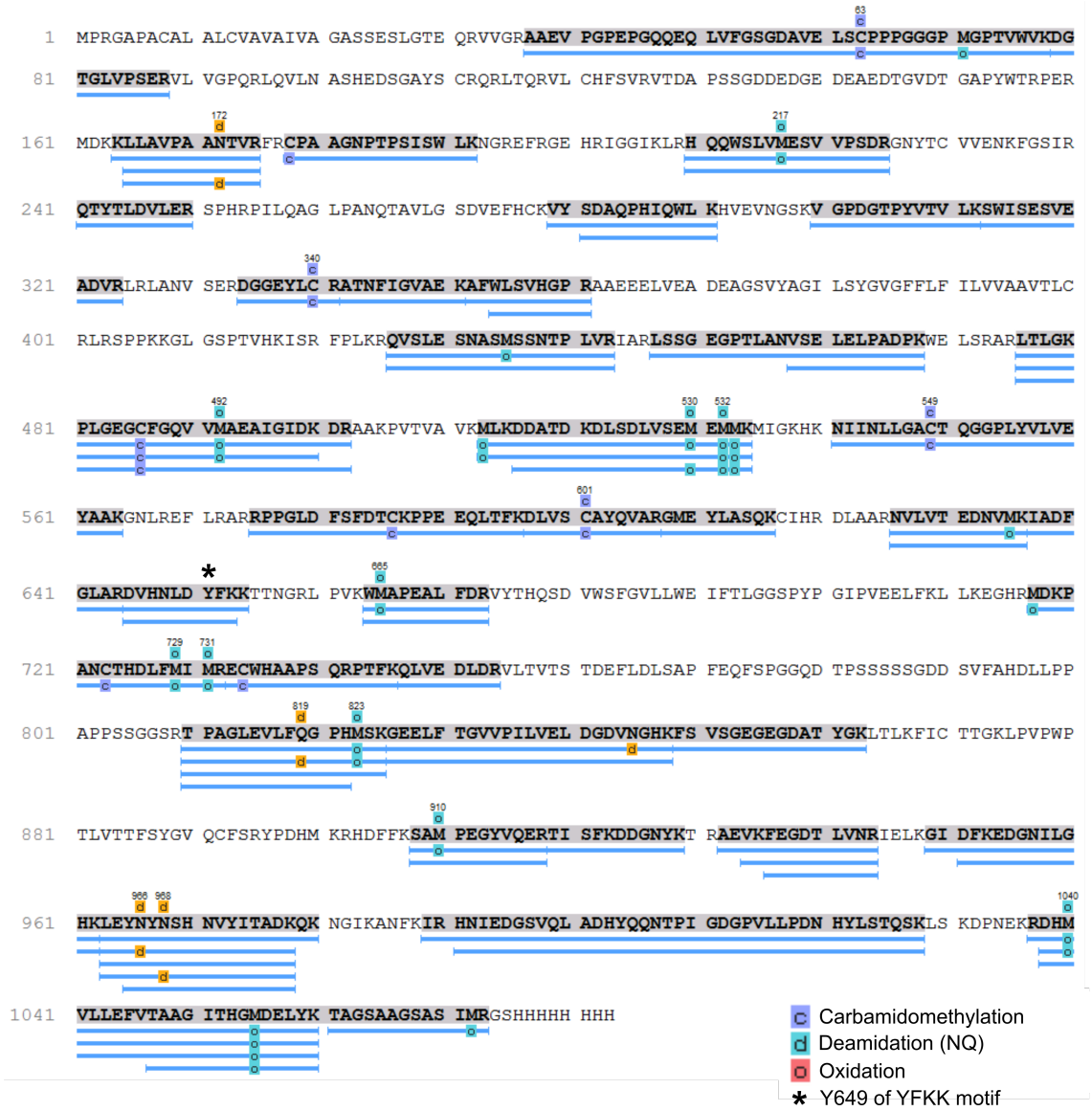
Supplementary Figure 1: Raw western blots used for the calculation of the relative expression levels of full-length FGFR3 constructs.
The relative expression level of different variants of full-length FGFR3 receptors (constitutively active (+K652E) and kinase-dead (+K510A) variants) with respect to FGFR3-5F1Y was assayed through densitometry analyses of the intensity of bands corresponding to full-length receptors following SDS-PAGE separation and anti-His western blotting. To do so, following infection of Sf9 cells with V0 viruses, samples corresponding to 1 million cells (as determined by cell counting) were collected and stored at -20 °C. Following cell lysis as detailed in 2.4.4, total and clarified cell lysates were separated by SDS-PAGE, then the His-tagged receptor visualised using anti-His western blotting. For each construct, the intensity of each full-length receptor band was determined by densitometry using ImageJ software. The relative expression level of each construct (as shown in (Figure 3.4)) was calculated from the day of highest expression for each construct with respect to FGFR3-5F1Y (as highlighted by asterisks).

