Structural and Computational Studies of Fibroblast Growth Factor Receptor Signalling Complex Formation

Santiago Manrique Zuñiga

Lady Margaret Hall
Nuffield Department of Clinical Medicine
Division of Structural Biology and Department of Biochemistry
University of Oxford

Trinity 2015

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford

I dedicate this thesis to my parents

Maria del Carmen & Felipe Manrique

and to my siblings

Pablo

Eduardo Moisés

Inés

Alfonso

Maria-Gracia

Andrés

Antonio

Rocío Jimena

Structural and Computational Studies of Fibroblast Growth Factor Receptor Signalling Complex Formation

Santiago Manrique Zuñiga

Abstract

Fibroblast growth factor receptor (FGFR) signalling plays a crucial role in embryogenesis, adult tissue development and tissue repair. It has been implicated in a wide range of human pathologies such as cancers and skeletal disorders which is the reason for a particular interest over recent years. Concerted efforts in academia and industry are targeting the FGFR signalling networks at different stages. The intracellular canonical downstream signalling pathways in healthy and pathogenic cells are relatively well understood. A definition of the precise mechanism for ligand induced FGFR dimerisation remains elusive. Structure biological efforts of the past 20 years have yielded a range of insightful but contradictory structures of FGFRs in complex with their ligands. It is key to solve uncertainties about the mode of extracellular and transmembrane receptor dimerisation in order to promote our abilities of targeted therapeutic manipulation of FGFR signalling in human physiology. The present study uses multiple methods from structural biology to shed more light on the effect of mutant FGFR transmembrane domains (TMDs) on signalling (Chapter 3) and on the architecture of the extracellular FGFR1 signalling complex (Chapter 4-6).

Coarse-grained molecular dynamics simulations of FGFR3 and FGFR4 TMDs have confirmed inherent propensity to form symmetric dimers mediated by Gx_3G motifs. Interestingly, disease related mutations do not map on the primary TMD dimer interfaces. Simulations of selected mutant TMDs show no significant effect on receptor dimerisation propensity. Atomistic simulations of the transmembrane FGFR3 Achondroplasia mutant reveal a significant shift of the dimer relative to the bilayer to allow interactions of the mutant Arg380 side chain with lipid head group phosphates. This finding supports mechanisms of signal transduction by conformational coupling across the membrane.

Biophysical characterisation of the FGFR signalling complex was performed to work out a co-crystallisation strategy using receptors expressed in mammalian cells. Crystallographic experiments yielded a FGFR1 structure bound to its ligands, revealing a full signalling complex with an architecture that contradicts two well-established, crystal structure-derived models. The presented structure features a central, decasaccharide-mediated Fibroblast Growth Factor dimer, which recruits a receptor on each end to form a quasi-symmetric pentameric signalling complex. Further biophysical validation makes the newly presented structure the first model that is supported by structural data from the complex in solution.

Acknowledgements

First and foremost I would like to thank my supervisors Professor E. Yvonne Jones and Professor Mark S. P. Sansom for their guidance and encouragement through the challenging past four years. Each January, Yvonne and Mark form part of the Wellcome Trust Structural Biology DPhil Programme interview panel. In January 2011 I was lucky enough to be interviewed by them and to be offered the Wellcome Trust scholarship. I was even luckier still that they both agreed to jointly supervise my DPhil project.

Secondly, I would like to thank all members in the Division of Structural Biology for a collaborative atmosphere and particularly those people who have directly been involved with my doctoral studies: Dr. Charlotte Coles for supervision in my first year at strubi, for kick-starting the FGFR project and for on-going useful discussions. Dr Cédric Leyrat for help with SAXS data analysis and discussions, Prof. Radu Aricescu for initiation of this project and on-going discussions, Prof. Christian Siebold for help with X-ray data interpretation, Dr. Karl Harlos for crystal handling and help with collection of X-ray diffraction data, Youxin Kong for help with cell culture and fluorescence microscopy, Dr Sergi Padialla-Parra for helping with cell imaging experiments, Weixian Lu and Dr. Yuguang Zhao for help with tissue culture. Furthermore I thank Dr Elena Seiradake, Dr Matthias Zebisch, Dr Jonahan Elegheert, Dr Tom Walter and all other members in Yvonne's lab.

I thank all members in the Division of Structural Bioinformatics & Computational Biochemistry, in particular those people who I had the privilege to collaborate with and be trained by: Dr. Tyler Reddy, Dr Antreas Kalli, Dr Heidi Koldsø, Dr Sarah Rouse and last but not least Amanda Buyan for computational expertise, chats and coffees.

I thank the Wellcome Trust for generous funding, as well as Lady Margaret Hall and the University of Oxford for their pastoral care.

Hearty thanks to Christelle & Louis van As with Liam and Gabrielle for their warm care, companionship, support in so incredibly many ways – simply for being amazing and becoming like family. The times with you will forever form part of my best memories.

Finally, I would like to thank my siblings Pablo, Eduardo, Inés, Alfonso, Maria-Gracia & Stephen, Andrés, Antonio & Crystal, Rocío & Marc, You and my parents Maria del Carmen (!) and Felipe (!) for all your love, encouragement and support in thoughts and prayers throughout my time in England and in times to come.

(Yes! My DPhil really happened. Not working for the CIA…)

Table of Contents

ABSTRACT 1

ACKNOWLEDGEMENTS 2

TABLE OF CONTENTS 3

CHAPTER 1 – INTRODUCTION 10

1.1 BACKGROUND 10

1.2 AN OVERVIEW OF THE RECEPTOR TYROSINE KINASE FAMILY 11

1.3 FGFR SIGNALLING 14

1.3.1 FIBROBLAST GROWTH FACTOR RECEPTORS 14

1.3.2 FIBROBLAST GROWTH FACTORS 15

1.3.3 HEPARAN SULPHATE PROTEOGLYCANS 17

1.3.4 FGFR SIGNALLING PATHWAYS 18

1.4 STRUCTURAL BIOLOGY OF THE EXTRACELLULAR FGFR SIGNALLING COMPLEX 20

1.4.1 THE BINARY FGFR–FGF COMPLEX 21

1.4.2 THE TERNARY FGFR–HEPARIN–FGF SIGNALLING COMPLEX 21

1.5 DYSREGULATION OF FGFR SIGNALLING IN HUMAN MALIGNANCIES 24

1.5.1 MUTATIONS IN THE EXTRACELLULAR DOMAIN 25

1.5.2 ACTIVATING MUTATIONS IN THE TRANSMEMBRANE DOMAIN 27

1.5.3 STRATEGIES FOR THERAPY 28

1.6 AIMS 29

CHAPTER 2 – MATERIAL AND METHODS 31

2.1 MOLECULAR DYNAMICS 31

2.1.1 COARSE GRAINED SIMULATIONS 31

2.1.2 ATOMISTIC SIMULATIONS 32

2.2 MOLECULAR BIOLOGY 33
2.2.1 COMPLEMENTARY DNA AND EXPRESSION VECTORS 33
2.2.2 CONSTRUCT DESIGN AND POLYMERASE CHAIN REACTION 35
2.2.3 DNA RESTRICTION, DIGESTION, LIGATION AND TRANSFORMATION 36
2.3 PROTEIN EXPRESSION AND PURIFICATION 37
2.3.1 PROTEIN EXPRESSION IN *E.COLI* 37
2.3.2 SMALL-SCALE PROTEIN EXPRESSION IN HEK293S CELLS AND WESTERN BLOT ANALYSIS 38
2.3.3 LARGE-SCALE EXPRESSION IN HEK293S CELLS 39
2.4 PROTEIN PURIFICATION 40
2.4.1 IMMOBILISED METAL AFFINITY CHROMATOGRAPHY 40
2.4.2 ANALYTICAL HEPARIN AFFINITY CHROMATOGRAPHY 40
2.4.3 SIZE EXCLUSION CHROMATOGRAPHY 41
2.5 MULTI-ANGLE LASER LIGHT SCATTERING 41
2.6 PROTEIN CRYSTALLISATION 42
2.6.1 INITIAL CRYSTALLISATION SCREENING 42
2.6.2 CRYSTAL OPTIMISATION 44
2.6.3 CRYSTAL CRYO-PROTECTION AND X-RAY DIFFRACTION DATA COLLECTION 45
2.6.4 X-RAY DIFFRACTION DATA PROCESSING 45
2.6.5 STRUCTURE DETERMINATION 47
2.6.6 STRUCTURE REFINEMENT 48
2.6.7 STRUCTURE VALIDATION AND ANALYSIS 48
2.7 SMALL-ANGLE X-RAY SCATTERING 49
2.8 SEQUENCE ALIGNMENTS 50
2.9 VISUALISATION OF GRAPHS 50

CHAPTER 3 – INVESTIGATION OF WILD-TYPE AND MUTANT FGFR3 AND FGFR4 TMD DIMERISATION 52

3.1 SUMMARY 52

3.2 INTRODUCTION 53

3.3 AIMS 53

3.4 SEQUENCE ANALYSIS AND SIMULATION SETUP 54

3.4.1 SEQUENCE ALIGNMENTS OF FGFR3 AND FGFR4 TMDS 54

3.4.2 MODEL BUILDING AND SIMULATION SETUP 57

3.4.3 HELIX DISTANCE TRACKING AND DIMERISATION TIME 60

3.5 WT AND MUTANT FGFR3 MD SIMULATIONS 63

3.5.1 ANALYSIS OF COARSE-GRAINED SIMULATIONS OF WT FGFR3 63

3.5.2 COARSE-GRAINED SIMULATIONS OF MUTANT FGFR3 70

3.5.3 ATOMISTIC SIMULATIONS OF FGFR3 WT AND G380R 77

3.6 ANALYSIS OF WT AND MUTANT FGFR4 MD SIMULATIONS 80

3.6.1 DIMERISATION INTERFACES OF WT AND MUTANT HELICES 80

3.6.2 REPRESENTATIVE TMD DIMER MODELS OF MUTANT FGFR4 82

3.7 DISCUSSION AND FUTURE PERSPECTIVES 84

CHAPTER 4 – BIOPHYSICAL STUDIES OF THE FGFR4-HEPARIN-FGF1 COMPLEX 90

4.1 SUMMARY 90

4.2 AIMS 91

4.3 EXPRESSION AND PURIFICATION OF FGFR4 D2D3$_{GLYC}$ 91

4.4 EXPRESSION AND PURIFICATION OF FGF1 96

4.5 ATTEMPTS AT FGFR4-HEPARIN-FGF1 COMPLEX FORMATION 98

4.5.1 GLYCOSYLATED FGFR4 AND FGF1 DO NOT FORM BINARY COMPLEX 98

4.5.2 HEPARIN INDUCED FORMATION OF FGFR4 SIGNALLING COMPLEX 101

4.6 FORMATION OF MORE STABLE SIGNALLING COMPLEXES 106

4.7 CRYSTALLISATION TRIALS OF VARIOUS FGFR4-HEPARIN-FGF1 COMPLEXES 109

4.7.1 COMPLEX FORMATION FOR CRYSTALLISATION TRIALS 109

4.7.2 CRYSTALLISATION TRIALS OF FGFR4 D2D3-HEPARINDP12-FGF1 111

4.8 SAXS ANALYSIS OF THE FGFR4-HEPARINDP12-FGF1 COMPLEX 114

4.9 CONCLUSIONS AND FUTURE PERSPECTIVES 116

CHAPTER 5 – STRUCTURAL INVESTIGATION OF THE FGFR1-HEPARINDP10-FGF1 SIGNALLING COMPLEX 119

5.1 SUMMARY 119

5.2 AIMS 119

5.3 ANALYSIS OF THE INTERACTION BETWEEN FGF1 AND HEPARIN 120

5.3.1 HEPARINDP12 FORMS STABLE FGF1 DIMER 121

5.3.2 HEPARINDP12 MEDIATED FGF1 DIMER FORMATION IS COOPERATIVE 123

5.3.3 SEC-MALLS IDENTIFICATION OF MOST HOMOGENEOUS FGF1 DIMER 125

5.4 FGFR1 SIGNALLING COMPLEX FORMATION 127

5.4.1 EXPRESSION AND PURIFICATION OF FGFR1 127

5.4.2 ENDOF1 DEGLYCOSYLATED FGFR1 DOES NOT BIND TO FGF1 131

5.4.3 HEPARINDP10 IN 2:1:2 STOICHIOMETRY FORMS MOST STABLE SIGNALLING COMPLEX 133

5.5 CRYSTALLISATION OF FGFR1-HEPARINDP10-FGF1 136

5.5.1 INITIAL CRYSTALLISATION SCREENING 136

5.5.2 CRYSTAL OPTIMISATION AND STRUCTURE DETERMINATION 139

5.5.3 REFINEMENT, CRYSTAL PACKING AND HEPARIN SOLUTION 142

5.6 N-LINKED GLYCOSYLATION OF FGFR1 147

5.6.1 SOLVING OF N-LINKED GLYCANS 147

5.6.2 LOCALISATION OF FGFR1 N-LINKED GLYCOSYLATION IN SIGNALLING COMPLEXES 149

5.7 THE ROLE OF HEPARIN IN THE FGFR1 SIGNALLING COMPLEX FORMATION 154

5.7.1 NONREDUCING HEPARIN ENDS FORM SAME INTERACTIONS TO RECEPTOR-LIGAND INTERFACE IN ALL THREE MODELS 154

5.7.2 C2-SYMMETRY AXIS IN NEW MODEL ABOLISHES ORIENTATIONAL RESTRAINTS OF HEPARIN 157

5.7.3 SOLELY THE NEW MODEL IS CONSISTENT WITH COOPERATIVE FGF DIMERISATION 161

5.8 CONCLUSIONS AND FUTURE PERSPECTIVES 164

CHAPTER 6 – STRUCTURAL ANALYSIS OF THE FGFR SIGNALLING COMPLEX IN SOLUTION 170

6.1 SUMMARY 170

6.2 INTRODUCTION 170

6.3 AIMS 171

6.4 SAXS CURVES FITTED TO ALLOSMOD STRUCTURE ENSEMBLES 172

6.4.1 MODELLING OF FLEXIBLE TERMINI AND $MAN_5GLCNAC_2$ MOIETIES 172

6.4.2 THE SCHLESSINGER MODEL IS NOT FOUND IN SOLUTION 174

6.4.3 AB INITIO MODELLED SAXS ENVELOPES CONFIRM ELONGATED COMPLEX ARCHITECTURE 176

6.5 SAXS INTERPRETATION ON THE BASIS OF MD SIMULATIONS 179

6.5.1 GENERATION OF A MD SIMULATION DERIVED STRUCTURAL ENSEMBLE 179

6.5.2 THE NEW SIGNALLING COMPLEX MODEL REPRESENTS THE COMPLEX ARCHITECTURE IN SOLUTION 181

6.6 CONCLUSION AND FUTURE PERSPECTIVES 184

CHAPTER 7 – FURTHER DISCUSSION AND FUTURE DIRECTIONS 187

7.1 SUMMARY 187

7.2 FURTHER DISCUSSION 187

7.2.1 THE SCHLESSINGER MODEL 188

7.2.2 THE PELLEGRINI MODEL 188

7.2.3 THE NEW MODEL AND STRUCTURAL ANALYSIS IN SOLUTION 189

7.2.4 FGFR SIGNALLING THROUGH THE MEMBRANE 191

7.2.5 FUTURE WORK 193

APPENDIX 195

APPENDIX A – EXPRESSION VECTORS 196

APPENDIX B – HFGFR SEQUENCE ALIGNMENT 198

APPENDIX C – OLIGONUCLEOTIDES USED FOR PLASMID PRODUCTION 200

APPENDIX D – NETNGLYC 1.0 SERVER PREDICTION 201

APPENDIX E – CA CLOSEST INTERHELIX DISTANCE PLOTS 203

APPENDIX F – LIST OF FORMULAE **206**

APPENDIX G – ABBREVIATIONS **209**

REFERENCES **211**

Chapter 1

Introduction

Chapter 1 – Introduction

1.1 Background

The developmental process from a single fertilised egg to a mature independent organism relies on a highly complex and tightly regulated series of changes in shape, composition and distribution of tissue. Each individual cell is integrated in a larger cellular network to secure smooth temporal and spatial control of gene expression through the process called embryogenesis – the miracle of nature that all eukaryotic life on earth depends on. Developing fingers that begin as five tiny buds sprouting from each side of a four-week-old human embryo will soon have grown little rods of cartilage in their tips. These are processes that only masterly orchestrated molecular signal transduction mechanisms between cells can guarantee. The same mechanisms define precise number, position and shape of a myriad of other developing structures, resulting in an adult human body consisting of about 4×10^{13} (40,000,000,000,000) cells.

Cellular communication is based on secretion of signal molecules such as hormones, neurotransmitters or morphogens (e.g. bone morphogenic proteins (BMPs), Wnt proteins or fibroblast growth factors (FGFs)). A multitude of specialised membrane-embedded receptor proteins read and process the extracellular signalling information leading to activation or inhibition of intracellular signalling pathways. This results in a cellular response by alteration of gene transcription and expression patterns. The transmembrane fibroblast growth factor receptors (FGFRs) are members of the prominent receptor tyrosine kinase (RTK) family. FGFRs play a key role in processes such as cellular growth, differentiation, metabolism and motility. Therefore, deregulation of FGFR signalling leads to developmental defects resulting in skeletal disorders or to cancer in the case of gain-of-function mutations.

1.2 An overview of the Receptor Tyrosine Kinase family

Biological membranes are composed of phospholipid double layers that enclose and compartmentalise cells of all organisms. They act as effective insulators and selective filters between the inner cytoplasm and the extracellular matrix. Each membrane accommodates membrane proteins or protein complexes that constitute about 30% of all genes in eukaryotic cells (Stevens & Arkin, 2000). They are involved in many basic cellular processes from ion transport (Gamper & Shapiro, 2007; Grigoryan, Moore, & DeGrado, 2011), through cell-cell recognition and adhesion (Maître & Heisenberg, 2013) to signal transduction (Lemmon & Schlessinger, 2010; Rosenbaum et al., 2009). This is why membrane proteins are important targets for current drugs. However, since they are harder to handle than soluble proteins, our understanding for this important class of proteins lags behind the understanding of soluble proteins.

In humans, membrane-embedded receptor tyrosine kinases form a superfamily of 58 known members, which fall into 20 subfamilies (**Figure 1.1**). All RTKs share a common molecular architecture with an extracellular ligand-binding domain, a single-pass transmembrane domain (TMD), which makes RTKs part of the large group of single-spanning membrane proteins (Hubert et al., 2010), and a cytoplasmic kinase domain that is linked to the TMD via a flexible juxtamembrane domain (JMD). Well-studied RTKs are amongst others epidermal growth factor receptors (EGFRs), insulin receptors (InsR), ephrin receptors (Ephs) and the FGFRs, which are essential determinants during embryonic development and adult homeostasis. Functional and structural studies have revealed unexpected diversity in the mechanisms by which polypeptide ligands – mainly growth factors – trigger RTK oligomerisation and activation. However, their signalling pathways are highly conserved in evolution from nematodes to humans, underlining their central role in development. Before activation by ligand-induced oligomerisation most RTKs exist in a monomeric inactive state. Lateral receptor dimerisation brings the cytoplasmic tyrosine kinase domains into close proximity (Schlessinger et al., 2000) enabling them to *trans*-autophosphorylate tyrosine residues in the kinase activation loop (H. Chen et al., 2008). This stabilises the active conformation of the kinase domain (Hubbard, 2004) that then phosphorylates further C-terminal tyrosine residues in *cis* and in *trans*. The generated phosphotyrosine residues serve as recruitment sites for a multitude of downstream signalling proteins.

In recent years, studies on the role of the transmembrane and juxtamembrane domain have added new levels of complexity to the understanding of RTK signalling. The analysis of triggering of activation by conformational coupling across the membrane is now revealing dimerisation processes on a more subtle level (Endres et al., 2013; Kaszuba et al., 2015). Since the identification of a TMD sequence motif that drives dimerisation of many integral membrane proteins (Lemmon et al., 1994; Schneider & Engelman, 2004), cellular, biochemical and computational studies have shown that the same motif drives dimerisation of RTK TMDs (Fleishman et al., 2002; Hubert et al., 2010; E. Li et al., 2005). Experiments using the TOXCAT assay (Russ & Engelman, 1999) have shown that the TMDs of all 58 human RTKs alone have an intrinsic propensity to form stable dimers within a membrane (Finger et al., 2009). The authors of this study suggest that the interaction of TMDs is a general property of RTKs that contributes to activation and signalling. Single amino acid mutations in RTK TMDs induce unregulated signalling and consequently pathological phenotypes (E. Li et al., 2005). Examples include a variety of FGFR3 mutants (He et al., 2010; 2011; Placone & Hristova, 2012; You et al., 2012) and FGFR4 mutants (Bange et al., 2002; Roidl et al., 2010; J. Wang et al., 2008). For many RTK TMDs computational analysis identified accurate TMD dimer interfaces that were in agreement with experimental reports (Reddy et al., 2014; Volynsky et al., 2013).

Although RTK JMDs were considered inherently flexible sequences, methodological advances provided a deeper understanding of their function. A computational study across all human RTKs has revealed conserved interactions of the JMDs to anionic membrane lipids. This membrane association decreases JMD flexibility and has the potential of reorienting the C-terminal kinase domain (Hedger et al., 2015). Structural studies have revealed that JMD regions can bind to and negatively regulate kinase domain activity, a finding that has been described as a general mechanism for RTKs (Hubbard, 2004). In addition, the JMD has been shown to take centre stage in EGFR activation by formation of an antiparallel α-helix dimer of the JMD's N-terminal half, whilst the C-terminal segment latches to the activated kinase domain (Jura et al., 2009; Brewer et al., 2009).