Appendix 4: FGFR2-KD*



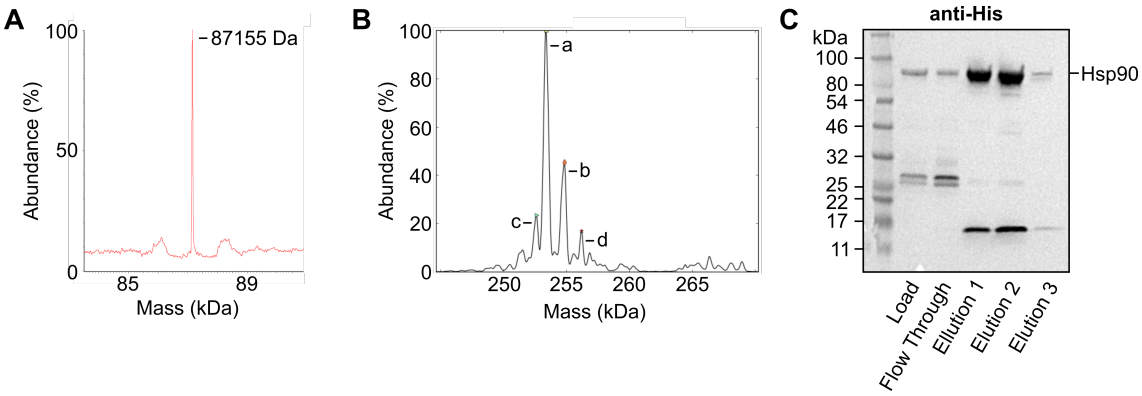
Supplementary Figure 2. Evidence of FGFR2-KD* phosphorylation.
As discussed in the main text (see 3.2.8), a variant of FGFR2 kinase domain that was phosphorylated at its YFKK motif of the activation loop (kindly provided by Sasha Evans) was used during *in vitro* autophosphorylation activity testing of FGFR3-5F1Y as a control for the anti-phosphoFGFR antibody. To demonstrate that this kinase domain (FGFR2-KD*) was phosphorylated at Y656 of its activation loop, a Coomassie-stained gel sample of this protein was sent for in-gel digestion and protein identification and post-translational modification analysis by mass spectrometry. Blue bars show peptide coverage obtained over the sequence of FGFR2-KD* (M459-T821 containing Y466F, Y561F, Y575F, Y586F, Y58F, Y608F, Y616F, Y657F, Y733F, Y779F, Y783F, Y805F and Y812F mutations, i.e. all accessible tyrosine residues except Y656 of the YYKK motif (mutated to YFKK here)), indicating that Y656 is phosphorylated.

Appendix 5: Attempted identification of FGFR3-5F1Y phosphorylation



Supplementary Figure 3. Mass spectrometry analysis to attempt to identify phosphorylation of FGFR3-5F1Y.
FGFR3-5F1Y purified in DIBMA appeared to be phosphorylated at an unidentified Tyr residue according to anti-phosphoTyr western blotting during *in vitro* activity assays of the purified receptor. To explore whether this phosphorylation modification site could be identified and therefore identify whether it was located at Y649 of the YFKK motif of the kinase activation loop, a Coomassie-stained sample of DIBMA-purified FGFR3- (corresponding to lane 1 of Figure 3.17B) was submitted for protein identification and post-translational modification identification by mass spectrometry. Blue bars indicate the obtained peptide coverage over the sequence of FGFR3-5F1Y. No phosphorylation modifications were observed, even though peptide coverage was obtained over Y649 of the YFKK motif (highlighted with an asterisk). No peptide coverage was obtained for the tyrosine residues 889 and 896 of the receptor C-terminal tail, possible alternative tyrosine phosphorylation sites.