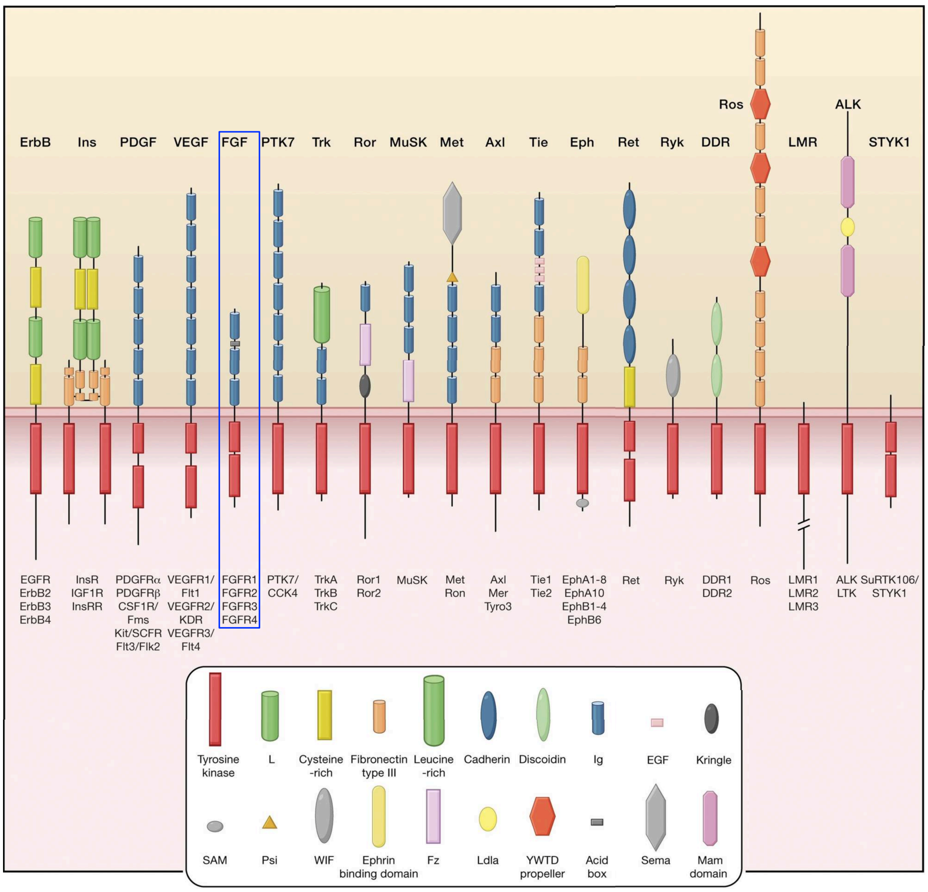


Figure 1.1 Receptor tyrosine kinase subfamilies

Schematic domain map of all 20 subfamilies of human receptor tyrosine kinases (RTKs). The family members are listed beneath each receptor. Structural domains in the extracellular regions were identified by structure determination or sequence analysis and are marked according to the key. The intracellular kinase domains are shown as red rectangles. The FGFR family is highlighted with a blue box. (Reprinted from Lemmon & Schlessinger, (2010) with permission from Elsevier)

Besides these rather recent insights into RTK signalling mechanisms, the kinase domain dimerisation and phosphorylation, and ligand-induced dimerisation of extracellular domains have been the subject of intense studies through multiple decades. Mutations on all three segments (ectodomain, TMD, kinase domain) of RTKs have been implicated in a wide variety of diseases including cancers and skeletal disorders. This places RTKs among the most intensely pursued drug targets.

1.3 FGFR signalling

The fibroblast growth factor receptor was discovered 40 years ago and has since then been found to be expressed in a wide range of cell types. FGFR signalling plays an essential role in a multitude of cellular processes such as cell proliferation, differentiation and survival, where it controls tightly regulated embryonic and adult developmental processes including branching morphogenesis, organogenesis, brain patterning, limb development, angiogenesis and homeostasis (Aricescu & Jones, 2007; Beenken & Mohammadi, 2009).

1.3.1 Fibroblast Growth Factor Receptors

There are four FGF signalling receptors (FGFR1-4), which consist of an extracellular ligand-binding region, a single-pass α-helical transmembrane domain and a C-terminal cytoplasmic kinase domain (see blue box **Figure 1.1**; (Goetz & Mohammadi, 2013)). The extracellular domain comprises three immunoglobulin (Ig)-like domains (Aricescu & Jones, 2007) D1 to D3. D2 and D3 are crucial for binding to one of 18 locally or hormonally acting fibroblast growth factor family members (Ornitz et al., 1996; Ornitz & Itoh, 2001; Plotnikov et al., 1999). Alternate splicing of FGFR homologs results in 48 different isoforms that vary in their ligand-binding properties (Duchesne et al., 2006; Holzmann et al., 2012). Tissue-specific splicing leads to restriction of FGFRb isoform expression in epithelial cell lineages and FGFRc isoform expression in mesenchymal cell lineages (Orr-Urtreger et al., 1993; Wuechner et al., 1996). Each receptor isoform exhibits a distinct profile for ligand-binding specificity (see **Figure 1.2**) creating a complex network of regulatory options. As FGFRb isoforms mainly bind to FGFs that are secreted from mesenchymal tissue, and FGFRc isoforms to FGFs secreted from epithelial tissue, a reciprocal receptor-ligand expression pattern leads to specific signalling loops between the mesenchyme and the epithelial tissue. This enables finely tuned orchestration of developmental processes and tissue homeostasis in adults (Xu et al., 1998; Zhang et al., 2006).

1.3.2 Fibroblast Growth Factors

FGFs are clustered into one non-secretory (FGF11; (R. Xu et al., 2013)) and 6 secretory subfamilies, which can be further categorised into one endocrine and five paracrine subfamilies (Belov & Mohammadi, 2013). Endocrine FGFs are involved in more global metabolic processes such as phosphate homeostasis and glucose metabolism whereas paracrine FGFs act locally in processes such as organogenesis and most notably in the repair and remodelling of tissues (Jonker et al., 2012). A chain of 120 amino acids forms the core homology domain of FGFs, which folds into a globular β-trefoil. The β-trefoil domain of paracrine FGFs is typically composed of 12 β-strands (Mohammadi et al., 2005), whereas endocrine FGFs lack the β-11 strand and therefore have an atypical β-trefoil. The highly divergent amino-terminal and carboxy-terminal sequences that flank the conserved core domain have a crucial role in various functional properties of FGFs (Goetz et al., 2010; Kalinina et al., 2009; Olsen et al., 2006). Just as the receptor, FGF activity is also regulated by various mechanisms such as alternative splicing, homodimerisation and most importantly heparan sulphate binding.

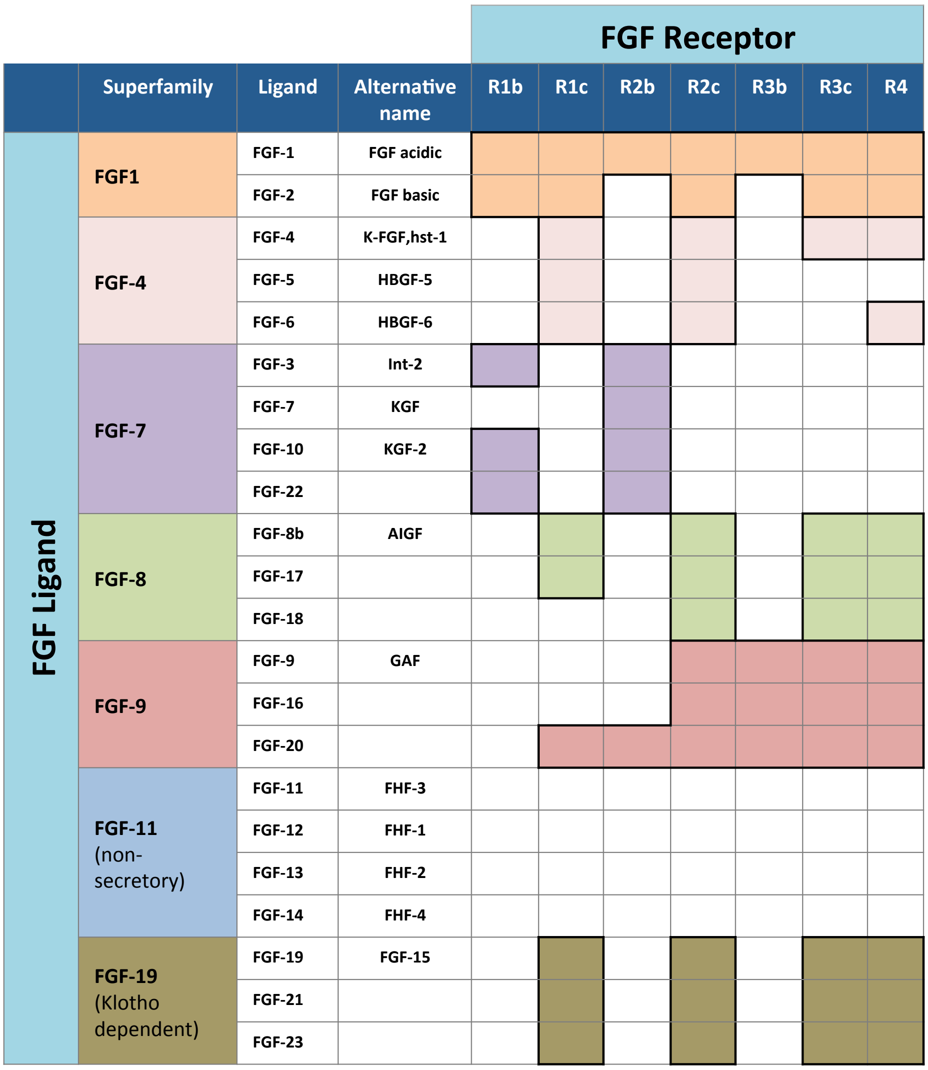


Figure 1.2 FGFR-FGF interaction matrix

The extracellular Ig-like domains of FGFRs 1 to 3 are alternately spliced, creating "b" and "c" isoforms. They differ in tissue distribution and ligand specificity (reviewed in (Holzmann et al., 2012)). FGFs are grouped into seven subfamilies based on sequence similarity and functional characteristics. Subfamilies FGF1, 4, 8, 9 and 19 act paracrine and subfamily FGF7 acts autocrine. Members of subfamily FGF11 are not secreted and members of subfamily FGF19 depend on Klotho-cofactor for signalling (Goetz et al., 2007). FGFs bind to heparan sulphate proteoglycans (HSPGs) and FGF receptor tyrosine kinases to activate intracellular signalling pathways triggered by receptor dimerisation, autophosphorylation and subsequent recruitment and docking of SH2 domain-containing proteins. FGFRL as a fifth, soluble subtype is excluded from this overview. FGFRL lacks an intracellular kinase domain but is still capable of binding to FGFs (Sleeman et al., 2001).

1.3.3 Heparan Sulphate Proteoglycans

Heparan sulphate proteoglycans (HSPGs) are abundantly present both at the cell surface and in the pericellular and extracellular matrix. They are split into three families. Syndecans and glypicans are cell surface HSPGs that are linked to the cell membrane via a TMD or a glycosyl phosphatidyl inositol (GPI) anchor, respectively. Perlecans are secreted HSPGs that reside in the extracellular matrix. All HSPGs are decorated with heparan sulphate (HS) chains, which form a vital regulator for the formation of a morphogen gradient. Among other functions, the interaction between HS and paracrine FGFs serves to retain ligand near the site of action, store the ligand, enhance the ligand's stability and limit the area of ligand signalling (Sommer & Rifkin, 1989). Most importantly, heparan sulphate is required for binding of FGF to FGFR and for formation of the FGFR signalling complex on the plasma membrane (Ornitz et al., 1992; Powell et al., 2002; Rahmoune et al., 1998; Spivak-Kroizman et al., 1994). HS is a glycan composed of linearly arranged, repeating units of N-substituted glucosamine and glucuronic acid (see bracket in upper structure in **Figure 1.3**; (Harada et al., 2009; Perrimon & Bernfield, 2000; Yan & Lin, 2009). Along the glycan chain, regions of 5 to 10 N-sulphated disaccharide domains harbouring various modifications (including 2-O-sulphation of glucuronic or iduronic acid and 3-O- and 6-O-sulphation of N-sulphated glucosamine) alternate with N-acetylated regions that are generally unmodified (Feyerabend, Li, Lindahl, & Rodewald, 2006). The blue box in **Figure 1.3** indicates the fully sulphated disaccharide unit that represents a heparin with degree of polymerisation (DP) of *2*. According to the nomenclature used in previous publications and in this thesis, this heparin is called heparinDP2. Thus, heparinDP8, 10 and 12 consist of 4, 5 or 6 fully sulphated IdoA-GlcNac disaccharides. On the cell surface, these highly sulphated domains constitute the main binding sites for FGFs (Böttcher & Niehrs, 2005; Gallagher, 2001; Turner & Grose, 2010). FGF activity is notably influenced by tissue-specific modification of heparan sulphate, producing heparan sulphates with a distinct "fine structure".

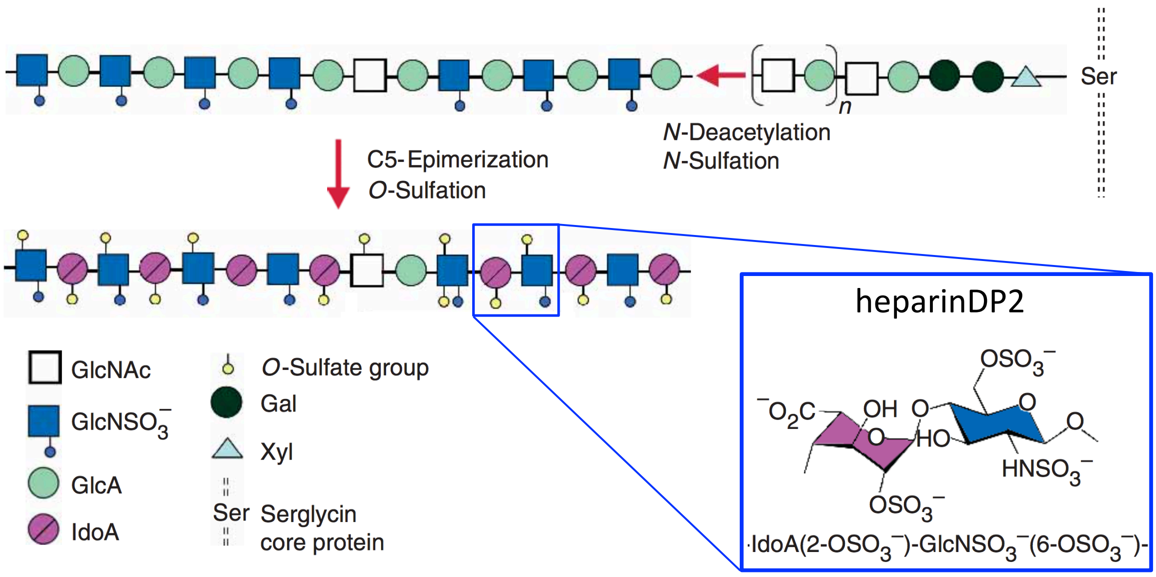


Figure 1.3 Overview of heparin biosynthesis and nomenclature

An N-sulphated intermediate polysaccharide (upper structure) is normally modified through GlcA C5-epimerization and O-sulphation at various positions to yield heparin (lower structure). The heparan sulphate chain is linked to the core protein via a serine residue that is usually in the sequence Ser-Gly-X-Gly (where X can be any amino acid but proline). (Adapted from Feyerabend et al., 2006 with permission from Elsevier)

1.3.4 FGFR signalling pathways

FGF- and HS-mediated activation of FGF receptor kinase results in activation of its intracellular substrates. This leads to initiation of distinct but potentially interactive signalling pathways, typically resulting in diverse cellular responses described as cell proliferation, cell survival or cell motility (Böttcher & Niehrs, 2005; B. Thisse & Thisse, 2005; Turner & Grose, 2010). FGFR substrate 2α (FRS2α), which is constitutively bound to the receptor kinase, and phospholipase Cγ1 (PLCγ1) are the major substrates of FGFR kinases (Carpenter & Ji, 1999; Gotoh, 2008). The adaptor protein growth factor receptor-bound 2 (GRB2) binds to activated FRS2α and recruits either the guanine nucleotide exchange factor son of sevenless (SOS) (see **Figure 1.4 A**) or the adaptor protein GRB2-associated binding protein 1 (GAB1) (see **Figure 1.4 B**). SOS activates the protein Rat sarcoma (RAS) GTPase, which triggers activation of the mitogen-activated protein kinase (MAPK) cascade (**Figure 1.4 A**; (Kouhara et al., 1997)). MAPK subsequently translocates from the cytoplasm to the nucleus in order to phosphorylate and activate immediate early gene transcription factors. This induces transcription of specific genes, which primarily lead to cell proliferation. However, this pathway can also lead to other responses such as cell migration or cell differentiation (Böttcher & Niehrs, 2005; B. Thisse & Thisse, 2005; Turner & Grose, 2010). On the other hand, GAB1 leads to activation of protein kinase B (AKT kinase) (Lamothe et al., 2004), which inhibits pro-apoptotic effectors such as forkhead box class O (FOXO) transcription factors and the BCL-2 antagonist of cell death (BAD). Thus, activation of AKT kinase leads to promotion of cell survival (**Figure 1.4 B**; Brunet et al., 1999; Datta et al., 1997). **Figure 1.4 C** shows recruitment and activation of PLCγ1, which initiates a different signalling pathway that is involved in cell differentiation and cell migration. Phosphorylated PLCγ1 catalyses the hydrolysis of the membrane phospholipid phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)(P_2) into diacylglycerol (DAG) and inositol-1,4,5,-trisphosphate (IP_3) (Carpenter & Ji, 1999). DAG activated protein kinase C (PKC) phosphorylates its substrates, including the myristoylated alanine-rich C kinase substrate (MARCKS), which is a regulator of cell motility (Hartwig et al., 1992). Calcium ions from intracellular stores are released upon phosphorylation of IP_3, which triggers activation of calcium-dependent proteins such as the phosphatase calcineurin. This leads to subsequent nuclear translocation of transcription factor nuclear factor of activated T cells (NFAT), which triggers transcription of genes that are essential for cell motility (D. Li et al., 2011).

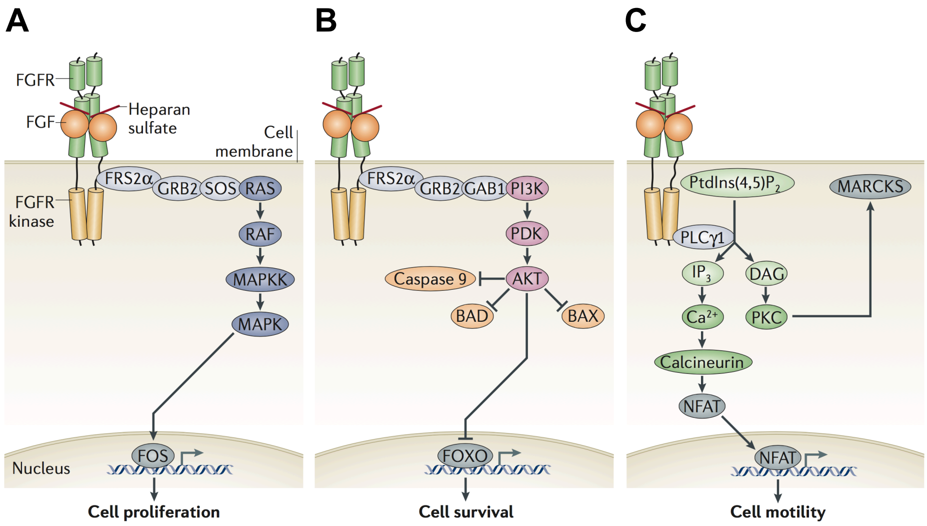


Figure 1.4 FGFR-FGF signalling pathways

Heparan sulphate proteoglycan bound FGFs on the surface of the target cell bind to the corresponding FGFRs, triggering formation of ternary FGFR–HSPG–FGF complexes in the membrane plane. This brings the intracellular kinase domains into close proximity, enabling them to cross phosphorylate each other. Phosphotyrosines recruit adaptor proteins that initiate one of three main FGFR signalling pathways. The FRS2α dependant Ras-MAPK pathway leads to cell proliferation (**A**), the FRS2α-PI3K dependant AKT pathway results in a cellular response of survival (**B**) and PLCγ1 dependant release of calcium ions triggers cell motility (**C**). (Abbreviations not explained in the main text: BAX, BCL-2-associated X protein; MAPKK, MAPK kinase; PDK, phosphoinositide-dependent protein kinase). (Reprinted from Goetz & Mohammadi, 2013; with permission from Elsevier)

1.4 Structural biology of the extracellular FGFR signalling complex

"A heparin string with FGF beads, heparin glue between FGF monomers, or heparin glue between FGF and its receptor? Two receptor molecules brought together by a single FGF monomer, by an FGF dimer, or in higher order aggregates?" (Wiesmann & de Vos, 1999). The arrangement of the active FGFR signalling complex has been matter of controversial discussion for over 15 years. Despite huge experimental efforts aiming to establish a plausible mechanism of complex assembly and despite multiple published receptor-ligand crystal structures, the ambiguity about the receptor dimerisation mode and the role of heparin could not be convincingly resolved.

1.4.1 The binary FGFR–FGF complex

There are currently crystal structures of eight different FGF-FGFR combinations available (without heparin). These include structures of FGFR1, 2 and 3, with constructs that comprise Ig domains D2 and D3 only. Within one year, two independent laboratories published crystal structures of all possible complex combinations of FGFR1c or FGFR2c with FGF1 or FGF2, which enabled accurate mapping of the receptor-ligand interface residues (pdb ID: 1CVS, (Plotnikov et al., 1999); pdb ID: 1DJS, (Stauber et al., 2000); pdb ID: 1EV2 and 1EVT, (Plotnikov et al., 2000)). A study based on the crystal structure of FGFR3c-FGF1 demonstrated that FGFR D1 and the D1-D2 linker (also called acidic box) are dispensable for FGF binding, that the acidic box auto inhibits FGF binding through interactions with the minimal FGF binding site, and suggested that the FGF1 N-terminus confers receptor binding affinity and promiscuity (pdb ID: 1RY7, (Olsen et al., 2004)). Two biochemical FGFR mutant studies followed, suggesting pathophysiological mechanisms for various skeletal disorders on the basis of the previous structural data (Ibrahimi et al., 2004; Ibrahimi et al., 2005). The crystal structure of FGF8b in complex with FGFR2c revealed the structural basis by which alternative splicing of the FGF8 modulates its organising activity (pdb ID: 2FDB, (Olsen et al., 2006)). Finally, structural studies that compared FGFR "b" and "c" splice variants of FGFR1 and FGFR2 in complex with FGF1 revealed the structural basis for the ability of FGF1 to cross-react with all isoforms of FGFR (pdb ID: 3OJ2, 3OJM, 3OJV, (Beenken et al., 2012)). These structural insights into the nature of 1:1 receptor-ligand dimers began to illuminate the molecular complexity of the FGF signalling system.

1.4.2 The ternary FGFR–heparin–FGF signalling complex

For signal transduction, FGFR-FGF dimers need to be brought into higher order complexes. This dimerisation of dimers involves heparan sulphates for fine-tuning (Ornitz & Itoh, 2001) and as a ternary signalling complex they produce specific signalling outputs in diverse biological contexts. To date only two complex structures have been published that comprise an HSPG-mimicing heparin sulphate oligosaccharide associated with two ligands and the extracellular ligand-binding regions (Ig domains D2 and D3) of two FGF receptors. These structures represent different arrangements of the FGFR signalling complex.

The 2:2:2 FGFR1-heparinDP10-FGF2 complex published by Schlessinger et al in 2000 (see **Figure 1.5 A**) is formed of two binary FGFR1-FGF2 complexes that are associated by direct receptor-receptor contacts. This arrangement creates a deep, positively charged canyon that is saturated with two heparinDP10 chains. The same 2:2 protein arrangement was previously observed for the packing of the crystal lattice in binary FGFR2-FGF1 complexes without heparin (Plotnikov et al., 1999). The authors reported that a single heparin chain is likely to occupy the complete heparin binding cleft in a 2:1:2 complex, an arrangement which to date has not been presented. Schlessinger and co-workers obtained the 2:2:2 complex structure instead, by preformation of 2:2 receptor-ligand crystals of FGFR1 and FGF2 and subsequent soaking with heparinDP10. Each of the two oligosaccharides occupies one half of the cleft with the nonreducing ends pointing at each other. According to this model each heparin molecule promotes the formation of a stable 1:1:1 FGFR-heparin-FGF ternary complex by cross-linking. Receptor dimerisation occurs upon recruitment of a second 1:1:1 ternary complex, assembling to a 2:2:2 complex (bridging both ternary complexes) stabilised by direct FGFR-FGFR contacts and by secondary FGF-FGFR interactions. In the literature this model is referred to as the Schlessinger model, symmetric model, two-end model or the canyon dimer model.

The contrasting Pellegrini model was published in the same year and suggests an asymmetric model, built of FGFR2-heparinDP10-FGF1 in a 2:1:2 ratio (see **Figure 1.5 B**; Pellegrini et al., 2000)). Crystals leading to this structure were grown from a solution containing all purified components in a predefined 2:1:2 stoichiometry. The signalling complex formed by two receptor-ligand units is completely devoid of any secondary protein–protein contacts. The two FGFR–FGF halves are therefore held together solely by heparin. Furthermore, heparin is positioned asymmetrically between both receptor-ligand dimers, interacting differently with each ligand, and binds to one receptor only. Hence, only a single 1:1:1 FGFR–heparin–FGF ternary complex is present in this model and heparin enhances ligand–receptor binding for only one of the two FGF–FGFR binary complexes. The unoccupied heparin–binding site is saturated with sulphates from the crystallisation condition. This model is referred to as the Pellegrini model or the asymmetric model.

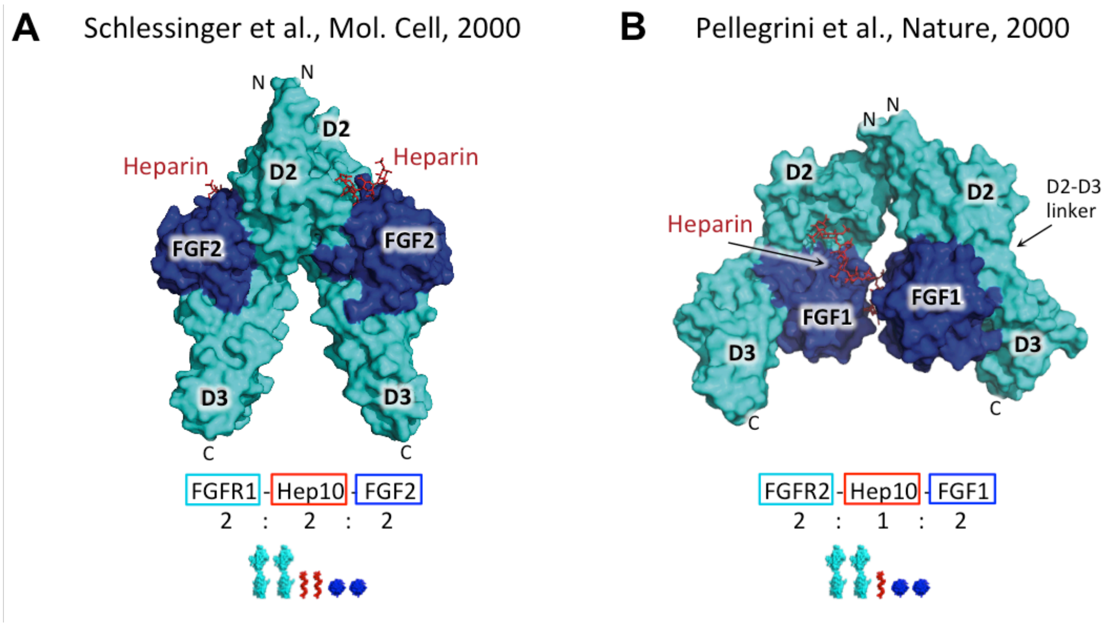


Figure 1.5 FGFR signalling complexes according to Schlessinger and Pellegrini

Comparison of the architecture of the 2:2:2 FGFR1–heparinDP10–FGF2 two-end Schlessinger model (**A**) (pdb ID: 1FQ9, Schlessinger et al., 2000) and the asymmetric 2:1:2 FGFR2–heparinDP10–FGF1 Pellegrini model (B) (pdb ID: 1E0O, Pellegrini et al., 2000). Both structures were solved by X-ray crystallography at resolutions of 3.0 Å and 2.8 Å respectively. The receptor molecules are coloured cyan and the ligands are coloured blue. Heparin chains are in red. Colour coded boxes indicate the complex components. The differing stoichiometries of the respective complex components are highlighted in numbers and in small cartoon representations below the structures.

It is still the subject of controversial discussion, which of these two models is formed on the cell surface for effective signalling through the membrane. The Pellegrini camp suggests that both models can be formed in solution and that both models might be of biological relevance (Brown et al., 2013). They argue that the 2:1:2 complex which forms upon a single HS chain and the 2:2:2 complex that requires nonreducing ends from two separate HS chains might represent different states of activation, regulated by a trans- to cis-isomerisation of a conserved proline in the D2-D3 linker invariant proline. However, the Schlessinger camp argues that the two-end model represents the only relevant signalling complex architecture. They have supported their hypothesis with detailed structural and functional studies of the D2–D3 trans-conformation and have argued that there are further implausible structural features in the Pellegrini model (Belov & Mohammadi, 2013; Mohammadi et al., 2005).

Importantly, besides regulatory mechanisms that involve interactions between receptor, ligand and heparin, there are multiple more levels of regulation. These involve amongst others JMD activation and inhibition (resulting in regulation of kinase dimerisation), down regulation of activated receptors and varying propensity of TMD dimerisation. However, intracellular processes exceed the scope of this study.

1.5 Dysregulation of FGFR signalling in human malignancies

The various developmental abnormalities and cancers that occur due to gain-of-function of RTK signalling underline their diverse and essential regulatory roles (Blume-Jensen & Hunter, 2001; Peters et al., 1994; Robertson et al., 2000). In humans, there are three main causes for signalling irregularities:

a) generation of oncogenic fusion proteins (chromosomal translocation),
b) mutations or deletions causing overactivation and
c) overexpression as a result of gene amplification.

The importance of FGFR gene expression was confirmed in the occurrence of dramatic consequences upon targeted gene inactivation. FGFR1 gene disruption leads to embryonic death (Deng et al., 1994). Inactivation of FGFR2 results in abnormal pulmonary development (Peters et al., 1994) and inactivation of FGFR3 causes bone overgrowth, which makes FGFR3 a negative regulator of bone growth (Deng et al., 1996). In contrast, no abnormal phenotypes could be found for the FGFR4 null mice (Weinstein et al., 1998). Interestingly, in screening of over 1,000 exon mutations of protein kinase genes from 210 different malignancies, the FGFR signalling pathway was identified as the most commonly mutated tyrosine kinase signalling pathway (Gallo et al., 2015).

There are multiple mechanisms that underlie aberrant signalling through activating FGFR mutations. These include enhanced kinase domain activation, ligand-independent receptor dimerisation and altered affinity for FGF ligands. The long lists of FGFR mutations have been compiled and categorised in different ways for review (Carter et al., 2015; Gallo et al., 2015; Touat et al., 2015; Wesche et al., 2011). Whereas mutations on the FGFR kinase domains are rather rare, the most common activating mutations affect either the FGFR ectodomain or the TMD. Most mutations for which functional experiments demonstrated loss- or gain-of-function are located on FGFR2 and FGFR3 (Gallo et al., 2015).

1.5.1 Mutations in the extracellular domain

Disulphide bond disruption in the extracellular domain by cysteine mutations is a common cause of tyrosine kinase overactivation found in FGFR2, FGFR3 and FGFR4. Mutations C278F and C340S/Y on FGFR2 perturb a disulphide bond and cause human genetic disorders e.g. Crouzon and Pfeiffer Syndrome by increased receptor activation, exhibiting premature closure of cranial sutures. These are craniosynostosis syndromes, which are usually accompanied by defects in brain development and chondrocyte signalling.

Mutations FGFR2 W290C and S352C introduce new cysteine residues causing again Pfeiffer Syndrome or Crouzon Syndrome, respectively (Galvin et al., 1996; Reardon et al., 1994). Similar mutations that have been identified in various cancers such as endometrial carcinoma and breast cancer include R203C, Y281C, S320C, Y338C, and S373C. These are all gain-of-function mutations that cause constitutive activation of FGFR and drive cellular proliferation and tumour progression.

Novel cysteine residues introduced by mutations also occur in FGFR3. R248C and S249C cause Thanatophoric Dysplasia Type I (TDI), a severe form of Achondroplasia (dwarfism) that typically causes neonatal death. These mutations have also been implicated in a multitude of cancer types (including bladder cancer, cervical cancer, gallbladder cancer, lung squamous cell carcinoma, head and neck squamous cell carcinoma and multiple myeloma). Importantly, these mutations are able to cause ligand-independent receptor activation (Liao et al., 2013).

The structurally and functionally best-studied non-cysteine mutations in the extracellular domain are the analogous D2–D3 linker mutations of FGFR1 and FGFR2 at position P252 or P253, respectively. P252R/S/T substitutions activate FGFR1 and lead to melanoma, lung adenocarcinoma and spermatocytic seminoma. Autosomal dominant mutations at this codon lead to Pfeiffer Syndrome (d'Avis et al., 1998). On FGFR2, the P253R/S mutations are associated with Apert Syndrome (Nagata et al., 2011; Webster & Donoghue, 1996) and have also been implicated in cancers such as lung adenocarcinoma and oral squamous cell carcinoma. Crystal structures and biophysical analysis of FGFR2c P253R have revealed that the Apert Syndrome mutation enhances the ligand-binding affinity without influencing the FGFR conformation (Ibrahimi et al., 2001; 2004; Olsen et al., 2006). S252W in FGFR2 is another well-studied non-cysteine substitution, which occurs in approximately 67% of patients with Apert Syndrome (Ibrahimi et al., 2001). This mutation interferes with proper endochondral bone development through aberration of MAPK signalling activation Chen et al., 2014).

1.5.2 Activating mutations in the transmembrane domain

The effects of pathological TMD mutations in FGF receptors have been the subject of multiple cellular and computational studies, trying to dissect common underlying mechanisms of over activation (L. Chen et al., 2010; Sarabipour & Hristova, 2013; You et al., 2007). Activating mutations within the transmembrane domain of RTKs have been suggested to induce re-orientations in the dimer interface of receptors, resulting in increased receptor activation (Bell et al., 2000). This section describes the most prominent mutations located on the TMDs of FGFR3 and FGFR4, which have been implicated in human pathologies (these mutations are the subject of study in **Chapter 3**).

A single point mutation on the FGFR3 gene results in a glycine to arginine substitution at position 380 in the TM domain, which is reported to cause Achondroplasia, the most common genetic form of dwarfism (Naski et al., 1996; Oberklaid et al., 1979; Shiang et al., 1994). This mutation induces constitutive activation of the tyrosine kinase domain without apparently altering the TM domain dimerisation energetics (E. Li et al., 2005; Webster & Donoghue, 1996; You et al., 2006). Thus, the molecular mechanism by which the G380R mutation triggers over activation of the receptor is still not understood. The FGFR3 Y373C mutation is reported to be responsible for the lethal skeletal disorder TDI (d'Avis et al., 1998). The mutant cysteine residue is thought to allow receptor dimerisation by formation of abnormal disulphide bonds, subsequently leading to ligand independent constitutive activation.

The FGFR4 G388R substitution (similar to the FGFR3 G380R mutation) is caused by a common single-nucleotide polymorphism (SNP). This variant results in sustained receptor activation due to impaired lysosomal degradation (Wang et al., 2008). The FGFR4 Arg388 allele, occurring in approximately 50 % of the human population, is innocuous in healthy individuals but predisposes cancer patients for significantly accelerated disease progression (Bange et al., 2002). FGFR4 Y367C is analogous to the FGFR3 Y373C mutation but has not yet been explicitly connected to any pathological disorder. However, it is hypothesised to lead to a constitutively active receptor tyrosine kinase and thus to enhance tumour growth (Roidl et al., 2010).

1.5.3 Strategies for therapy

FGFR mutations that are mainly linked to cancers and severe bone disorders cause aberrant activation of their intracellular signalling pathways. The importance of FGF signalling has driven development of new drugs targeting these pathways by blocking or attenuating kinase activity. Many therapeutics are in active development and a few more are entering the clinic.

The overall success of clinical treatment with tyrosine kinase inhibitors is challenged by secondary mutations resulting in acquired mechanisms of chemotherapy resistance (Sierra et al., 2010). Another level of complexity in terms of specificity of small-molecule inhibitors is added by mutant FGFR fusion proteins. It has also been shown for FGFR1 that the kinase domain gatekeeper residue mutation V561M confers in vitro drug resistance to the multikinase inhibitor PP58 (Blencke et al., 2004). The ATP-competitive multikinase inhibitor dovitinib has proved to be active against FGFRs (Renhowe et al., 2009) and has shown promising pre-clinical anti–tumour activity in FGFR driven cancers (Trudel et al., 2005). A number of FGFR2 mutations were observed in response to dovitinib (N550H/K/S/T, V565I, E566A/ G and K642N) that lead to enhanced activity of the receptor kinase. Residues N550, E566 and K642 form a network of hydrogen bonds called the molecular brake, which have an autoinhibitory effect. Drug-resistant mutations that are located at this residue network disrupt the molecular brake in the kinase hinge region of FGFR2 (Byron et al., 2013). There are a number of drugs that are successfully used in the clinic that also show promising effects in connection with FGFR mutations. For instance, a patient with oral squamous cell carcinoma expressing FGFR2 P253R was treated with pazopanib, a multikinase inhibitor that effectively reversed cellular transformation and contributed to tumour shrinkage (Liao et al., 2013). The development of effective cancer treatments depends on the prediction of drug-resistance mechanisms evolving in cancerogenous cells. The efficacy of chemotherapy agents is compromised without such predictions, contributing to a decrease in treatment progression and finally resulting in tumour relapse.

1.6 Aims

As discussed above, FGFR signalling is implicated in many disorders and diseases with extensive potential for severe developmental, homeostatic or oncogenic consequences. There is a lack of understanding of the FGFR signalling complex architecture and underlying mechanisms of assembly on the cell surface, many of which are of high medical importance. Accurate structural information is required for efficient structure-based drug design. The main aim of this thesis is a structural investigation of the fibroblast growth factor receptor association with its extracellular ligands FGF and HSPG. X-ray crystallography and molecular dynamics simulations, in combination with a range of biophysical techniques will serve to reveal new molecular details of the FGFR ectodomain and TMD complex assembly.

Chapter 2

Material and Methods

Chapter 2 – Material and Methods

2.1 Molecular Dynamics

2.1.1 Coarse Grained Simulations

Atomistic models of the α-helical TM fragments of wt and mutant FGFR3 and FGFR4 were built in PyMOL (version 1.4, DeLano Scientific). The N- and C-termini were chosen according to sequence alignments and TMD predictions. The same fragments where chosen for FGFR4, aligning the sequences according to the predicted hydrocarbon core embedded segments (for exact TMD sequence boundaries see **Table 6**). In order to perform CG simulations these atomistic structures were converted to CG models as described previously (Bond & Sansom, 2006; Bond, Wee, & Sansom, 2008). All simulations were performed using a modified version of the Martini force field with an approximate 4:1 mapping of heavy atoms (i.e., not H) to CG particles (Bond et al., 2008; Bond & Sansom, 2006; Monticelli et al., 2008). Each CG particle is assigned to one of four particle types, namely "polar" (P), "mixed polar/apolar" (N), "hydrophobic apolar" (C) and "charged" (Q) with further subtypes for the N and Q particles to allow fine-tuning of Lennard-Jones interactions. To mimic hydrogen bonding in α-helices, the distance between backbone particles was restrained with a harmonic distance restraint with an equilibrium length of 6 Å and a force constant of 1000kJ mol^{-1} nm^{-1} (Bond et al., 2008). All simulations were performed using GROMACS 4.5.4 (Van Der Spoel et al., 2005) with CG parameters for palmitoyl oleoyl phosphatidyl choline (POPC, zwitterionic headgroup), palmitoyl oleoyl phosphatidyl serine (POPS, partially negatively charged headgroup), water molecules, amino acids and sodium or chlorine ions as described previously (Bond & Sansom, 2006; Kalli et al., 2010). All CG simulations were set up in the same fashion. Two copies of the same α-helical transmembrane domain were placed in a box (100x100x100Å) along the z-axis in a distance of approximately

70 Å. About 320 CG lipid molecules were placed randomly around the helices before filling the box with CG water. Each CG model was energy minimized using < 100 steps of the steepest descent method. In order to run a 200 ns self-assembly simulation to obtain a lipid bilayer in the x-y-plane, both helices were position restrained. After correct formation of a bilayer around each TMD pair, restraints were removed, energy minimisation and 20 ns equilibration was performed in order to finally run 20 production simulation repeats of 1 μs of each system. Analysis of the trajectories was performed with GROMACS 4.5.4, and with adapted MDAnalysis-dependant (Michaud-Agrawal et al., 2011) Python functions (kindly provided by Tyler Reddy; (Reddy, 2012)). VMD 1.9 and PyMol 1.4 were used for visualisation of molecules.

2.1.2 Atomistic Simulations

Initial atomistic (AT) simulation systems were generated from representative CG TMD dimer structures (obtained using the g_cluster module of GROMACS) using the in-house developed CG2AT protocol as described previously (Stansfeld & Sansom, 2011.). This step included conversion of water molecules, ions, lipids and protein chains. The resulting AT systems were energy minimised using 5,000 steps of steepest descent and equilibrated (duration of 1 ns) with a position restraint on protein Cα atoms (force constant of 10 kJ mol^{-1} $Å^{-1}$). Unrestrained production simulations were performed using different initial velocities with a time step of 2 fs and a total duration of 100 ns per simulation for TMD simulations (**Section 3.5.3**) and of 50 ns for signalling complex simulations (**Section 6.5.1**). A semi-isotropic pressure of 1 bar was applied using a Parinello-Rahman barostat (Parrinello, 1981) and the temperature was maintained at 323 K with a Berendsen thermostat (Berendsen et al., 1984). Electrostatic interactions were modelled using a particle mesh Ewald summation (cut-off distance of 10 Å, also used for van der Waals interactions; (Cerutti et al., 2009)). The bond lengths were restricted using the LINCS algorithm (Hess et al., 1997). GROMACS version 4.5.4 (Hess et al., 2008) was used in combination with the GROMOS96 43a force field (Scott et al., 1999) in TMD simulations

(**Section 3.5.3**) and with the Charmm27 force field (MacKerell et al., 2001) for signalling complex simulations (**Section 6.5.1**). Heparin parameters were generated using he SwissParam server (Zoete et al., 2011).

2.2 Molecular Biology

2.2.1 Complementary DNA and Expression Vectors

Complementary human DNA (cDNA) encoding Fibroblast Growth Factor 1 (UniProt accession code P05230), Fibroblast Growth Factor 2 (UniProt accession code P09038), Fibroblast Growth Factor Recptor 1 (UniProt accession code P11362) and Fibroblast Growth Factor Receptor 4 (UniProt accession code P22455) were purchased from Source BioScience (see **Table 1** for summary). For expression in mammalian HEK293S cells, multiple pHLsec vectors with different functions based on the pLEXm backbone (Aricescu et al., 2006) were used (see **Table 2** and **Appendix A.1** for details on the pHLsec vector). DNA inserts coding for a specific target sequence were typically inserted into the vector using the Age1 and the Kpn1 restriction sites (added to the DNA insert fragment via primers; see **Appendix C** for primer sequences). The pHLsec vector generates expression constructs with an N-terminal Kozak sequence, a start codon followed a secretion signal, an optional tag at the 5' end, followed by the target protein sequence with an optional C-terminal tag and the stop codon at the 3' end. The pHLsec and pHLsec-10H vectors were used for protein purification. The pHLsec-mV and the pHLsec-mTFP vectors were used for fluorescence microscopy experiments. For expression in *E.coli*, respective DNA constructs were inserted into a modified version of pET22b(+) (Invitrogen) which was a kind gift by Dr. Radu Aricescu (University of Oxford, Division of Structural Biology). He introduced an Age1 restriction site directly before the insert, and a Kpn1 site directly behind the insert in order to enable fast subcloning between mammalian expression vectors (pHLsec based vectors) and the bacterial expression vector. The pET22b(+) vector introduces a C-terminal hexahistidine tag whereas pET22b(+)2 bypasses the tag with an additionally introduced stop codon (see **Table 2.2**; and **Appendix A.2**)

Protein Name	**Gene Name**	**UniProt Accession Code**	**NCBI Reference Sequence**
Human Fibroblast Growth Factor 1	FGF1	P05230	NM_000800.2
Human Fibroblast Growth Factor 2	FGF2	P09038	NP_001997.5
Human Fibroblast Growth Factor Receptor 1c	FGFR1	P11362	NM_023110.2
Human Fibroblast Growth Factor Receptor 4c	FGFR4	P22455	NM_002011.4

Table 1 – Complementary DNA of genes used in this work

Name	**Tag**	**5' site**	**3' site**	**Tag cleavage**
pHLsec	C-term 6His	Age1	Kpn1	-
pHLsec-10H	C-term 10His	"	"	-
pHLsec-mV	C-term mV-8His	"	"	3C
pHLsec-mTFP	C-term mTFP-8His	"	"	3C
pHLsec-avi	C-term avi	"	"	-
pET22b(+)	C-term 6His	Age1	Kpn1	-
pET22b(+)2	-	"	"	-

Table 2 – Expression vectors for mammalian and bacterial protein expression

2.2.2 Construct Design and Polymerase Chain Reaction

A number of sources of information were used for the design of protein constructs that were produced in this study. The main sources were existing structural data entries in the Protein Data Bank (PDB), theoretical prediction of N-linked glycosylation sites (NetNGlyc 1.0 Server, CBS, Denmark, (Gupta & Brunak, 2002)), theoretical prediction of O-linked glycosylation sites (NetOGlyc 4.0 Server, (Steentoft et al., 2013)) theoretical recognition of secretion signal sequence (SignalP 4.0 server; (Petersen et al., 2011)) and domain boundaries listed on the UniProtKB protein sequence database. Generated constructs will be described in detail in the respective chapters. A full list of all oligonucleotides used in this thesis, in the polymerase chain reactions (PCR) is available in **Appendix C**. Primers were typically designed to have a length of 21-30 coding bases with a melting point between 60-63 ˚C and a GC content of around 60 % using the online tool Oligo Calc (http://www.basic.northwestern.edu/biotools/oligocalc.html). Flanking primers were extended with a restriction site for direct ligation of new constructs into the appropriate vector after the PCR reaction. All primers were ordered from Eurofins Genomics (Munich, Germany). PCR reactions were performed in volumes of 50 μl containing template DNA (1ng/μl), forward and reverse primer (0.2 pmol/μl each), 1x *Pyrobest*™ buffer II (Takara Co.), 0.4 mM dNTPs and 0.1 units *Pyrobest*™ DNA polymerase (Takara Co.), which is a thermostable DNA polymerase possessing 3'-5' exonuclease activity. PCR cycles were programmed in 6 steps as described in **Table 3** (**Saiki et al., 1988**).