Appendix 6: Analysis of tagged Hsp90



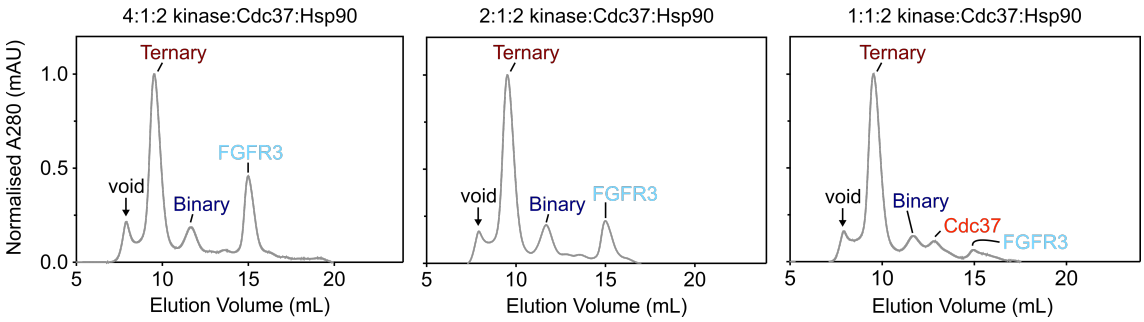
Supplementary Figure 4. Analyses of tagged Hsp90 showing protein degradation. (A) Intact mass spectrum of Hsp90 without N-terminal His6-tag removal, finding a mass 11 Da heavier than expected (87144.51 Da). (B) Native mass spectrum of a reconstituted ternary complex using tagged Hsp90, finding a group of species around the expected mass of a 1:1:2 (kinase:Cdc37:Hsp90) ternary complex, thus showing probable truncation of one of the complex components. Species are labelled as in Supplementary Table 1. (C) Anti-His western blotting of fractions from IMAC purification of Hsp90 without His6-tag removal, illustrating the presence of protein degradation.

Supplementary Table 1. Species observed by native mass spectrometry.

Species	Relative Abundance (%)	Observed Mass (Da)	Mass Difference (Da)	Annotation
a	100	253,385	− 2,810	1:1:2*
b	45	254,485	− 1,710	1:1:2*
c	23	252,635	− 3,560	1:1:2*
d	17	256,205	+ 10	1:1:2

Note: the expected mass of a 1:1:2 complex (kinase:Cdc37:Hsp90) containing tagged Hsp90 is 256,195 Da; complex species with truncations (*); species are labelled as in Supplementary Figure 4.

Appendix 7: Optimisation of ternary complex reconstitution



Supplementary Figure 5: Reconstitution of FGFR3 kinase-bound Hsp90 ternary complexes using various molar equivalents of the kinase.
Ternary complexes of FGFR3 kinase-bound Hsp90 were first reconstituted using a molar ratio of 1:1:2 (kinase:Cdc37:Hsp90) given the reported stoichiometry of a kinase-bound complex (148,191). SEC analysis of these complex reconstitution samples suggested that Hsp90 was not fully saturated by binary complexes at this ratio. Cdc37 was also not entirely consumed. In attempt to optimise complex reconstitution and saturation of Hsp90 with kinase and Cdc37, reconstitutions were repeated at molar ratios of 2:1:2 and 4:1:2 (kinase:Cdc37:Hsp90) and analysed by injection on to a Superdex 200 Increase 10/300 size exclusion chromatography column. SEC traces obtained from these injections indicated that while free Cdc37 was consumed by the presence of more kinase, excess binary kinase-Cdc37 complexes were being observed that were not being incorporated into ternary complexes with Hsp90. As an equimolar concentration of Cdc37 with respect to Hsp90 dimers was used in these reconstitutions (and therefore was not in excess), this implied that Hsp90 was not completely saturated in these reconstitutions and that free Hsp90 remained.