Step	T [˚C]	Time	Process
1	96	90 s	Melting of template DNA
2	96	45 s	Denaturing: yielding single-stranded DNA
3	58	45 s	Annealing: primers anneal to template DNA
4	72	120-360 s	Elongation: 1-2 min for each kb
5	-	-	Repeat steps 2-4 30 x
6	72	10 min	Elongation: final step to fill all gaps

Table 3 – Standard protocol for a polymerase chain reaction

To introduce point mutations into constructs, site directed mutagenesis was performed in two steps, using internal primers containing altered codons. In the first PCR step two DNA fragments upstream and downstream from the mutation were amplified in two separate reactions. The reaction producing the N-terminal fragment contained the N-terminal forward primer paired with the reverse version of the mutagenesis primer, whereas the second reaction contained the forward mutagenesis primer paired with the C-terminal reverse primer. PCR products were gel purified using the QIAquick gel extraction kit (Qiagen) before using them for the second PCR step. In this step the N- and C-terminal fragments were used as mega primers, annealed to form the full DNA construct and amplified in PCR cycles without any additional flanking primers. The N- and C-terminal flanking primers were added directly into the reaction mix for amplification of the full construct with a standard PCR reaction as described in **Table 3**. Agarose gel electrophoresis was used to evaluate the outcomes of PCR reactions and all PCR products were gel purified.

2.2.3 DNA Restriction, Digestion, Ligation and Transformation

Vectors and PCR products were digested for two hours with the corresponding restriction enzymes (typically Age1 and Kpn1, New England BioLabs; see restrictions sites in **Appendix A**) using the recommended temperature and buffer conditions. Gel electrophoresis was performed to stop the reaction, to separate small DNA fragments from the target fragment and to check the outcomes of the digestion. Appropriate bands were isolated from the gel and purified with the QIAquick gel extraction kit (Qiagen). The Quick Ligation kit (New England Biolabs) was used to ligate the purified fragment into the respective vector using following protocol: 0.5 μl Quick ligase, 5 μl Quick ligase buffer, 3.5 μl DNA fragment, 1 μl vector were mixed and incubated at room temperature (RT) for 20 min. A transformation mixture was prepared adding 1 μl of the ligation product to 30 μl Library Efficiency DH5α Competent Cells (Invitrogen) before incubation on ice for 10 minutes. After a heat shock at 42 ˚C for 45 seconds, the transformed cells were incubated on ice for 5 minutes before adding 70 μl of SOC medium (Invitrogen). The reaction mix was incubated at 37 ˚C for 1 hour and then plated on LB-Agar plates containing 100 μg/ml ampicillin. After incubation over night at 37 ˚C, three colonies were picked of each construct for inoculation of 10 ml LB cultures, that incubated over night at 37 ˚C, containing 100 μg/ml ampicillin. The plasmid DNA of interest was then prepared using the QIAprep spin miniprep kit (Qiagen) according to the manufacturer's protocol. A digestion test was typically performed (using the same restriction enzymes that were used to digest the insert and vector before ligation) with subsequent agarose gel evaluation for identification of clones that contain the insert of expected size. These clones were subjected to sequence analysis (Source Bioscience) for final quality control of the Plasmid DNA sequence.

2.3 Protein Expression and Purification

2.3.1 Protein Expression in *E.coli*

All FGF constructs were expressed in the *E.coli* expression system. DNA inserts were ligated into a modified pET-22b(+) vector (Novagen) expressing a C-terminal 6xHis tag, using the Age1 and Kpn1 restriction sites. Plasmids were amplified with the QIAprep spin miniprep protocol (**Section 2.2.3**) and transformed into *E.coli* Rosetta 2 (DE3) pLusS (Novagen). This cell strain expresses rare tRNAs to enhance the expression of eukaryotic proteins. Five litre bacterial cell cultures in lysogeny broth (LB) medium containing 50 μg/ml ampicillin were grown to an optical density at 600 nm of 0.8, induced with 0.2 mM isopropyl β-thiogalactopyranoside (IPTG) (Melford Labs) and incubated in a shaker over night at 15 °C. At around 16 hours post-induction, the medium was spun down for 45 minutes at 3000 x g. Cell pellets were re-suspened in PBS with a complete protease inhibitor cocktail for expression in prokaryotic cells (Roche) and stored at -80 °C.

2.3.2 Small-Scale Protein Expression in HEK293S Cells and Western Blot Analysis

All FGFR constructs for *in vitro* experiments used in this thesis were expressed in an N-acetylglucosaminyltransferase I (GnTI) deficient cell line of 293S human embryonic kidney (HEK) cells. GnTI deficiency disables HEK293S cells to synthesize Asn-linked polysaccharides, which leaves posttranslational modification in the $Man_5GlcNAc_2$ glycosylation state. This limits heterogeneity across molecules and enables further enzymatic deglycosylation with endoglycosidase F1 (EndoF1) to a single N-acetylglucosamine residue, which is highly favourable for crystallisation (Reeves et al., 2001; Aricescu et al., 2006). Cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Sigma), supplemented with non-essential amino acids (Gibco), L-glutamine and 10 % fetal bovine serum (FBS, Invitrogen) in a humidified 37 ˚C incubator with 5 % CO_2. Small-scale expression and secretion trials in HEK293S cells were performed for all constructs prior to scaling up the expression. Approximately $3x10^5$ cells and 2 ml of DMEM containing 10 % FBS were seeded in 6-well plates (Corning) and grown to confluence of 80-90 %. Then media was removed and exchanged with 1.5 ml DMEM per well, containing 2 % FBS and supplemented with 500 μl transfection mix containing 2 μg plasmid DNA and 4μl transfection reagent lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Cells were incubated at 37 ˚C with 5 % CO_2 for 48 h prior to collection of cell suspensions for a check of secretion level of protein constructs by western blot analysis. In order to separate cell debris, the cell media was centrifuged at 15000 g for 10 minutes at 20 ˚C. Cells were resuspended in 1 ml PBS and for each construct 10 μl of cell suspension and 10 μl of medium were boiled in a reducing loading buffer for 5 min. These samples were subjected to sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) for protein separation according to molecular weight (Sambrook & Russell, 2000). BenchMark^{TM} pre-stained and hexa histidine tagged protein ladders (Invitrogen) were included on the SDS-PAGE as molecular weight markers. The protein bands from the SDS-PAGE were transferred onto a nitrocellulose membrane (GE Healthcare Life Sciences) at 30 V for 60 minutes using the XCell SuperLock^{TM} Mini-Cell in 10 % [v/v] methanol NuPAGE transfer buffer (Invitrogen). The nitrocellulose membrane was incubated in phosphate buffered saline (PBS) supplemented with 5 % fat-free milk powder for 10 minutes to reduce nonspecific binding of antibodies. Western blot analysis was performed using a mouse monoclonal anti-Penta-Histidine primary antibody (1:1000 dilution, Qiagen) and a goat anti-mouse IgG horseradish peroxidase conjugate (1:5000 dilution, Sigma) secondary antibody as previously described (Alegria-Schaffer et al., 2009). The enhanced chemiluminescence western blotting detection kit (ECL, GE Healthcare Life Sciences) was used for visualisation of His-tagged Protein bands.

2.3.3 Large-Scale Expression in HEK293S Cells

Only protein constructs that were identified to express, to fold correctly and to secrete efficiently in small scale trials were considered for large scale protein production which was typically carried out by Weixian Lu or Dr. Yuguang Zhao (University of Oxford, Department of Structural Biology) according to the standard method (Aricescu et al., 2006). Adherent HEK293S cells were grown in expanded surface polystyrene roller bottles (2125 cm2, Greiner Bio-One) to 90 % confluence and transfected with a mixture of 2 mg endotoxin free plasmid DNA (produced with Plasmid Mega or Giga Kits, Qiagen) and 4 mg polyethyleneamine (PEI, Aldrich) per litre of cell media. Roller bottles were typically incubated for 7-10 days at 37 ˚C with 5 % CO_2. Medium was then clarified by centrifugation at 7000 g for 40 min, filtered using SteritopTM filter units (0.22 μm pore size, sterile) and stored at 4 ˚C. Prior to purification media were typically buffer exchanged and concentrated using a QuickStand system (GE Healthcare Life Sciences) according to the manufacturer’s protocol.

2.4 Protein Purification

2.4.1 Immobilised Metal Affinity Chromatography

Most FGF and FGFR protein constructs were expressed with a C-terminal polyhistidine tag (6xHis, 8xHis or 10xHis). These proteins were purified in a two-stage purification protocol, starting with immobilised metal affinity chromatography (IMAC) followed by a size exclusion chromatography step.

E.coli cells were defrosted and lysed by a French press (Cell Disruption Solutions, Constant Systems), centrifuged at 20,000 rpm for 45 minutes at 4 ˚C and filtered with a 45 μm syringe filter (Appleton Woods). The protein supernatant was directly loaded onto a 5 ml Ni-loaded chelating sepharose HisTrapTM column (GE Healthcare Life Science) pre-equilibrated in PBS (pH 7.5). To limit unspecific protein binding, medium that contained protein expressed in HEK293S cells was supplemented with 20 mM imidazole prior to application to the Ni-loaded chelating sepharose column. The column was typically washed with 10-20 column volumes PBS (pH 7.5) containing 40 mM Imidazole. Protein elution was performed with PBS (pH7.5) containing 400 mM Imidazole. During elution fractions of 1-1.5 ml were collected using an Äkta high performance liquid chromatography (HPLC) system (GE Healthcare). SDS-PAGE was used to check protein composition in eluted fractions. Fractions containing protein of appropriate size and purity were pooled and concentrated to a concentration between 25-50 mg/ml using Amicon-Ultra centrifugal filter units (Millipore).

2.4.2 Analytical Heparin Affinity Chromatography

In order to improve small angle X-ray scattering (SAXS) data quality and to favour crystallisation by reducing flexible regions, one FGF1 construct (FGF1-P241) was cloned into the pET-22b(+)

vector lacking the 6xHis tag. Cells expressing FGF1-P241 were lysed as described in **Section 2.4.1** and the protein supernatant directly applied to a 5ml HiTrap™ heparin HP column (GE Healthcare Life Science) pre-equilibrated in a low salt buffer (50 mM HEPES pH 7.5, 50 mM NaCl). The column was washed with approximately 20 column volumes low salt buffer and eluted with a sodium chloride gradient from 50 mM to 2M over 20 column volumes. The eluted volume was collected in 1.5 ml fractions and analysed by SDS-PAGE to identify protein size and purity.

2.4.3 Size Exclusion Chromatography

As a second purification step after IMAC or analytical heparin affinity chromatography, gel filtration was performed to achieve higher protein purity and to bring the protein into the appropriate buffer. Depending on the protein size μl of protein solution with a protein concentration of 25-50 mg/ml was injected onto either a Superdex S75, a HiLoad S200 HR10/30 or a Superdex S200 16/60 size exclusion chromatography column (Pharmacia Biotech). The respective column was previously equilibrated in the running buffer containing 10 mM HEPES (pH 7.5) and 150-300 mM NaCl. Protein fractions were analysed by SDS-PAGE in order to pool and concentrate the fractions containing pure protein sample. The pure protein samples were optionally deglycosylated by enzymatic treatment with EndoF1 (Chang et al., 2007; Grueninger-Leitch et al., 1996) before subjection to complex formation, crystallisation and further biophysical methods described below.

2.5 Multi-Angle Laser Light Scattering

The combination of size exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) was applied to determine the molar mass of single proteins and protein complexes in different states of dissociation. Concentrated purified protein samples prepared as described in **Section 2.4** were diluted in MALLS running buffer (10 mM Hepes pH 7.5, 150-300 mM NaCl) and samples of 90 μl containing typically 3 nM of protein were applied to SEC-MALLS analysis. For protein complex samples, exact amounts of all complex components were mixed, diluted in MALLS running buffer to achieve a final concentration of 3 nM of the FGF component and 90 μl of the sample were directly injected to SEC-MALLS analysis. Size exclusion chromatography was performed on a 30 ml Superdex S200 HR 10/30 column (GE Healthcare) attached to an Agilent chromatography system. A Dawn Helios II Multi-Angle Light Scattering detector, using a 120 mW solid-state laser operating at 658 nm as a light source, and an Optilab rEX Refractive Index detector the light scattering and the refractive index of the samples upon elution from the size exclusion column. The recorded data were analysed with the ASTRA5 program suite (Wyatt Technology). Experiments were performed with assistance of Dr. Tomas Malinauskas (University of Oxford, Division of Structural Biology).

2.6 Protein Crystallisation

2.6.1 Initial Crystallisation Screening

Different strategies were applied to achieve protein complex formation and subsequent protein complex crystallisation of the FGFR4-Heparin-FGF1 and the FGFR1-Heparin-FGF1 complexes. Heparin oligosaccharides of selected lengths (Iduron), concentrated protein from appropriate SEC fractions of FGFR4 or FGFR1 and >95% pure FGF1 (tested with SDS-PAGE) were mixed in different stoichiometric ratios to form a pentameric (2:1:2) or a hexameric (2:2:2) complex. The complex was either applied to SEC purification or directly applied to crystallisation trials. EndoF1 was optionally added directly to the protein complex sample previous to crystallisation, in order to cleave off the N-glycans that were added to the receptor during expression in HEK293S cells.

Pre-crystallisation experiments using the PCT™ Pre-crystallisation Test kit (Hampton Research) were used for each new protein complex preparation in order to determine an appropriate protein concentration for crystallisation. For crystallisation trials in a sitting drop vapour diffusion format, a Cartesian Technologies robot (Walter et al., 2005) dispensed 100 nl reservoir solution to 100 nl protein complex solution in 96-well Greiner plates. The drops were equilibrated against 95 μl reservoir solution in sealed chambers and the plates were placed in temperature-controlled storage vaults at either 21˚C (TAP Homebase) or 4˚C (Formulatrix). Up to 1150 different reservoir solutions from commercially available crystallisation screening kits (see **Table 4**) were tested for initial crystallisation trials. Crystallisation conditions for FGFR4-HeparinDP10/12-FGF1 and FGFR1-HeparinDP10/12-FGF1 are provided in the corresponding chapters.

	Masterblock	**Commercial Screen**	**Wells**	**Manufacturer**
1	Block 1	Crystal Screen	A1-D12	Hampton Research
		Crustal Screen 2	E1-H12	
2	Block 2	Wizard 1	A1-D12	Emerald Biosystems
		Wizard 2	E1-H12	
3	Block 3	PEG/Ion Screen	A1-D12	Hampton Research
		PEG 6000 Grid Screen	E1-H12	
		Ammonium Sulphate Grid Screen	G1-H12	
4	Block 4	Natrix Screen	A1-D12	Hampton Research
		Crystal Screen Cryo	E1-H12	
5	Block 5	PEG/LiCl Grid Screen	A1-B12	Hampton Research
		NaCl Grid Screen	C1-D12	
		MPD Grid Screen	E1-F12	
		Quick Screen	G1-H12	
6	Index	Index Screen	A1-H12	Hampton Research
7	SaltRX	SaltRX Screen	A1-H12	Hampton Research
8	PACT	PACT Premier	A1-H12	Molecular Dimensions
9	PEGRx	PEGRx Screen	A1-H12	Hampton Research
10	Morpheus	Morpheus Screen	A1-H12	Molecular Dimensions
11	PGA	PGA Screen	A1-H12	Molecular Dimensions
12	ProPlex	ProPlex Screen	A1-H12	Molecular Dimensions

Table 4 – 1152 commercially available crystallisation conditions combined in 12 screens

2.6.2 Crystal Optimisation

Poorly diffracting crystals obtained from initial crystallisation screens were optimised in order to collect high resolution X-ray diffraction data. A number of different methods for crystal optimisation were used. Hampton Additive Screens (Hampton Research) contain volatile and non-volatile compounds, which were added to the original crystallisation condition. The range of additive compounds was extended by using the Silver Bullets additive screen (Hampton Research). Manual optimisation was performed systematically by varying the pH, the concentration of components and the ratio of protein solution to reservoir solution. The latter two were varied in a three-row optimisation screen (Walter et al., 2005). Each row had a specific protein to reservoir ratio (1:1, 2:1, 1:2) and the reservoir solution was diluted within each row in 3% steps from 100% to 67%. The final crystallisation conditions and corresponding cryo protectant for protein structures presented in this thesis are presented in **Table 5**.

Construct	**Space Group**	**Protein Concentration**	**Crystallisation Condition**	**Temp (˚C)**	**Freezing Procedure**
FGFR1 D2D3-HepDP10-FGF1	C2	6 mg/ml	0.1 M amino acid 0.1 M imidazole MES pH 6.5 30 % PEGMME 550 PEG 20K	20 ˚	25% Glycerol
FGFR1 D2D3-HepDP10-FGF1	$P2_12_12_1$	6 mg/ml	0.1 M Tris pH 7.0 4 M NaCl	4 ˚C	25% Glycerol

Table 5 – Crystallisation conditions and freezing procedure for presented structures

2.6.3 Crystal Cryo-Protection and X-ray Diffraction Data Collection

Crystal deterioration by radiation damage was avoided by X-ray data collection in cryogenic conditions of 100 K (Garman, 2010), which requires cryoprotection. Crystals were fished from the mother liquor solution with cryo-loops and kept there until data collection. Depending on the crystallisation condition crystals were incubated in crystallisation solution supplemented with cryoprotectants such as ethylene, glycerol or low molecular weight polyethylene glycol (PEG) in order to avoid formation of water crystals during flash freezing in liquid nitrogen. Screening of crystal diffraction properties and data collection was performed at Diamond Light Source (DLS; Oxfordshire, UK). Well diffracting crystals were indexed by the automatic image processing pipeline at DLS. Before collecting a complete diffraction data set, a collection strategy suitable for the respective space group was programmed (specifying initial crystal orientation to beam, beam intensity, exposure time per frame, degree of oscillation per frame and number of frames). All data were collected using a single-photon-counting PILATUS detector.

2.6.4 X-ray Diffraction Data Processing

Typically three representative diffraction images, recorded at oscillation ranges separated by 90°, were indexed to assign Miller indices (hkl) to each recorded reflection and to determine the space group as well as the unit cell dimensions of the crystal lattice. Then, the reflection intensities from the entire dataset were integrated, merged and scaled using the Mosflm (Leslie, 2006) or xia2 (Winter, 2010) data processing suites. In order to assess the high resolution limits as well as the data quality several criteria were used which are briefly described below.

1) Completeness (%). The indexing process allows determination of the number of expected unique reflections. Completeness is the percentage of reflections, which a dataset contains of the total expected reflections (**Formula 1**). The ideal completeness in the highest resolution shell was set at >90 % and for overall completeness at >95 %.

2) Redundancy of a dataset describes the ratio of total measured reflections to unique measured reflections (**Formula 2**). A higher redundancy enhances the accuracy of the intensity for each measured reflection.
3) I/σI is a signal to noise parameter in crystallography which describes the ratio of signal intensity to the error in measurement of the signal intensity. The I/σI value typically decreases with increasing resolution. This parameter is also used to determine the high-resolution limit for the dataset. Error model adjustment to a χ^2 value (goodness of fit, **Formula 3**) of approximately 1 for all resolution shells of a dataset resulted in reliable estimates of σI.
4) R_{merge} (%) is a linear measure of the internal consistency of the dataset, which represents the deviation between equivalent measurements (**Formula 4**). For the entire dataset R_{merge} values of less than 10 % and for the highest resolution shell less than 100 % were targeted. However, this parameter does not account for the redundancy of the dataset. R_{pim} (the precision-indicating merging R-value) was used for highly redundant data. R_{pim} corrects for the increasing precision of the intensity measurements as more measurements are made (**Formula 5**). Pros and cons of these two R-values are further discussed by Manfred S. Weiss (French & Wilson, 1978; Weiss, 2001).

For each dataset, structure factor amplitudes were calculated from the reflection intensities using TRUNCATE (French & Wilson, 1978), which is part of the CCP4 software package (Winn et al., 2011). The intensity of a reflection is proportional to the square of the amplitude of the structure factor (**Formula 6**). However, in order to determine the phase for each structure factor, molecular replacement was required in this work.

2.6.5 Structure Determination

The so-called "phase problem" for the crystal structures presented in this thesis has been solved with the method called molecular replacement (MR) (Rossmann & Blow, 1962) using PHASER (McCoy et al., 2007), which is part of the CCP4 suite. The MATTHEWS_COEF program (Matthews, 1968) was used for initial estimation of the number of protein molecules in the asymmetric unit. The choice of suitable MR models will be discussed in the results **Section 5.5.2**. Ligands and waters were removed from these models and molecules were split into their single domain components. Each domain component was an input model for a separate rotation search (**Formula 7**) against the experimental data. Solutions with top scores were used for a translation search (**Formula 8**) and top scoring solutions were evaluated for clashes between symmetrically related molecules. The solutions passing this test were subjected to rigid-body refinement before providing the final solution statistics, models and maps.

Three criteria were applied to test for successful molecular replacement solutions. The **Log-likelihood gain** (LLG) is a quantitative measure for the fit of the molecular replacement to the experimental data compared to the same set of atoms in a random distribution. The LLG value describes the difference between the calculated log-likelihood derived from the model and the log-likelihood derived from a Wilson distribution, which is a probability distribution function for structure factors of known atom types but unknown atom positions. Valid solutions do typically result in positive LLG values increasing with each cycle of the molecular replacement procedure. The **Z-score** is the second validation criterion for molecular replacement solutions. It measures the number of standard deviations above the mean LLG score for a particular LLG value. Valid solutions should have a positive Z-score greater than 7, which increases with each cycle of the molecular replacement procedure. The third and clearest evaluation is finally possible by visual inspection of the produced **electron density map**. The solution is likely to be correct if the electron density map features unmodelled electron density for missing domains, additional N- and C-terminal amino acids to the modelled domains and potential ligands.

2.6.6 Structure Refinement

Phasing solutions from molecular replacement that satisfied the criteria described in **Section 2.6.5** were subjected to iterative model building in coot (Emsley & Cowtan, 2004) and restrained refinement using REFMAC (Murshudov, Vagin, & Dodson, 1997) and PHENIX (Adams et al., 2010). This process was guided by geometrical constraints, atomic positions, non-crystallographic symmetry and B-factor distributions. R_{work} (**Formula 9**), R_{free} (**Formula 10**) and the quality of the resulting map were closely monitored during the refinement process. The gap between R_{work}, which represents the deviation of the model structure factors from the structure factors of the data using only 95 % of the reflections, and R_{free}, which is similar to R_{work} but calculated from only 5 % of randomly assigned reflections that are excluded from the refinement process, is a quantitative measure for introduction of bias during the manual model building processes. The difference for final refined models should not be higher than 4-5 %. Both R-values do typically drop if the refinement process is leading to a model that is consistent with the measured structure factors.

2.6.7 Structure Validation and Analysis

MolProbity (Davis & van der Merwe, 2006) assesses the stereochemical properties of structural models and was used for structural validation of refined models. The Ramachandran analysis plots the ϕ angle (the dihedral angle between N and Cα) against the ψ (the dihedral angle between Cα and C=O) for each protein residue. Theoretical calculations and experiments have identified “allowed” and “prohibited” combinations of ϕ and ψ, which can be directly identified on the Ramachandran plot. The percentage of residues that are not within the allowed regions, so-called outliers, is an independent measure for the quality of model geometry, particularly as ϕ- and ψ- angles are not restrained during refinement. MolProbity also analyses any other steric clashes formed by H-bond and van der Waals contacts in molecular interfaces. Refinement and validation statistics are presented in the corresponding chapters. Figures were rendered in PyMol (DeLano Scientific LLC) or CHIMERA (Pettersen et al., 2004).

2.7 Small-Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) experiments were performed to assess the behaviour of protein complexes in solution. Protein complex samples were prepared directly before subjecting them to the beamline for collection of scattering data on the BM29 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at 20° C. The distance between the detector and the sample was 1 m and the beam had a wavelength of 0.0995 nm. During SEC protein purification, buffer fractions were collected after protein elution. This buffer was measured before and after every SAXS measurement in order to minimise the background noise after buffer subtraction. Each sample was measured in 10 consecutive frames of 10 seconds at 20° C. The sample flow was adjusted according to the degree of radiation damage. Only measurements that did not show any radiation damage through all 10 frames were used for analysis. Initial processing was performed directly at the beamline using the automatic pipeline of AUTOSUB (Pernot et al., 2013). GNOM (Semenyuk & Svergun, 1991) was used to yield the distance distribution function (P(r)) and to evaluate the maximum particle dimension (D_{max}). Ab initio reconstructions were generated by the data processing pipeline available at BM29 using DAMMIF (Franke & Svergun, 2009) and DAMAVER (Volkov & Svergun, 2003). Molecular Dynamics simulations of different FGFR-Heparin-FGF complex architectures were performed (using GROMACS, see **Section 2.1.2**) for creation of structural ensembles, which resulted in a pool of about 2020 models. Flexible glycans and missing loops were modelled in combination with energy minimisation using the AllosMod-FoXS server (Guttman, Weinkam, Sali, & Lee, 2013). CRYSOL (Svergun et al., 1995) was used to calculate theoretical SAXS patterns and ensemble fitting was performed using GAJOE (Bernadó et al., 2007).

2.8 Sequence Alignments

Sequence alignments were performed using ClustalW (Larkin et al., 2007), WebLogo (Crooks et al., 2004). For visualisation of sequence alignments Jalview (Waterhouse et al., 2009) and Microsoft Word were applied. Predictions of secondary structure and transmembrane regions were performed using the SPLIT-server 4.0 (Juretić et al., 2002) and MPEx (Snider et al., 2009) and the Uniprot server (The UniProt Consortium, 2015).

2.9 Visualisation of Graphs

Gnuplot 4.2, Xmgrace and Microsoft Office Excel were applied for display of data.

Chapter 3

Investigation of wild-type and mutant FGFR3 and FGFR4 TMD dimerisation

Chapter 3 – Investigation of wild-type and mutant FGFR3 and FGFR4 TMD Dimerisation

3.1 Summary

Fibroblast growth factor receptors form a subfamily of the receptor tyrosine kinases that are involved in signalling through the plasma membrane. Extracellular ligand binding leads to lateral dimerisation and receptor activation to stimulate cellular proliferation and differentiation. The transmembrane domains play a crucial role in dimerisation processes and several mutations located in the lipid bilayer have been identified to stabilise FGFR dimers and thus to promote unregulated signalling.

In this work, coarse-grained and atomistic molecular dynamics (MD) simulations are used to assess the role of diverse mutations in the TMDs of FGFR3 and FGFR4. Simulations show that the FGFR3 G380R mutant, which causes Achondroplasia, has no severe impact on the helix dimer interface but induces a shift towards the extracellular matrix, which could be confirmed with atomistic simulations. The two analogous FGFR3 Y373C and FGFR4 Y367C mutants, causing a lethal type of dwarfism and involved in breast cancer progression respectively, do neither interfere with the dimer interface nor induce a helix shift. This suggests that the Cys residues activate receptors by formation of disulphide bonds via the “flip sides” of the TM domains. Moreover, this work presents the first FGFR4 wild-type (wt) and G388R TM domain dimer model, which reveals helix-helix association via a G_{388}xxxG motif. Arg388 entirely abrogates this binding interface setting both C-termini apart from each other but maintaining the helix-helix orientation by shifting the main binding interface in N-terminal direction. The presented results further knowledge about the role of disease related mutations in RTK TM domains and provide new insights into possible mechanisms of abnormal FGFR signalling.

3.2 Introduction

As discussed in **Section 1.3** FGF signalling plays an essential role in a multitude of tightly regulated biological processes such as embryonic development, cell proliferation, differentiation and migration (Beenken & Mohammadi, 2009). FGFR activation is not solely regulated via ligand binding but also dependent on juxtamembrane autoinhibition (Hubbard, 2004; Mohammadi et al., 1996), down-regulation of active receptors and proper dimerisation of the transmembrane helices to bring the kinase domains in the right spatial arrangement. Deregulation of these processes can lead to diverse human diseases including skeletal disorders as well as cancer (Wesche et al., 2011).

The role of single TM domains in RTK signalling is only partly understood but a list of pathogenic mutations in these domains provides strong supporting evidence that TM helices play a critical role in receptor dimerisation processes and thereby in cell signalling of RTKs (Li et al., 2006). While all TM domains of human RTKs show an intrinsic propensity to form stable dimeric structures within a membrane (Finger et al., 2009), their affinity must be inherently weak in order to allow regulation (Peng et al., 2009). Mutant dependent alterations of the monomer-dimer equilibrium can lead to constitutive activation or diminished receptor turnover (Turner & Grose, 2010). Some prominent TMD mutations of FGFR3 and FGFR4, which this work focuses on, were subject to various studies and provided initial insights into the different functional effects of single TM amino acid substitutions. However, molecular mechanisms remain elusive.

3.3 Aims

The main aim of the work presented in this chapter is to identify the homodimer interfaces of wild-type FGFR3 and 4 TMDs through course-grained molecular dynamics simulations. The next step will be to run simulations with mutant TMDs to observe possible changes of the primary interfaces hinting on pathogenic mechanisms.

3.4 Sequence analysis and simulation setup

3.4.1 Sequence alignments of FGFR3 and FGFR4 TMDs

Various sequence alignment tools were applied to identify which atomistic FGFR3 and FGFR4 TMD models to build as initial α-helical constructs for coarse-grained MD simulations. Sequence identity calculations across human FGFR1-4 show that all members across this receptor family have high sequence identities (**Figure 3.1 A**) ranging between 59.5 % (FGFR2/FGFR4) and 71.1 % (FGFR1/FGFR2). However, sequence identities calculated for the same receptors but from limited 45 residue sequences of the predicted TMD regions, including flanking unstructured residues, reveal much lower values. The TMD regions of FGFR3 and FGFR4 have only 17.4 per cent sequence identity. Direct sequence comparison of these 45 residue regions reveal a well conserved hydrophobic core (**Figure 3.1 B**) flanked by an N-terminal extracellular acidic patch and a C-terminal intracellular basic patch, which are indeed an inherent feature of all eukaryotic single pass transmembrane proteins (George Hedger, 2015). However, the conservation score shows that these flanking regions are less conserved compared to the hydrophobic core. Overall only four residues are conserved across all four receptors; three of those residues are located in the hydrophobic core. The fourth conserved residue is a tyrosine, which is positioned in the extracellular membrane-water interface. Interestingly, the N-terminal acidic patch, which holds the TMD in correct position (Killian & Heijne, 2000), varies in charge and localisation. FGFR1, 2 and 4 have the first acidic residue (Glu, Glu and Asp respectively) two residues downstream from this conserved tyrosine (red arrow in **Figure 3.1 B**). The first acidic residue of FGFR3 sits 5 residues upstream from the conserved tyrosine (black arrow in **Figure 3.1 B**). This could have implications for the behaviour of the TMDs in MD simulations and needs to be taken into account when choosing the TMD models for simulations. The position of the intracellular basic patch is well conserved and shows a very similar pattern of residues across all four receptors.

Transmembrane domain prediction tools gave different results for FGFR3 and FGFR4. UniProt (The UniProt Consortium, 2015) TM-prediction identified regions of 21 residues, which were perfectly aligned across both receptors and located between the negatively and positively charged pattern. The MPEx TM-prediction tool (Snider et al., 2009) predicted a very similar TMD for FGFR4 consisting of only 19 residues (see highlighted sequences in grey, **Figure 3.1 C**). In contrast the TMD prediction for FGFR3 is shifted 5 residues towards the extracellular region. This indicates that the TMD of FGFR3 could behave differently in the membrane, in comparison to the other FGF receptors, potentially caused by the shift of the extracellular acidic patch towards the N-terminus. If the MPEx TMD prediction were to be accurate, the linker between the extracellular domain D3 (see sequences highlighted in cyan) and the TMD would be 5 residues shorter in FGFR3. This again could lead to a more direct coupling of positional changes in the extracellular domains to the TM domains.

Taking these results together, TMD models were used that conform to the following rules:

a) Keep the average of both predicted TMD regions in the centre.
b) Encompass all known positions of pathogenic TMD mutations (see red residues in **Figure 3.1 C**).
c) Encompass two acidic residues from acidic patch.
d) Encompass at least one basic residue from basic patch.

A

FGFR FL sequence identity matrix

	R1	R2	R3	R4
R1		71.1	65.5	60.5
R2	71.1		70.1	59.5
R3	65.5	70.1		62.5
R4	60.5	59.5	62.5	

FGFR TMD sequence identity matrix

	R1	R2	R3	R4
R1		52.2	28.3	17.4
R2	52.2		30.4	17.0
R3	28.3	30.4		17.4
R4	17.4	17.0	17.4	

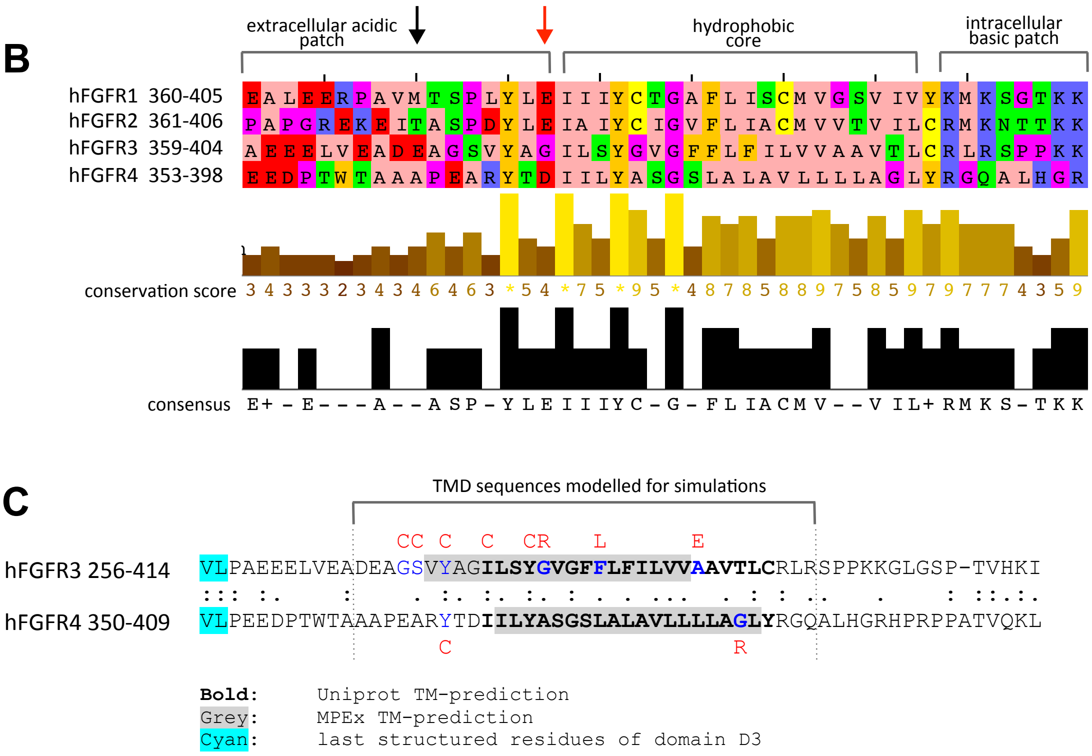


Figure 3.1 Sequence alignment and TMD prediction

(A) Sequence identity matrices across FGFR family members for full length receptors (left panel) and for extended TMD regions of 45 residues (right panel). Pairs with highest conservation are highlighted green and pairs with lowest conservation are highlighted red. Sequence conservation values between FGFR3/FGFR4 are framed blue. All sequence pairs were aligned and analysed using the CustalW program (Larkin et al., 2007). **(B)** Sequence alignment of extended homologous TMD regions of FGFR1-4 (Colour code: pink=aliphatic or hydrophobic; orange=aromatic; blue=basic; red=acidic; green=hydrophilic; magenta=conservationally special; yellow=cysteine). The red arrow indicates the acidic residue for FGFR1, 2 and 4 which is located most proximal to the hydrophobic core. The black arrow indicates the same, but for FGFR3. The middle panel presents the conservation score calculated by hierarchical analysis (Livingstone & Barton, 1993). The bottom panel shows the Jalview consensus sequence. The sequence alignment was done in ClustalW and the representation was generated using Jalview (Waterhouse et al., 2009). **(C)** Transmembrane domain predictions for FGFR3 and FGFR4, performed by UniProt ((The UniProt Consortium, 2015); bold residues) and MPEx (Snider et al., 2009); residues highlighted in grey). Residues that have been related to human pathologies are highlighted in dark blue colour and the corresponding mutant residues are indicated with red letters. Residues highlighted in cyan indicate the last structured residues of the respective D3 domains.

3.4.2 Model building and simulation setup

Table 6 lists the TMD sequences, which were used for modelling of initial, ideal α-helical constructs. Bold letters represent the TMD prediction from the UniProt server. Both Tyr to Cys substitutions are located at the same position close to the N-terminal water-membrane interface. In contrast, the Gly to Arg substitution in FGFR3 is located closer to the TMD centre and in the case of FGFR4 rather on the C-terminal water-membrane interface. For coarse-grained simulations TMDs and their flanking residues were assumed to form an ideal α-helix. Atomistic starting structures were built in PyMOL (**Figure 3.2 A**), converted to CG models (**Figure 3.2 B**) with an approximate 4:1 mapping of heavy atoms to coarse-grain particle as previously described (Bond et al., 2008; Bond & Sansom, 2006) and used to set up simulations using the GROMACS 4.5.4 package (Van Der Spoel et al., 2005). **Figure 3.2 C** shows the atomistic and the corresponding coarse-grained model of POPC that was used for all simulations. **Table 7** lists the composition of all coarse-grained simulations that were setup in this study. Simulation system number 2 contained 80 % POPC and 20 % POPS and was used as a control for TMD behaviour in presence of lipids with partially negatively charged headgroups. For a detailed description of the simulation setup procedure and simulation parameters see **Section 2.1.1**. An ensemble of 20 simulation repeats of 1 μs was run for each simulation system listed in **Table 7**.

Receptor	Mutation	TM sequence
FGFR3	wt	D$_{367}$EAGSVYAG**ILSYGVGFFLFILVVAAVTLC**RLR
	G380R	D$_{367}$EAGSVYAG**ILSYRVGFFLFILVVAAVTLC**RLR
	Y373C	D$_{367}$EAGSV**C**AG**ILSYGVGFFLFILVVAAVTLC**RLR
FGFR4	wt	A$_{361}$APEARYTD**IILYASGSLALAVLLLLAGLY**RGQ
	G388R	A$_{361}$APEARYTD**IILYASGSLALAVLLLLARLY**RGQ
	Y367C	A$_{361}$APEAR**C**TD**IILYASGSLALAVLLLLAGLY**RGQ

Table 6 – Sequences of the TMD models used for MD simulations

List of protein sequences applied in CG-MD simulations. The regions predicted to be covered by the hydrophobic membrane core are bold. Selected amino acid positions, which were identified to be related to diseases are highlighted in blue. The mutated residues are highlighted in red.

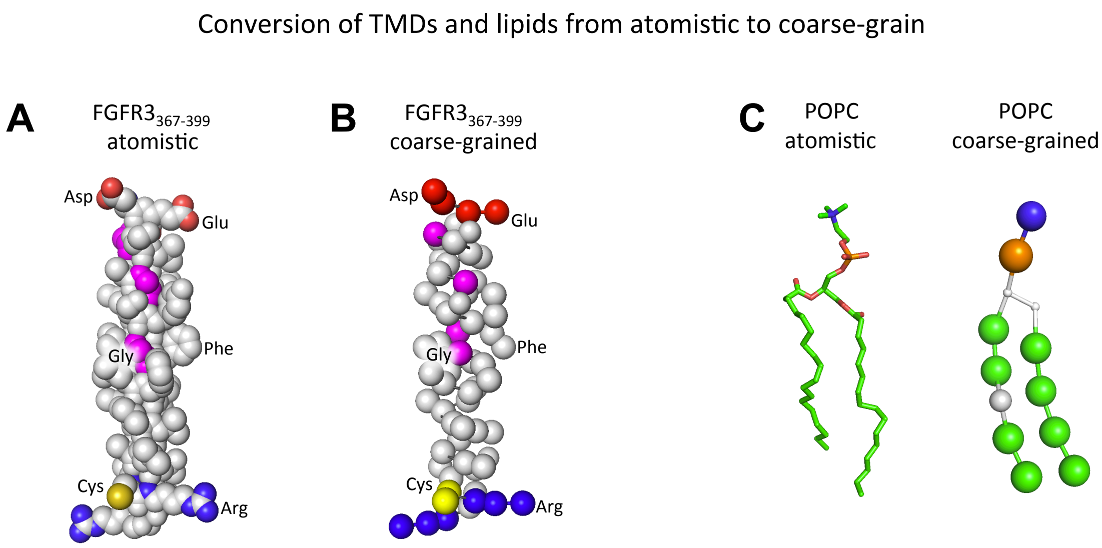


Figure 3.2 Conversion of atomistic TMD and lipid models to coarse-grained models

(**A**) An atomistic model of FGFR$_{3367-399}$ generated in PyMOL, shown as representative for all 6 TMD models, which were created from the TM sequences listed in **Table 6**. (**B**) The same α-helix as in **A**, after conversion to coarse-grain. In **A** and **B**, all residues are shown in spheres representation, N-terminal acidic aspartate and glutamate residues are coloured red, glycines are coloured magenta, cysteines are coloured yellow and C-terminal positively charged arginine residues are coloured in blue. (**C**) An atomistic POPC molecule shown in sticks representation and the corresponding coarse-grained model shown in spheres representation. Sulphates are coloured blue and nitrogens are coloured in orange.

System name	System components	n simulations forming helix dimer	Particle number
1 FGFR3 wt	2 $FGFR3_{367\text{-}399}$ wt + 319 POPC	19 of 20	9028
2 FGFR3 $wt_{POPC/PS}$	2 $FGFR3_{367\text{-}399}$ wt + 256 POPC + 64 POPS	18 of 20	10401
3 FGFR3 G380R	2 $FGFR3_{367\text{-}399}$ G380R + 323 POPC	16 of 20	9403
4 FGFR3 Y373C	2 $FGFR3_{367\text{-}399}$ Y373C + 317 POPC	16 of 20	8935
5 FGFR4 wt	2 $FGFR4_{361\text{-}393}$ wt + 317 POPC	20 of 20	8940
6 FGFR4 G388R	2 $FGFR4_{361\text{-}393}$ G388R + 318 x POPC	20 of 20	9081
7 FGFR4 Y367C	2 $FGFR4_{361\text{-}393}$ Y367R + 319 x POPC	20 of 20	8894

Table 7 – Summary of coarse-grained simulation systems.

20 simulation repeats of 1 μs duration were run for each simulation system. Apart from the simulation system 2 (FGFR3 $wt_{POPC/S}$) all simulations were run in a lipid bilayer formed of 100% POPC. System 2 contained 80% POPC and 20% POPS in order to control and validate the quality of simulations run in 100% POPC.

3.4.3 Helix distance tracking and dimerisation time

Visual inspection of a finished simulation typically gave a first impression about membrane and helix behaviour. The four frames presented in **Figure 3.3 A** were extracted from a representative FGFR3 wt dimerisation simulation. At t = 0 μs both helices are embedded in the self-assembled lipid bilayer at a distance of 70 Å. Through the course of the simulation both helices dissociate without constraints in a periodic box, which eventually leads to formation of a helix dimer (see t = 0.4 μs in **Figure 3.3 A)**. The first analysis performed for all simulation within an ensemble was quantitative tracking of the distance between both helices in each trajectory, which yields information on a number of TMD properties. The assessment of the duration until dimer formation is the most obvious. If dimer formation occurs quickly for a given TMD construct, the tracked separation between both helices will decrease sooner after simulation start. These measurements were performed by adopting MDAnalysis-dependent (Michaud-Agrawal et al., 2011) Python functions (Reddy et al., 2014)) to my data. The helix distance was defined by the closest interhelix approach of two Cα particles from each peptide. A distance matrix for all interhelical Cα particle interactions was calculated for each 10th frame of a given simulation. The Minimum value for each frame was recorded as closest distance and plotted against the simulation time in μs. The difference between the starting time of the simulation and the time of first contact of both helices was defined as the dimerisation time (see thin arrow in **Figure 3.3 B**). The simulation time of 1 μs per repeat is relatively short. Therefore not all simulations formed a TMD dimer (for helix distance plots of individual simulations see **Appendix E, 1-6**). As previously seen for Glycophorin A (GpA) (Psachoulia et al., 2008) there was stochastic variation in the time required for initial dimer formation corresponding to the use of random initial velocities (consistent with a Maxwellian distribution at the stated temperature) for each simulation within an ensemble. The average dimerisation time was calculated for each ensemble. Plotted on a histogram (**Figure 3.3 C**), there is no significant difference in the time to dimer formation among

FGFR3 and FGFR4 wt and mutant simulations. Much larger ensembles (typically of ≥100 simulations) would be needed to explore any differences in dimerisation times more accurately (Hall et al., 2014; Wassenaar et al., 2015). In many simulations, both helices rotate and move substantially in the x-y-plane before dimer formation. Thus, the stochastic dimerisation times and substantial translational and rotational motions of the helices relative to one another within the bilayer plane are consistent with a simulation protocol that is not biased toward formation of a particular dimer interface. Tracking of the helix distance also provides information on the stability of the dimers. Dimer dissociation events can easily be identified using the Cα closest approach method. Both the FGFR3 and the FGFR4 formed stable dimers in all CG simulations with no observations of subsequent dissociation of the helices.

Having confirmed the formation of stable dimers in an unbiased fashion in our simulations, separate sub-chapters for FGFR3 and FGFR4 TMDs will deal in more detail with dimer configurations and interfaces.

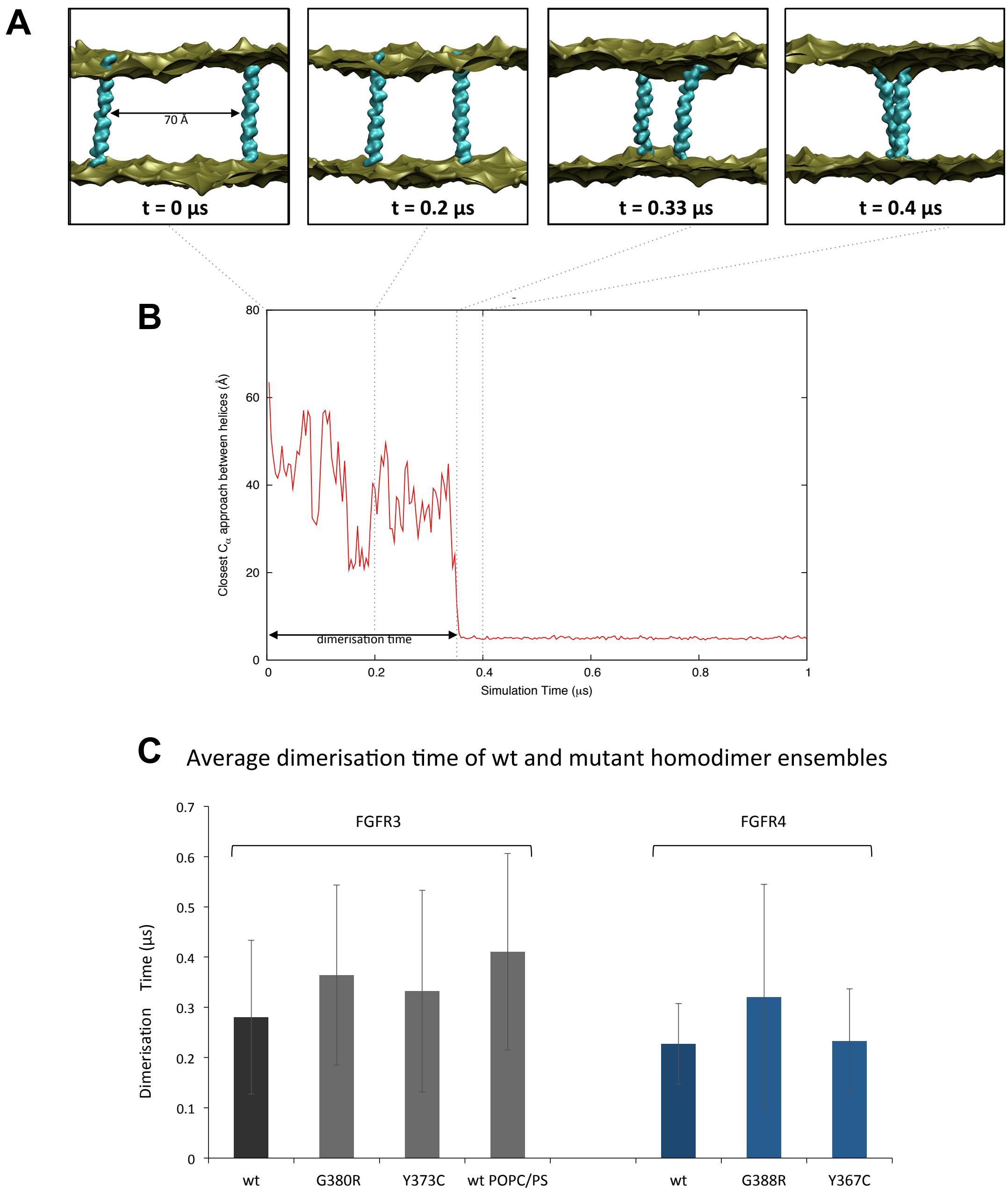


Figure 3.3 Helix distance tracking and dimerisation time

(**A**) Snapshots from a wt FGFR3 CG simulation showing TMD configurations at different times (t = 0 µs, t = 0.2 μs, t = 0.33 μs, t = 0.4 μs). The α-helices are shown as Cα traces in cyan with the bronze spheres corresponding to a surface representation of the lipid phosphate particles. For the sake of clarity other lipid and solvent particles have been omitted. The initial distance of both helices of 70 Å is indicated with an arrow. Dotted lines pointing towards **B** connect each snapshot with the corresponding time point on the helix-distance plot. (**B**) Plot of the distance Cα helix–helix approach shown as a function of time for the same wt FGFR3 CG simulation as in panel A. Helix distance plots of all individual simulations in each ensembles are presented in **Appendix E, 1-6**. (**C**) Plot of the average dimerisation time for each simulation ensemble run in a 100% POPC lipid bilayer (if not differently specified) of FGFR3 wt, FGFR3 G380R, FGFR Y373C, FGFR3 wt (in 80% POPC 20% POPS lipid bilayer), FGFR4 wt, FGFR4 G388R and FGFR4 Y357C. Tracking of the first Cα helix-helix distance that measured below <7Å identified the dimerisation time of individual simulations. Error bars show the standard deviation across all 20 simulation repeats across each simulation ensemble.

3.5 Wt and mutant FGFR3 MD simulations

3.5.1 Analysis of coarse-grained Simulations of wt FGFR3

3.5.1.1 Identification of the TMD dimer interface of wild-type FGFR3

With the analysis performed in **Section 3.4**, no significant differences between FGFR3 wt and mutant dimer formation could be observed. In order to reveal how mutant residues are involved in dimer formation, it is key to study the nature of the helix-dimer interfaces more closely. Firstly, to identify whether FGFR3 wt TMDs form a predominant dimeric interface, a positional probability analysis was performed. Again a MDAnalysis and Python based in-house code (Reddy et al., 2014) was adopted, which centres and rmsd-fixes helix 1 of all frames of an ensemble and plots the relative configuration of the second helix on a thermal graph that represents the xy-plane of the lipid membrane (**Figure 3.4 A**). The plot for the wild-type ensemble in **Figure 3.4 B** shows that helix 2 is distributed anisotropically with a very low positional probability (indicated with blue colour) at a radial distance of 25-60 Å and a relatively high positional probability (indicated with yellow colour) within a 15 Å radius from the central helix 1. A close up view of the same plot with a more sensitive probability range clearly reveals a primary dimer interface for FGFR3 wt (yellow/orange area in **Figure 3.4 C**). The results of this analysis are encouraging as they provide evidence that FGFR3 transmembrane domains do not just stick to each other in a random fashion but interact with each other via one preferred interface. However, this analysis does not give direct information on which residues are involved in formation of helix-helix contacts.

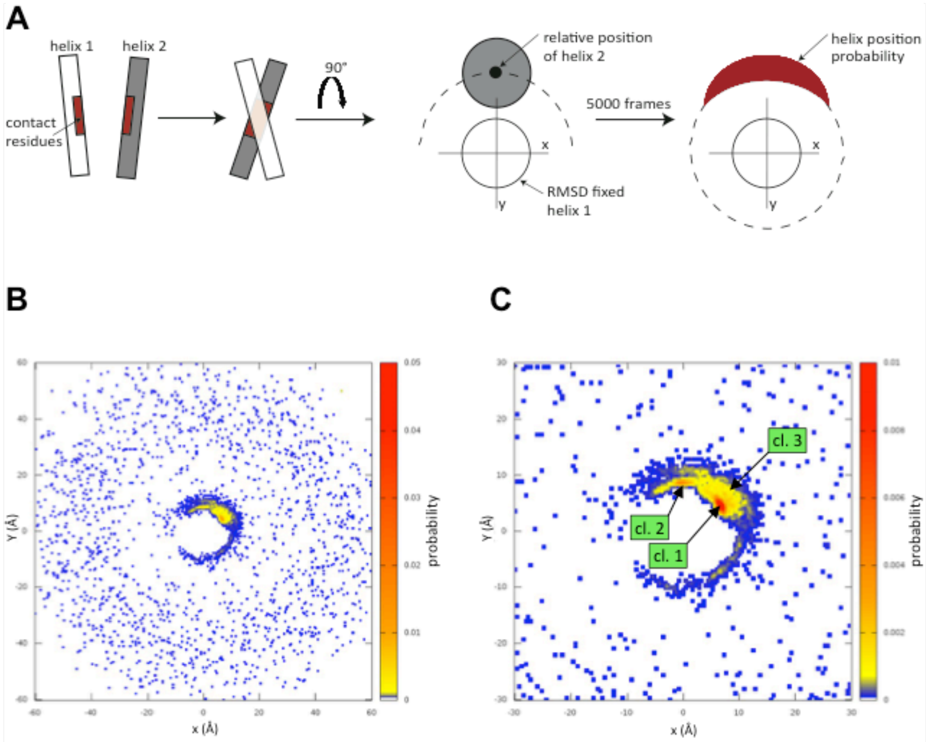


Figure 3.4 Thermal plot of positional probability

(**A**) The thermal plot of positional probability is a simple way of visualising preferred dimer interfaces between two helices. This is achieved by plotting of the relative position of helix 2 to rmsd-fixed helix 1. Including every 10th frame to this analysis gives 5000 data points for ensembles with a total simulation time of 20 μs. Calculation of the probability for each xy-position allows a heat map representation which colours data points according to their probability. Areas of lower probability are coloured in blue and areas of higher probability are coloured in red. (**B**) Spatial distribution of helix 2 in the reference frame of rmsd-fixed helix 1 (centred at the origin) in 20 repeats of 1 μs coarse-grained simulations. X- and y-axis represent the membrane plane of the full simulation box size with a side length of 120 Å. Each dot represents one frame. The non-linear probability scale is indexed as P = 0.0 white, P = 0.000001 blue, P = 0.001 yellow, P = 0.01 orange, P = 0.05 red. (**C**) Close up of **B** with a radius of 30 Å and a thermal plot with maximum heat at a probability of p = 0.01.

3.5.1.2 Identification of predominant Dimer forming Residues

Given that there is no experimental high–resolution structure available for the FGFR3 TM domain, these simulations may provide more accurate information on important dimer interface residues and give some insights into the molecular mechanism of pathological mutations. To identify which residues feature most prominently at the dimer interface for FGFR3 wild-type, an analysis was used that has been proven to identify predominant interacting residues from GpA dimer simulations (Reddy et al., 2014) consistent with previously experimentally identified interfacial contacts (Treutlein et al., 1992). Firstly, in order to calculate the normalised probability for each Cα residue to be in the top five closest interhelical contacts over the whole ensemble of simulations, a distance matrix of all possible Cα-Cα combinations between helices was created. For each frame, the residue identifiers of the five Cα particles that form the smallest distances, were recorded. The total number of frames in which one residue is found amongst the five closest contacts was divided by the total number of contacts (5 x number of frames), which yielded the probability for each residue to be in the five closest dimer contacts. In order to assess symmetry, probability calculations were done separately for matching residues on helix 1 and 2. This approach will be referred to as the "top five contacts" analysis.

The result of the top five contacts analysis for FGFR3 wt is plotted in **Figure 3.5** and reveals unambiguously an interface formed of mostly small and hydrophobic amino acids, separated by gaps of three non-interface residues. This is consistent with the identification of a helical interface. As described previously, the so-called GxxxG Engelman motif is featured in helix-helix association of many protein TM domains (Russ & Engelman, 2000). This motif has been proposed by (Schneider & Engelman, 2004) to be not only valid for interfacing glycines but also for other small residue (Ser, Thr, or Ala). FGFR3 does not comprise a GxxxG motif in its helix-helix interface but has an ample motif matching the (small)xxx(small) motif (**Figure 3.5**). The FGFR3 TMD interface can be separated into an N-terminal interface formed by the sequence D_{367}xxGSxxAxxxSxxxG and a C-terminal interface that can be assigned to residues V_{393}xxxR. Interestingly, the first residue forming part of the interface (D367, negatively charged) and the last residue forming part of the interface (R397, positively charged) are both hydrophilic and located in the extracellular or intracellular water-membrane interface respectively. This suggests, that the helix-helix interface is not just formed by hydrophobic interactions in the hydrophobic membrane core but that interacting residues outside the membrane support the interface.

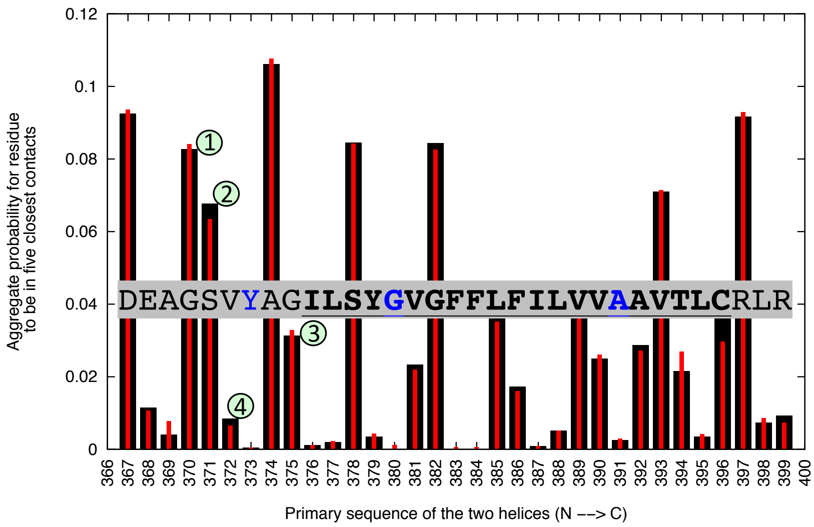


Figure 3.5 Residue top five contacts analysis

The normalised frequency of occurrence for a residue in the five closest contacts between helix 1 (black) and helix 2 (red) when the interhelix (Cα) separation is within 7 Å over all coarse-grained replicate simulations for FGFR3 wt. The peptide sequence is shown in the grey box with bold letters for the UniProt-predicted TMD and dark blue letters for selected residues, which have been identified to be mutated in skeletal diseases but are not involved in helix-helix interactions. Residues labelled with green circles are mutated to cysteine in TD. Numbers represent the experimentally proven ranking of these mutant cysteine's potential to enhance dimer formation.

Surprisingly, all residues in the hydrophobic core of the TMD that have been identified to be mutated in various skeletal disorders (Y373C, I376C, Y379C, G380R, F384L and A391E) do not feature prominently in the closest contacts in FGFR3 dimer configurations. It is noteworthy that residues G380 and A391, involved in Achondroplasia and Crouzon Syndrome (Meyers et al., 1995) respectively, sit exactly between two residues that form important parts of the interface (**S**xG_{380}x**G**, **V**xA_{391}x**V**), supporting the assumption that these residues are located on the flip side of the binding interface. This is consistent with the lack of an effect of the Achondroplasia G380R mutation on dimerisation propensity (You et al., 2006). In contrast G370, which is predicted to be located at the extracellular water-membrane interface, is prominent in the closest contacts and mutated to Cys in TDI. Compared to Achondroplasia, TDI is much more severe and normally neonatal lethal (Martínez-Frías et al., 2010). Thus, phenotype severity may, reflect the proximity of the point mutation to the dimer interface. As highlighted in **Figure 3.1 C**, there are several residues located N-terminally of G370, which have been found to be also featured in TD as cysteine mutations. As our simulations show, the intensity of dimer interface involvement is decreasing in the order G370 > S371 > G375 > Y373 (see green encircled numbers in **Figure 3.5**). The G370C mutation has been identified by site-specific infrared dichroism to be directly part of the dimer interface (Peng et al., 2009). Interestingly, the same study revealed that the propensity for disulphide formation in FGFR3 decreases in the same order as our simulations predict intensity of involvement in helix-helix contact formation.

Residue R397, which was identified as one of the top three FGFR3 dimerisation contact residues, is part of the C-terminal CRLR tetra-peptide. Peng et al. have shown that deletion of this tetra-peptide leads to increased dimerisation propensity of FGFR3, matching that of GpA. It is necessary to emphasise, that in order to identify the helix-helix interfaces, it was key to use the relative distances of the Cα particles, not contacts of the coarse-grained side chains. This means, that the R397 Cα particles form close contacts not due to, but despite the bulky arginine side chains. With the performed analysis it can just be assumed that both arginine side chains can not be pointing directly at each other but must be forming interactions in parallel or antiparallel orientation. The analysis presented in **Figure 3.5** enables us to suggest residues that form the main TMD interface. However, it does not provide structural models of the most representative helix configuration and of alternative dimer interfaces. More accurate description of side chain interactions requires detailed analysis of the different protein conformations and helix-helix configurations in each simulation ensemble.

3.5.1.3 Representative dimer models by cluster analysis

In order to identify structural models for the various areas of different probability on the positional probability plot, the following analysis will separate frames according to the structural similarity of the TMD dimers. To achieve that, the helix-helix configurations of all frames were compared on an rmsd distance matrix and clustered according to an rmsd cut-off using *g_cluster* of the GROMACS package. The value of the rmsd cut-off was varied between 2.5 and 4.5 Å in order to obtain a primary cluster, containing approximately 40 – 50 % of all dimer structures. Representative average structures were generated for the first three clusters using an RMSD cut-off of 3 Å (**Figure 3.6**). Adding these three clusters together, they compromise 61 % of all frames in the ensemble.

The representative dimer structure of cluster 1 (48 % of all structures, **Figure 3.6**) confirms, that the TMD interface is predominantly constituted by the residues of the ample (small)xxx(small) motif (as identified in **Section 1.5.1.2**; **Figure 3.5**). According to this highly symmetrical representative structure, the interface is formed via close interactions of Asp367, Ser371, Ala374, Ser378 and Gly382 at the N-terminal end and via Val393 and Arg397 at the C-terminal end of the TM domain (see residues in spheres representation in black colour **Figure 3.6**). As previously assumed, both bulky Arg397 residues are positioned parallel to each other and thereby enable their Cα particles to undergo close interaction. Strikingly, and as suggested by the top five contact analysis, all three disease-target residues Y373, G380 and A391 blue spheres with red labels) are positioned on the exact flip side of the dimer interface. Most residues reside between both dotted lines, which represent the coarse-grained PO_4 particles, and thereby inside of the self-assembled membrane bilayer. Our simulations show that the positively charged head group particles of coarse-grained Arg397, which mimic the arginine guanidino groups, are not located in the cytosol but rather buried below the level of the lipid headgroups, where they point towards the PO_4 layer. This is consistent with previous descriptions of the behaviour of positively charge residues at the membrane bilayer interface (Killian & Heijne, 2000; C. Kim et al., 2012; Sankararamakrishnan & Weinstein, 2000).

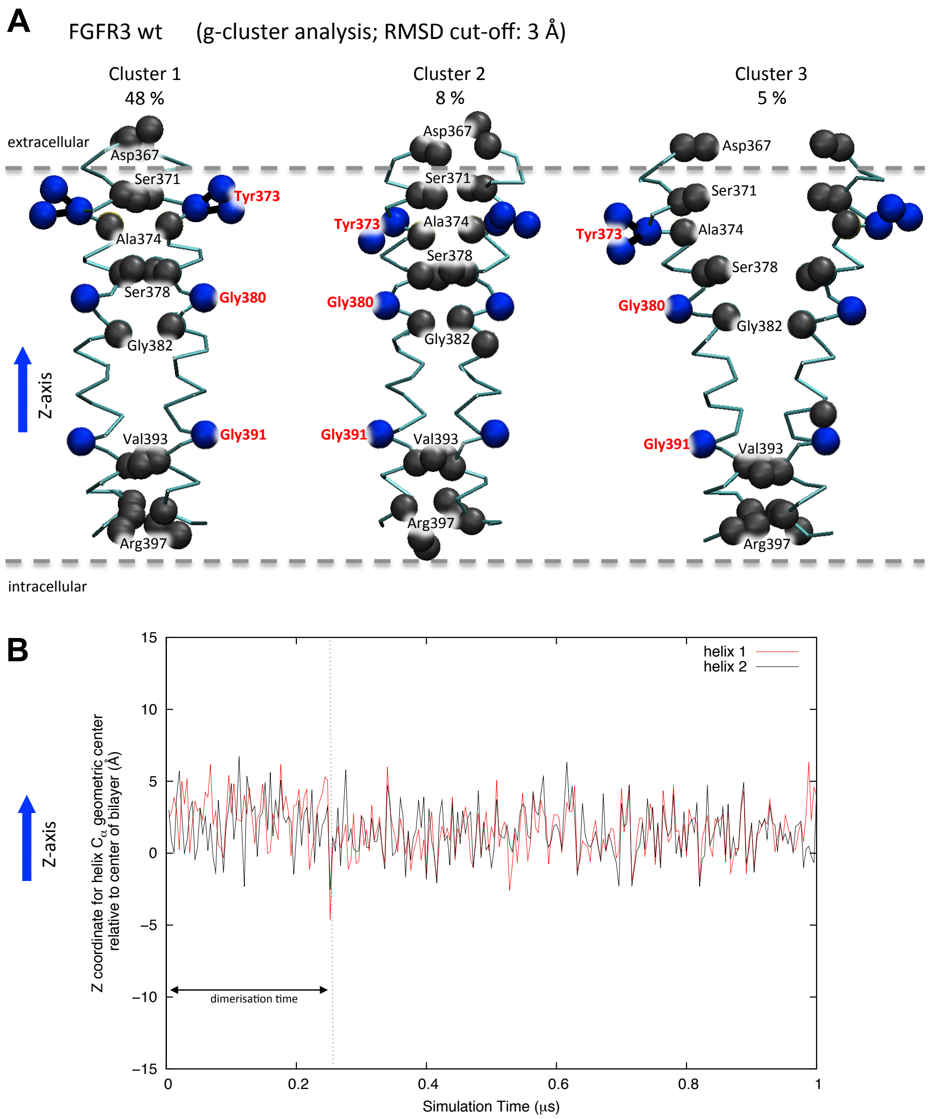


Figure 3.6 FGFR3 wild-type cluster analysis and z-tracking

(**A**) Representative average coarse-grained structures of the first three clusters of FGFR3 wt (identified with 3 Å RMSD cut-off by *g_cluster*, GROMACS), visualised with VMD. The Cα backbone is coloured in cyan. The closest contact building residues are shown by reduced van der Waals representation (black) and disease-target residues, which were predicted not to be involved in dimer formation (Y373, G380 and A391; see **Figure 3.50**) are shown in van der Waals representation in blue, with red label. Two dashed lines symbolise the approximate position of the PO_4 groups of the coarse-grained POPC lipid-bilayer. (**B**) Tracking of the z-coordinate for each helix Cα geometric centre along the bilayer normal, relative to centre of the bilayer during the first FGFR3 wt replicate simulation. Every 10th frame was parsed in a non-centred trajectory.

Cluster 2 (8 % of all structures) engages the same residues for dimer formation with the only difference that the residues in the TMD centre seem to interact more closely, whereas the residues at both termini seem to be slightly looser. Even cluster 3 (5 % of all structures) reveals the same symmetric interhelical orientation with the difference that all interactions on the

N-terminal side of the TMD are abrogated to leave only two close helix contacts on the C-terminal side of the TMDs (Val393 and Arg397).

The relative position of the helix dimer centre of mass (COM) to the geometric centre of the lipid bilayer, which corresponds to the dimer's position in the z-axis (blue arrows in **Figure 3.6**), does not vary across different cluster configurations. The z-position of each helix was tracked through the course of each simulation and no significant switches in z-position other than fluctuations in nanosecond scale of about +/- 2.5 Å around the average helix z-position of 2.5 Å were found. **Figure 3.6 B** shows an exemplary z-tracking plot of the first simulation repeat of the FGFR3 wt simulation ensemble, which forms a TMD dimer after about 0.25 μs (for reference see **Appendix E, A** simulation repeat 1). Fast fluctuations are undergone by both helices, mostly in a synchronised fashion; stationary changes in z-position can not be observed.

3.5.2 Coarse-grained simulations of mutant FGFR3

In this chapter, the FGFR3 mutant simulation ensembles are subjected to the same analysis as presented in **Section** 3.5.1 for the FGFR3 wt simulations. Some results of the wild-type analysis will be used as reference in a few of the figures in this chapter.

3.5.2.1 Identification of the TMD Dimer Interfaces of mutant FGFR3

The thermal plot of positional probability analysis and the top five contacts analysis in **Section 3.5.1** have shown that our wild-type helices form dimers mediated by an ample motif of mostly small and hydrophobic residues. Furthermore, the residues Gly380 and Tyr373 (which are mutated in skeletal disorders) are positioned on the exact opposite side of the dimer interface. Comparing the positional probability plots of FGFR3 wt to those of G380R and Y373C, it is evident, that both mutant simulation ensembles present a primary interface (labelled with green box as ***cl. 1***, (cluster 1) in **Figure 3.7 B** and **C**), which corresponds to the secondary interface of

FGFR3 wt interface (labelled ***cl. 2*** in **Figure 3.7 A**). Interestingly, as the representative FGFR3 wt structures for cluster 1 and 2 have shown (**Figure 3.6 A**), these clusters engage the same residues for helix-helix interactions despite being shifted by around 40° counter clockwise. This led to the conclusion that the primary binding interface observed for both mutants G380R and Y373C does not abrogate the interactions found for the wild-type primary dimer interface, but rather shifts the approaching angle of the second helix slightly. Indeed, the top five contacts analyses for both mutants identifies the same interhelical contact residues, forming the same (small)xxx(small) Engelman-like motif patterns (**Figure 3.7 D, E** and **F**). In both cases the residues D_{367}xxGxxxAxxxSxxxG remain the central residues for dimer formation. As a consequence of conservation of the dimer interface across wild-type and mutant simulations, the mutant residues remain excluded from the dimer interface. This suggests, that TMD-triggered mechanisms of over activation are either out of reach of the insight that our coarse-grained simulations can give, or that aberrant functioning of mutant FGFRs studied in this work is not directly expressed through major changes of the dimer interface.

Strikingly, whereas all other residues occur in the top five contacts to a certain degree, two phenylalanine residues Phe383 and Phe384, which are positioned at the centre of the FGFR3 TMD, do not form interhelical interactions at any time. This is consistent with findings that phenylalanine is highly enriched in the centre of TMDs (Wallin et al., 1997) and has low affinity for the membrane interface but highest affinity to the membrane core, where it keeps the TMD in correct position by reaching between the two membrane leaflets (Braun & Heijne, 1999; Killian & Heijne, 2000).

Comparing the closest contacts analysis, no severe shifts of probabilities can be identified. However, Arg380 slightly decreases the frequency of occurrence in the top five closest contacts of its surrounding interface residues S378 and G382 (SxR_{380}xG). In contrast, residues G370 and A374 clearly evolve higher values of probability of about 0.138 and 0.146 respectively, compared to 0.084 and 0.014 in the wild-type simulations. At the same time the C-terminal interaction pattern, which was identified as V_{393}xxxR in **Section 3.5.1.2** loses its strength with probability values of Val393 = 0.0672 and Arg 397 = 0.094 for the wild-type simulations to probability values of Val393 = 0.036 and Arg 397 = 0.068 for G380R simulations.

As **Figure 3.7 C** shows, a fairly big yellow cloud, covering a radial angle of about 90˚ around helix 1, represents cluster 1 for the Y373C simulations. At the same time, cluster analysis has shown that Y373C gives rise to a very weak secondary binding interface on the exact opposite side of the major binding region. These two effects together may describe why the closest contacts analysis has overall lost significant interface residues on the N-terminal end of the C-terminal region. It is noteworthy that the only residue that significantly gains importance as a dimer interface residue is Ala374, which is the adjacent residue to the mutant Cys373.

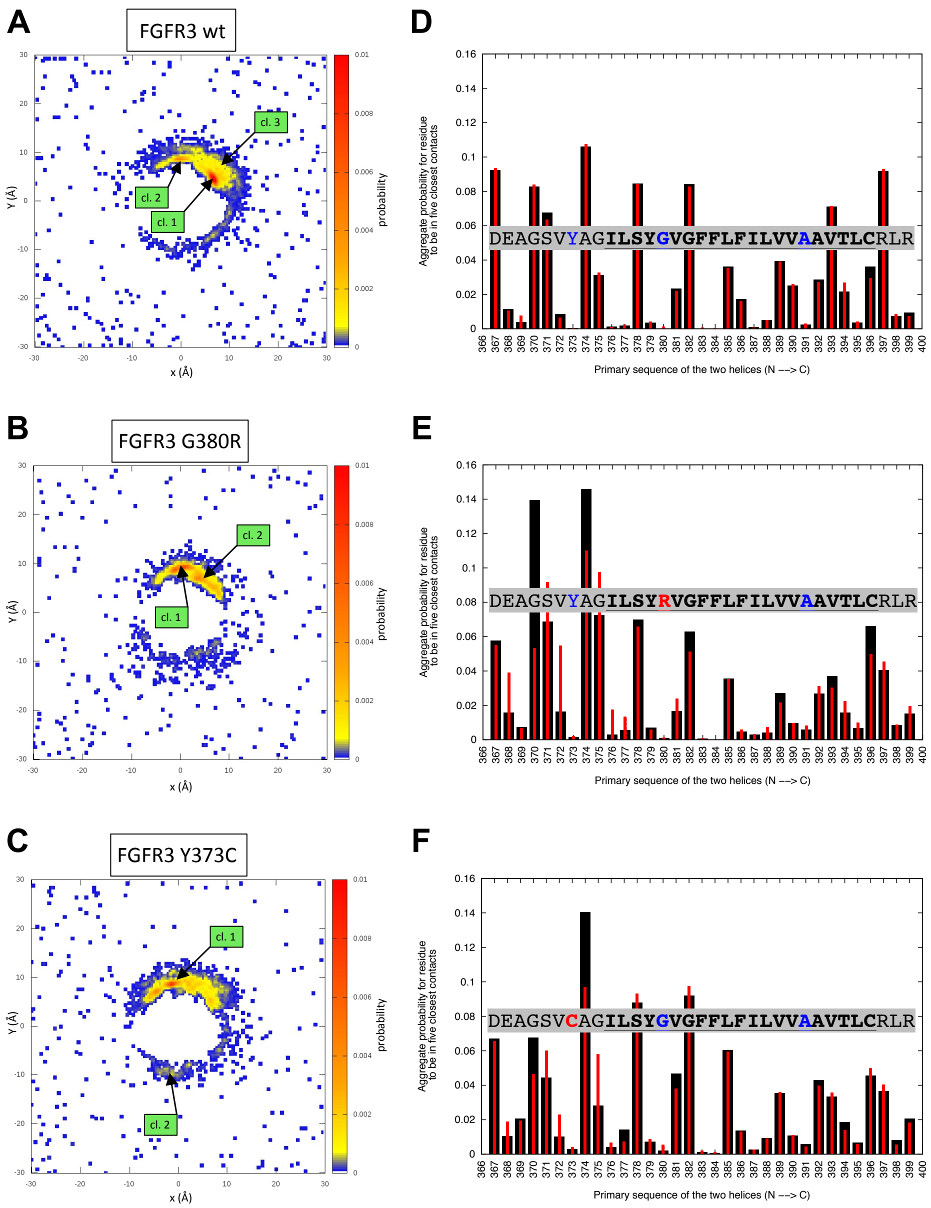


Figure 3.7 Positional probabilities and top five contact analysis

Spatial distribution of helix 2 in the reference frame of rmsd-fixed helix 1 (centred at the origin) in 20 repeats of 1 μs coarse-grained simulations for FGFR3 wt (**A**), G380R (**B**) and Y373C (**C**). X- and y-axis represent the membrane plane of the full simulation box size with a side length of 120 Å. Each dot represents one frame. The non-linear probability scale is indexed as P = 0.0 white, P = 0.000001 blue, P = 0.001 yellow, P = 0.01 orange, P = 0.05 red. Green boxes indicate clusters of higher probability, which were identified by g_cluster (GROMACS). All heat maps were generated using gnuplot. The normalised frequency of occurrence for a residue in the five closest contacts between helix 1 (black) and helix 2 (red) when the interhelix (Cα) separation is within 7 Å over all coarse-grained replicate simulations for FGFR3 wt (**D**), G380R (**E**) and Y373C (**F**). The peptide sequence is shown in the grey box with bold letters for the UniProt-predicted TMD, blue letters for selected residues, which have been identified to be mutated in skeletal diseases but are not involved in helix-helix interactions and red residues for the mutated residue in the respective simulation ensemble.

3.5.2.2 Representative TMD dimer Models of mutant FGFR3

In order to find more detailed conformational or configurational effects caused by the mutations, representative models of cluster 1 for both mutant TMD ensembles were compared. As discussed in **Section 3.5.2.1**, **Figure 3.8** confirms that the interfacing residues are pretty well conserved through all simulation types. The 40˚ rotational shift of cluster 1, which was observed in the positional probability analysis for both mutants (**Figure 3.7 B and C**) is clearly visible on a single residue basis. FGFR3 wt Ser378 and Val393 of helix 1 point away from the viewer and are covered by the corresponding residue of helix 2, which point towards the viewer (**Figure 3.8 A-C**). In contrast, the configuration of the same residues in both mutant simulations is changed in that Ser378 and Val393 of helix 1 covers the corresponding residue of helix 2. This is consistent with a 40˚ counter clockwise shift of the interhelical approach, at conservation of the main helix-helix interactions. No further significant differences could be observed for the FGFR3 Y373C simulations. The dimer remains symmetrical and the differences in contact intensities observed in the top five contacts analysis can all be described with the representative cluster 1 model: the three residue pairs of Ala374, Ser378 and Gly382 form more tight interactions compared to the wt TMD dimer and the bigger distances between the C-terminal residue pairs of Val393 and Arg397 confirm the loss of significance of that pattern.

The mutant FGFR3 Arg380 residues do not seem to have repulsive interactions with the hydrophobic lipid-bilayer core, as there is no disruption of the binding interface. Surprisingly, the full G380R TMD dimer reveals a movement in z-direction of about 4 Å towards the extracellular region, which must be a direct effect of the Arg380 residue. Throughout the simulation both of the partially positively charged coarse-grained Arg380 side chains are directed toward the upper membrane leaflet where they are able to interact with the negatively charged phosphates of the lipid headgroups of the outer membrane leaflet. The resulting force of these arginine residues pulling upwards could be responsible for the increase of distance between the residue pairs Ser378 and Gly382, which are the closest interface residues to Arg380, and for the simultaneous effect on the N-terminal residue pairs of Gly370 and Ala374, which seem to be pushed closer together. Calculations of the average TMD dimer position relative to the bilayer normal (z-position) for all simulation ensembles confirm that the G380R mutation moves the helix dimer 4 Å upwards, that the Y373C mutation has no influence on the z-position and that wild-type simulations, which were run in membranes containing 20 % POPS do not show a significant difference in motion along the z-axis (**Figure 3.8 D**). Interestingly, monomeric wt TMDs have an average z-position of 2.4 Å and move down 0.65 Å after dimerisation. The same motion is observed for Y373C and the wt 20 % POPS simulation. In contrast, the monomeric G380R TMDs, which start at an average z-position of 5 Å, move even further up after dimerisation, to reach an average dimer position of 5.7 Å. This might be due to a synergetic effect, where the interaction of Arg380 residues with PO_4 particles is facilitated when both helices pull up together.

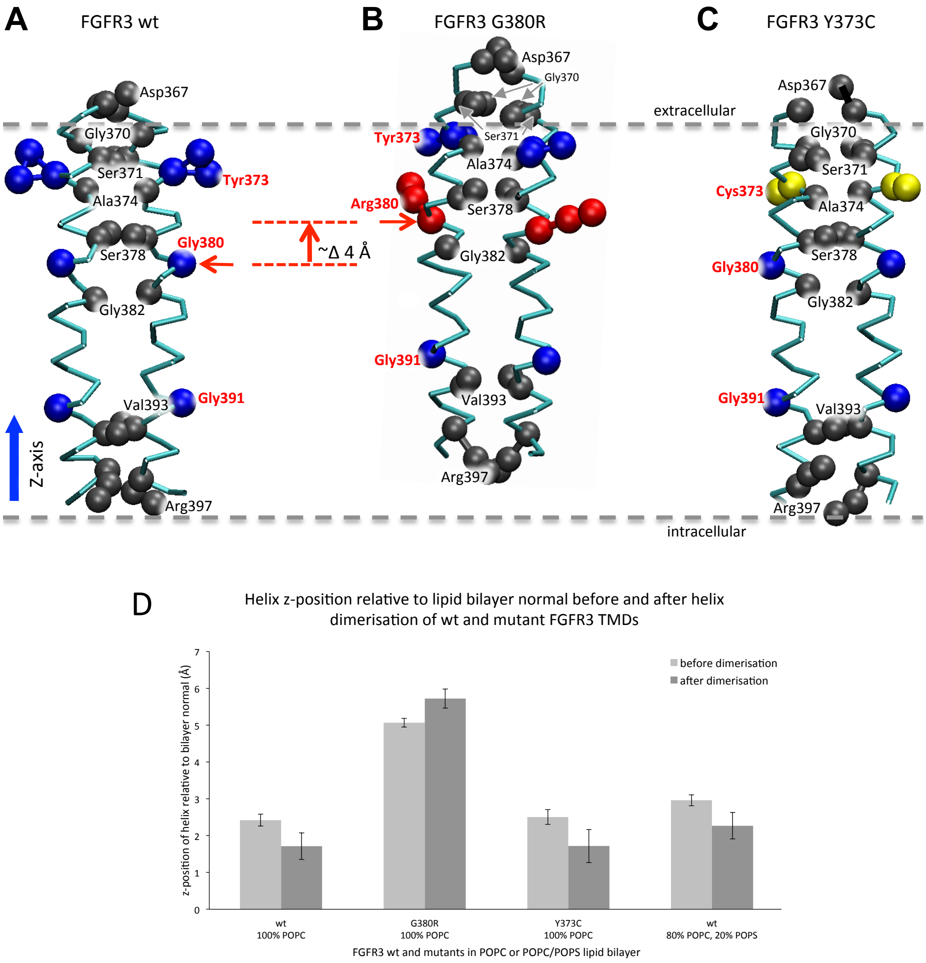


Figure 3.8 Representative coarse-grain models and helix positions in z-axis

Representative coarse-grain models corresponding to the primary helix–helix interface ((identified by *g_cluster*, GROMACS) **Figure 3.7 A-C**) for the wild-type (**A**), G380R homodimer (**B**) and Y373C homodimer (**C**) CG simulations. The Cα backbone is coloured cyan. The closest contact residues are shown as grey, spheres and disease-related residues (Y373, G380, and A391) as blue spheres. The two dashed lines indicate the approximate position of the phosphate particles of the CG lipid bilayer. The horizontal arrows indicate the residue at position 380 (Gly in the wild type and Arg in the mutant). Cysteine in the Y373C mutation is coloured yellow. The structure representations were generated with VMD. (**D**) Plot of the average z-position of TMDs, calculated from helix z-tracking data by measuring the distance of the centre of mass of each helix to the lipid bilayer normal for each single frame. The data of each single simulation repeat has been divided in the z-position before and after dimerisation (point of dimerisation was identified by tracking the closest Cα helix-helix distances below 7Å). Increase of the distance means movement towards the N-terminus.

3.5.3 Atomistic simulations of FGFR3 wt and G380R

The structural and functional implications of the FGFR3 G380R mutation were further explored with a particular focus on protein–lipid interactions on atomistic models. Therefore the representative CG structures shown in **Figure 3.8 A and B** were converted to the corresponding atomistic models (Stansfeld & Sansom, 2011) to run two short simulation repeats of 0.1 μs for FGFR3 wild-type and G380R. For each ensemble the average z-position of the COM of the helix dimer relative to the geometric centre of the lipid bilayer was plotted against the simulation time. The relative movements of wild-type and G380R dimers reveal an approximate difference in the z-position of ~5.5 Å (see vertical arrow in **Figure 3.9**), i.e., just a little more than in the CG simulations and consistent with the different resolutions of the two methods. This agrees well with the displacement of the TMD in the mutant by 5 Å revealed by neutron diffraction studies (Han et al., 2006).

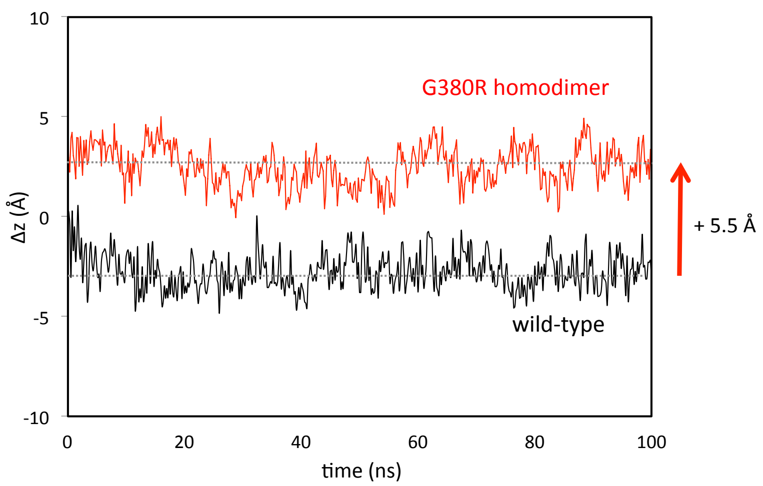


Figure 3.9 COM of helix dimer relative to COM of lipid bilayer

TM helix dimer movements along the *z*-axis (equivalent to the local bilayer normal) during the AT-MD simulations of the FGFR3 wild type (black) and G380R homodimer mutant (red; average of two simulations). The z-position of the geometric centre of each dimer is shown relative to the centre of mass of the lipid bilayer. The vertical red arrow shows the ~5.5 Å average difference in position of the TMD dimer along the bilayer normal.

The final frame of the atomistic wild-type simulation represents a stable dimer, fully integrated into the lipid bilayer (**Figure 3.10 A**). The primary dimer interface observed in the CG simulations is maintained. The final snapshot of the G380R mutant homodimer clearly shows the vertical translation of both Arg380 residues (see red circles in **Figure 3.10 B**), such that these side chains interact with the phosphate headgroups of the upper leaflet. This "snorkelling" effect of basic amino acid side chains in the hydrophobic membrane environment has been widely discussed to have a stabilising effect on the TMDs localisation within bilayers and has also been suggested to have a regulative role in integrin signalling across the membrane (Killian & Heijne, 2000; C. Kim et al., 2012).

In the wt TMD dimer, both of the Arg397 residues are located close to the lipid headgroup phosphates of the lower (inner) leaflet (**Figure 3.10 C**) where they have a horizontal, parallel position, which enables their Cα atoms to form close contacts. This interaction is in agreement with the CG simulations, which suggested that Arg397 plays an important role in (lipid-mediated) TMD dimer formation. Furthermore, the AT simulations confirm that both Arg399 residues reside in the cytoplasm and that both Cys396 residues are not involved in the dimer interface formation.

Interestingly, concomitantly with the 5.5 Å translation induced by the G380R mutant, the C-terminal residues (i.e., C_{396}RLR) are pulled towards the core of the lipid bilayer (**Figure 3.10 D**). Consequently, the Arg399 side chains occupy energetically favourable positions at the membrane–water interface that are occupied by the Arg397 residues in the wt TMD dimer. Intriguingly, the Arg397 residues are moved closer towards the membrane centre (resulting in some degree of local thinning of the bilayer), which forces them to adopt a snorkelling conformation to retain the interaction of the guanidinium group with the lipid phosphates. This destabilises the intermolecular Arg397–Arg397 interaction, resulting in more asymmetric rearrangement of the C-terminal region in which the Cys396 residues now form the main TMD dimer interaction. These findings support the suggestion derived from experimental data that conformational changes at the CRLR tetrapeptide may serve as a conformational switch between different states of activation and inactivation of the kinase (Peng et al., 2009).

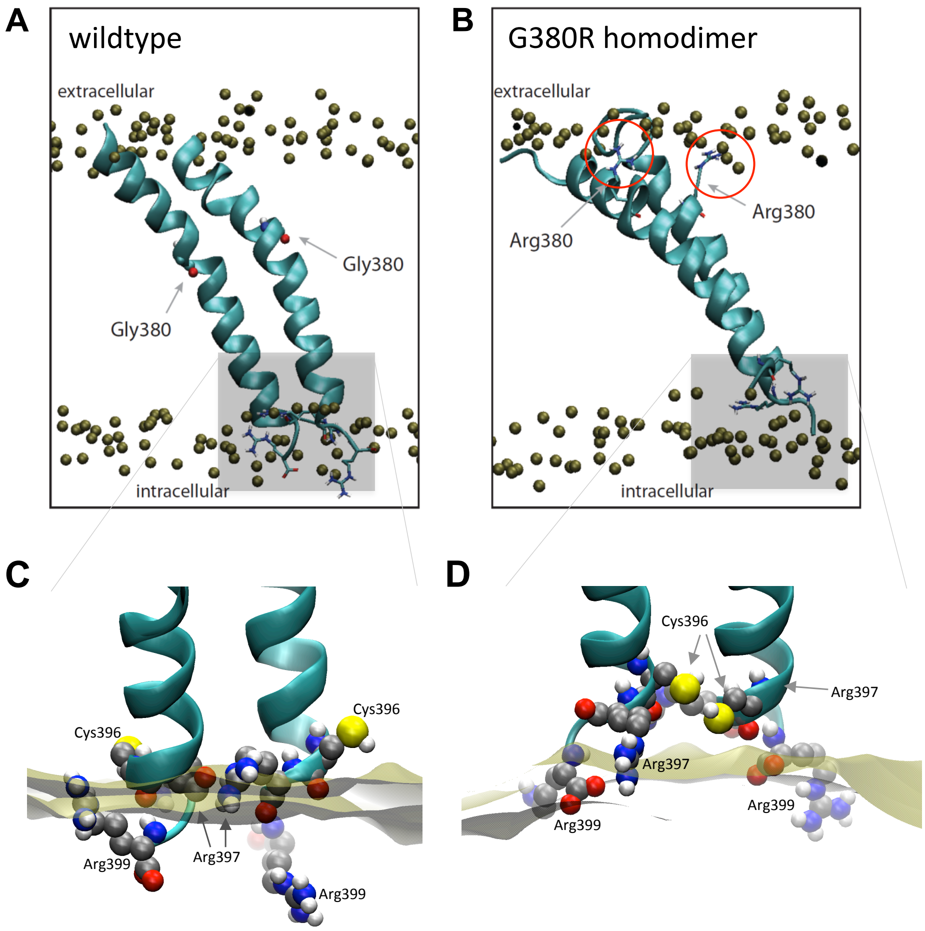


Figure 3.10 Snapshots of atomistic FGFR wild-type and G380R simlations

(**A**) Final frame of the 100 ns wild-type homo-dimer AT MD simulation. Gly380 and Arg399 are shown as spheres or bonds respectively. (**B**) Final frame of the 100 ns G380R homo-dimer AT MD simulation. The Arg380 side chains forming snorkelling interactions with lipid phosphates are shown in bond format and highlighted with red circles The phosphorus atoms of the lipids are shown as bronze spheres for both simulations (**C**-**D**) Detailed view of protein–lipid interactions at the inner membrane leaflet. The atoms of C396, R397, and R399 are shown as spheres, coloured according to atom type. The bronze surface represents the position of the membrane phosphate particles.

3.6 Analysis of wt and mutant FGFR4 MD simulations

3.6.1 Dimerisation interfaces of wt and mutant helices

All FGFR4 simulations have been analysed with the same tools as the FGFR3 simulations. The thermal plots of the spatial distribution of helix 2 around the rmsd-fixed helix 1 for FGFR4 wt and both mutants show very similar patterns of interaction (**Figure 3.11 A-C**). Two widely spread binding interfaces on opposite sides of the helix, with one strongly preferred side (always assigned to the first cluster, see green boxes) were identified for all three simulation systems. As the orientation of the rmsd-fixed start structure was not in the identical orientation, the thermal graphs need to be aligned according to the arrows in the centre of each plot. Interestingly, the TMD of FGFR4 comprises a real Engelman motif on the C-terminal side of the helix at positions G_{388}xxxG. Top five closest contacts analysis, clearly indicates that Gly388 and Gly392 interactions form the main anchor points for the dimerisation of the FGFR4 wild-type TM domains (**Figure 3.11 D**). Ala381 and Leu384 seem to have a supporting function in dimer formation. The histogram of the top five closest contacts for the mutant FGFR4 G388R (which was found to predispose cancer patients for significantly accelerated disease progression (Bange et al., 2002)) shows that the number of contacts between the glycines of the GxxxG motif strongly decreases (**Figure 3.11 E**). This suggests that this mutation has a direct impact on the dimer formation in that it shifts the binding interface towards the N-terminus by significantly increasing the close helix interactions of Ala381. However, this shift does not seem to have any influence on the approaching angle of both helices as the thermal plot analysis of wild-type simulations yield nearly the identic pattern as the G388R simulations. Also, the secondary interface opposite the primary interface remains the same.

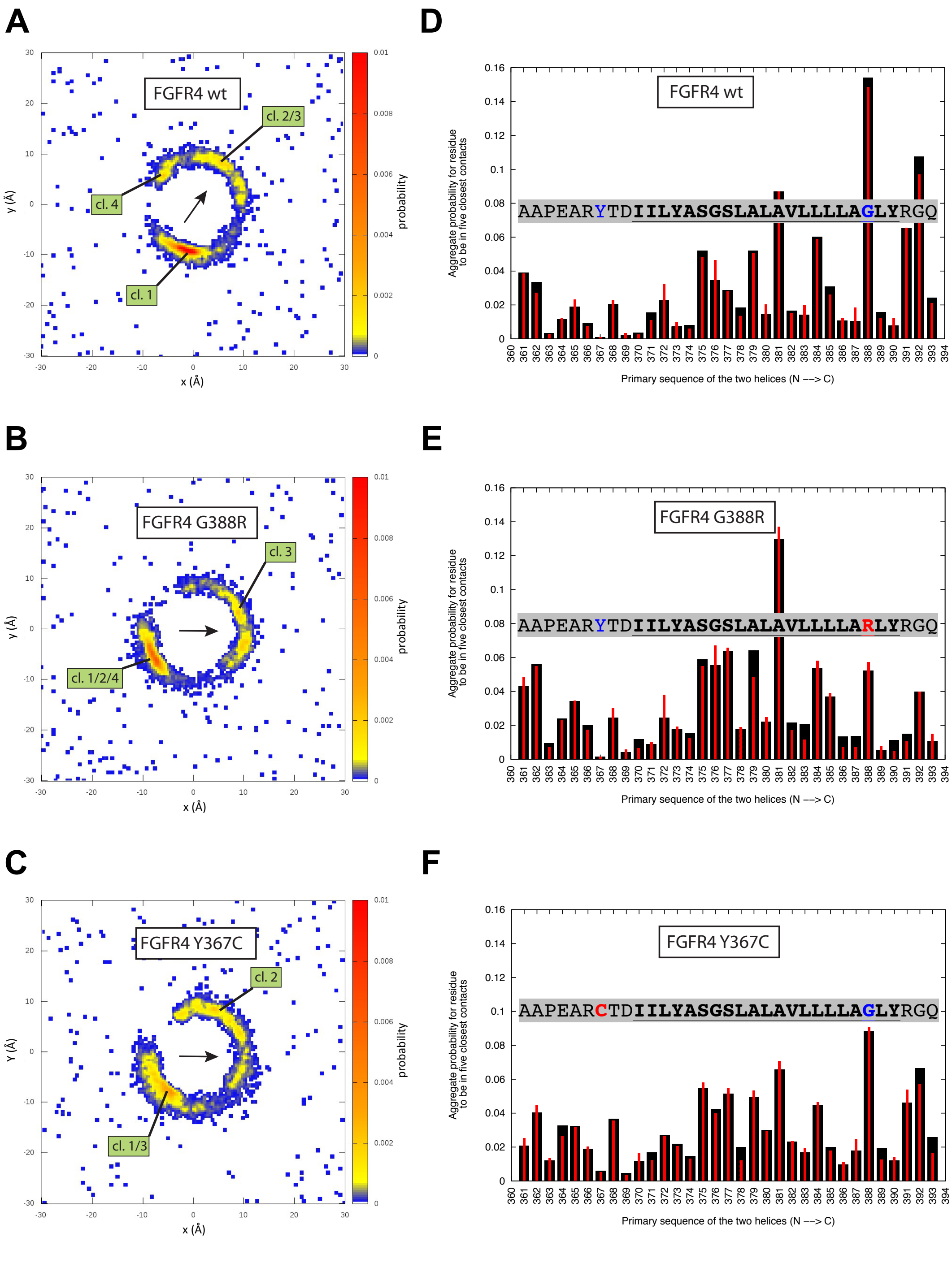


Figure 3.11 Positional probability and top five contacts analysis

(**A**-**C**) Thermal plot of the positional probability of FGFR4 wt (G388R or Y367C respectively) helix 2 in the reference frame of rmsd-fixed helix1. All trajectories of the 20 simulation repeats of each system were merged and every 10^{th} frame was included into the calculation. The non-linear probability scale is indexed as P = 0.0 white, P = 0.000001 blue, P = 0.001 yellow, P = 0.01 orange, P = 0.05 red. Green boxes indicate clusters of higher probability, which were identified by g_cluster (GROMACS). All heat maps were generated using gnuplot. (**D**-**F**) The normalised frequency of occurrence for a residue in the five closest contacts between helix 1 (black) and helix 2 (red) when the FGFR3 wt (G380R or Y373C respectively) helix-helix distance is ≤ 7 Å.

In contrast to the G388R mutation, the Y367C mutation did not show significant modulations of the TMD interface (**Figure 3.11 F**). This is not surprising, as according to our wild-type simulations, Tyr367 is not contributing to helix dimer formation. Thus, Cys367 did neither abrogate the C-terminal G_{388}xxxG motif, nor did it introduce heavy changes to the interaction pattern on the N-terminal pole of the TMD dimer. However, as the thermal positional probability plot for the Y367C mutation suggests (**Figure 3.11 C**), the prominence of the primary interface (labelled cluster 1/3) is weakened slightly, whereas the opposite secondary interface increases in importance. This suggests that there must be a fair percentage of asymmetric dimers in the simulation frames. Symmetric dimer frames as well as asymmetric dimer frames are all included in the top five contacts analysis, which might be the reason for a general decrease of probability for residues with higher values and a general increase of probability for residues with low values. As a result, for FGFR4 wt only 14 residues reached a probability value above 0.2, whereas for FGFR4 Y367C 21 residues passed the probability threshold of 0.2. In order to identify symmetric and asymmetric interface residues and in order to provide insight from our simulations on mechanistic effects, representative structures for each cluster need to be generated.

3.6.2 Representative TMD dimer models of mutant FGFR4

A representative structure for the most prevalent dimer conformation has been calculated with *g_cluster* (GROMACS) for the wt and mutant simulation systems. The closest contacts of the representative structure for the FGFR4 wt (**Figure 3.12 A**) (cluster 1, comprising 41.6% of the total 5000 structures) matches precisely the closest contacts identified by the top five closest contacts approach (**Figure 3.11 D**). The small and hydrophobic residues on the C-terminal side of each helix form a semi-symmetrical dimer with the interaction pattern A_{381}xxLLxxGxxxG, with the first glycine residue of the GxxxG motif (blue) being the disease-target residue G388. The N-terminal sides of the helices, comprising the second disease-target residue Y373, seem to form rather weak, if any interactions at all. Interestingly, representative structures for the second and third cluster (17.3 % and 9.5 % of the total frames pool respectively; structure not shown) reveal a symmetric interface mediated with the exact opposite helix side for cluster 2, and an asymmetrical attachment for cluster 3, in which one helix binds with the cluster 1 GxxxG interface and the other helix interacts with the opposite cluster 2 interface (for clusters see green boxes in **Figure 3.11 A** and **B**). Strikingly, as the representative structure of the FGFR4 G388R cluster 1 shows (**Figure 3.12 B**), Arg388 interferes with the primary GxxxG interaction pattern favouring asymmetrical interactions of small residues towards the N-terminus (Ala381, Leu384 and Leu385). The bulky Arg388 residues form tight interactions in anti-parallel orientation and snorkel down towards the phosphate headgroups. In that energetically favourable position the Arg388 residue pair crosses the space in which C-terminal residue pairs Arg391 and Gly392 of wild-type dimers form an important part of the interface, which results in their separation.

In contrast to the Y373C substitution on FGFR3, the representative structure of cluster 1 of the FGFR4 Y367C substitution obviously weakens the overall helix association, expressed by two helices, which do not form particularly tight interactions in any region. The G_{388}xxxG is weakened, which permits more flexibility in the binding interface without forming other significant interactions. Atomistic simulations of FGFR4 wt, G388R and Y367C need to be studied in order to get more insight into the importance of individual symmetric and asymmetric TMD dimers which were identified in this chapter by coarse-grained MD simulations.

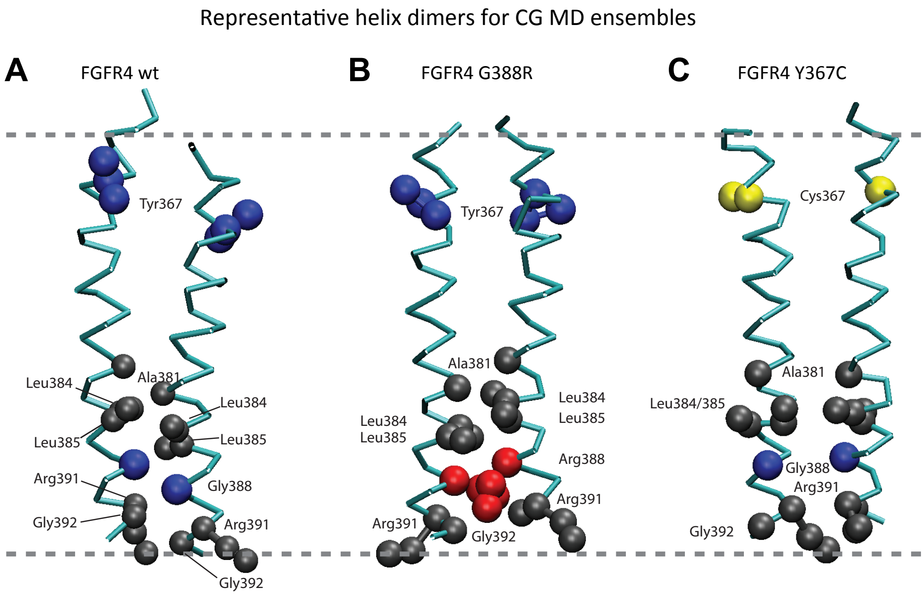


Figure 3.12 Representative TMD dimer models of wt and mutant FGFR4

Representative coarse-grain structures of the first cluster (identified by g_cluster, GROMACS) of FGFR4 wt (**A**), G388R (**B**) and Y367C (**C**) respectively, generated with VMD. The Cα backbone is coloured in cyan. The residues building the closest contacts are shown in grey van der Waals representation reduced in size. Disease-target residues are Y367 and G388 and A391 are shown in blue van der Waals radii. Two dashed lines symbolise the approximate position of the PO_4 groups of the coarse-grained POPC lipid-bilayer. In (B) and (C) the mutated residues are coloured red or yellow, respectively.

3.7 Discussion and future perspectives

In this work the behaviour of FGFR 3 and 4 transmembrane domains was studied by the means of coarse-grained molecular dynamics simulations. Introduction of disease-related mutations revealed changes in conformation and in dimer z-position, hinting on possible biological mechanisms of overactivation. The TM fragments that were choosen according to sequence alignments and TMD predictions (**Figure 3.1**), remained stable inside the lipid bilayer throughout all simulations and the majority of all simulations resulted in dimer formation (**Table 7, Figure 3.3**).

FGFR3 wild-type: Comparing our model of the FGFR3 TMD dimer with the NMR structure of the dimer in detergent micelles (Bocharov et al., 2013), the experimental data reveals a different interface, which involves an extended hydrophobic heptad motif. However, Bocharov et al. suggest an alternative packing mode for pathogenic mutations involving a (small)xxx(small) motif formed by residues G370-xxx-A374-xxx-S378-xxx-G382. This motif is identical with the interface that was found in the identified primary cluster for the wild-type TMD (**Figure 3.7 A** and **D**). This comparison of NMR studies with the presented simulation data suggests that the FGFR3 transmembrane domains may adopt different packing modes dependent on the environment (micelle vs. bilayer). However, the different packing modes could also represent different functional states of the TMD dimers.

FGFR3 G380R: The representative structure of the most abundant dimer configuration (**Figure 3.8 B**) strongly suggests, that both Achondroplasia mutant residues Arg380 are not engaged in dimer formation but are directly exposed to the lipid environment on opposing sides. This disagrees with a previously published model based on infrared spectroscopy and evolutionary conservation data (Peng et al., 2009), which suggested an asymmetrical dimer with different binding interfaces. A clear difference between the wild-type and mutant TM domains is their shifted position along the z-axis (**Figure 3.8 B**). While Gly380 in the wt helix is very close to the bilayer centre, the representative structure of FGFR3 G380R shows, that both arginine residues avoid the bilayer core by directing the positively charged arginine headgroups towards the lipid phosphate groups of the outer membrane leaflet. These results are consistent with previous work, which has shown that the positioning of arginine residues in the membrane core is energetically unfavourable (Dorairaj & Allen, 2007). In fact, lipid-exposed arginine side chains have rather been described to “snorkel” towards the membrane-water interface to coordinate with negatively charged phosphate groups (Sankararamakrishnan & Weinstein, 2000; Segrest et al., 1990). This snorkel movement of both Arg380 sidechains give an explanation for the observed 4 Å shift (compared to FGFR3 wt) of the dimer towards the outer leaflet. The snorkel effect is weaker before dimerisation (**Figure 3.8 B**) suggesting that the force of two G380R mutated dimers towards the membrane-water interface is higher than the force of one single helix. Both, the dimer shift and an observed coherent induced movement of the residue Arg397 into the hydrocarbon core are consistent with neutron diffraction studies, which describe a helix shift of 5 Å (Han et al., 2006). Our findings support previous reports that the Achondroplasia mutant does neither stabilise, nor destabilise the FGFR3 TM domain dimer (You et al., 2006). Thus, reasons for phenotypic changes may be described by more complex mechanisms. Cho et al. have demonstrated that the mutant receptors avoid lysosomal degradation and thereby prolong the signalling (Cho et al., 2004). Representative structures of wild-type and G380R systems (**Figure 3.8 A and B**) were converted to atomistic representations and used as starting models for atomistic simulations of 100 ns duration per run. Comparing the final snapshots of these AT simulations (**Figure 3.10 A** and **B**) provides further evidence, that both Arg380 side chains are responsible for the positive z-axis shift perpendicular to the lipid bilayer by "snorkelling" towards the lipid phosphate groups. Concomitantly, the C-terminal C_{396}RLR motifs (undergoing interactions with the Arg397 pair in the wild-type simulation) are pulled into the hydrophobic core of the membrane, where a rotation of these motifs breaks the Arg397 interactions and brings both Cys396 residues into close proximity, potentially allowing for formation of disulphide bonds. Future simulation studies looking at larger scale structural changes of the JM domains may probe whether this effect plays a role in the mechanism of pathogenesis of Achondroplasia. Furthermore, the G380R mutation has been observed to alter interactions of the JM domain with intracellular components (Cho et al., 2004), which possibly is the cause for a delay in FGFR3 downregulation (Monsonego-Ornan et al., 2000).

FGFR4 wt and G388R: This work presents the first model for the dimeric FGFR4 TM domain (**Figure 3.12 A**). The Gly388 residue was found to be the most prominent interfacing residue as part of a GxxxG motif (**Figure 3.11 D**). The mutant Arg388 totally disrupts the binding motif found for FGFR4 wt without disrupting the helix configuration (**Figure 3.11 B**). The bulky arginine residue rather shifts the region of closest backbone interactions towards the centre of the TM domain where Ala381 forms the closest Cα interaction (**Figure 3.11 E**). Neither a helix shift in the lipid bilayer nor a major shift of dimer interface could be observed. The crucial difference between the wt and the G388R mutant is the increased distance of the C-termini resulting from the electrostatic repulsion of the Arg residues. Atomistic simulations of the TM domains need to be performed to confirm the findings of the CG simulations and to get more detailed insights into the processes that promote sustained activation and subsequent promotion of prostate cancer initiation and progression, suggested by (J. Wang et al., 2008).

FGFR3 Y373C and FGFR4 Y367C: The analogous Y to C substitutions on FGFR3 and 4 are very poorly studied. Y373C on FGFR3 has been reported to cause Thanatophoric Dwarfism type1 (d'Avis et al., 1998; Rousseau et al., 1996) and the FGFR4 Y367C mutation has been found to enhance receptor activation and tumour growth of breast cancer cells (Roidl et al., 2010). While these studies report that the Tyr to Cys substitution predisposes disulphide bond formation, the studied FGFR3 and FGFR4 simulations suggest that the mutant cysteine residues are not part of the major dimerisation interface (**Figure 3.7 F and Figure 3.11 F**). Formation of disulphide bonds is unlikely to take place in the context of the cluster 1 interface. However, positional probability plots reveal a secondary interface for FGFR3 Y373C (**Figure 3.7 C** cl. 2) and FGFR4 Y367C (**Figure 3.11 C** cl. 2) that bring the Cys residues into closer proximity. Short contacts of two adjacent Cys residues would be enough to trap two helices in a dimer via disulphide bonding, potentially resulting in kinase domain overactivation. A different mechanism for enhancement of receptor activation by these Tyr to Cys mutations could be that activated receptor dimers, formed in cluster 1 mode, form covalent disulphide bonds to a third receptor in cluster 2 mode, resulting in a ligand independent activation of three receptors, subsequently leading to receptor clustering. Receptor clustering has been shown for other membrane receptors by fluorescent microscopy in live cells (**Seiradake et al., 2013**). This could be a feasible experimental approach to assess the role of mutant FGFR TMDs.

The identification of various interaction modes of TMDs is an emergent theme for a number of receptor tyrosine kinases. This probably reflects the looser interactions compared to the tightly binding, canonical interface of the GpA TMD dimer. Weaker interaction is likely to be of functional importance, allowing the RTK TMDs to form a “multistate switch”, which can be regulated by extracellular ligand-receptor or receptor-receptor interactions. It is important to consider possible methodological limitations of coarse-grained simulations. CG simulations do only provide an approximation to more detailed, atomistic treatments of TMD dimer interfaces. However, recent studies have suggested that the combination of CG and subsequent AT simulations can reproduce, e.g., the NMR structure of the integrin TMD heterodimer with high degree of accuracy (Kalli et al., 2011). One reason is that multiple CG simulations allow exploration of a wider range of helix-packing and conformational space than atomistic simulations on their own, provided the same limited computational cpu time. Furthermore, comparison of atomistic and coarse-grained simulations of GpA dimerisation suggests that free energy profiles for the dimerisation process yielded by the two methods are comparable (Janosi et al., 2010; Sengupta & Marrink, 2010). Most of the simulations presented in this work used POPC as a single phospholipid species, which is a oversimplification of what in biology is a highly complex composition of multiple lipid families. A study on GPCRs suggests that lipid composition of the membrane may influence receptor association and function (Botelho et al., 2006). This is likely to be translated to TMDs of receptor tyrosine kinases (Prakash et al., 2011). Parameters of many lipids are available and it will be important to consider including multiple lipid species in future studies, in order to study their interactions with the transmembrane and juxtamembrane domains of FGFR3. In particular, it seems that interactions of the intramembranous and juxtamembrane domains, with one another and with lipids, are likely to modulate the intrinsic TMD interactions (Endres et al., 2013; Matsushita et al., 2013).

Chapter 4

Biophysical studies of the FGFR4-heparin-FGF1 complex

Chapter 4 – Biophysical studies of the FGFR4-heparin-FGF1 complex

4.1 Summary

To this day the high-resolution structure of the FGFR4 ectodomain remains unresolved. Ig-domains 2 and 3 are hypothesised to bind to the ligand FGF in the same fashion as has been shown for FGFR1, FGFR2 and FGFR3. All published structures were obtained from refolded protein, expressed in bacterial systems and therefore lack N-linked glycosylations. Two controversially discussed crystal structures of the hypothetically active 2:2 receptor-ligand arrangement show very distinct architectures and involve highly sulphated oligosaccharide in different stoichiometric ratios.

This chapter describes the protocol for FGFR production in mammalian cells. SEC and SEC-MALLS analysis are used to find a strategy for stable signalling complex formation in the presence of N-linked glycans. The presented data suggests that glycosylated FGFR4 has a very weak affinity to FGF1. First the presence of heparin seems to enable FGF1 to bind to the well-studied FGF binding site on FGFR. The SEC-MALLS data also suggests that presence of heparin drives formation of a higher molecular FGFR4-Heparin-FGF1 complex in a 2:1:2 stoichiometric ratio. However, this complex is not tight enough to be isolated by SEC. Receptor deglycosylation by EndoF1 treatment led to partial degradation of the receptor. Deglycosylation of the receptor in the context of the full complex mix prevented degradation. FGFR4-Heparin-FGF1 complexes prepared using various FGFR4 constructs were screened extensively for crystallisation without success. Recorded SAXS data from the complex yielded a curve, which could not be fitted to any of the previously published signalling complex structures. Instead, ab initio and homology

modelling suggested a new arrangement of two receptor-ligand units in a maximally extended, heparin linked fashion.

4.2 Aims

The aim of this chapter will be to determine the architecture of the FGFR4-Heparin-FGF1extracellular signalling complex and to perform comparisons with previous structural data for the FGFR family. This requires establishment of FGFR4 and FGF1 protein production and purification, stable complex formation and crystallisation of the complex or alternatively SAXS data analysis.

4.3 Expression and purification of FGFR4 $D2D3_{glyc}$

Constructs were designed which comprise Ig domain 2 (D2) and Ig domain 3 (D3) of FGFR4. The construct design strategy is described in **Section 2.2** including choice of expression vector and cloning sites. Our domain boundaries were chosen based on sequence alignments with human homologous receptors (**Appendix B**), on secondary structure prediction and on published structures of FGFR1 and FGFR2 (Beenken et al., 2012; Plotnikov et al., 2000; Schlessinger et al., 2000). There were five N-linked glycosylation sites predicted (using the NetNGlyc 1.0 Server, CBS, Denmark) within the ectodomain of FGFR4, one in Ig domain 1 (N112) and four in Ig-domain 2 (N258, N290, N311, N322) (**Figure 4.1 A**). It is well known, that mutation of glycosylation sites or deglycosylation of proteins can favour crystallisation (Chang et al., 2007). In order to reduce heterogeneity caused by glycans and to test the contribution of each N-linked glycosylation to receptor solubility, each of the predicted glycosylation sites was mutated to Gln in a separate construct. Primers o12/o35 were used for generation of the wild-type FGFR4 $D2D3_{144\text{-}354}$ construct, mutagenesis primer pairs o1/o2 for N258Q, o3/o4 for N290Q, o5/o6 for

N311Q and o45/o46 for N322Q (see sequences in **Appendix C**). Efficient secretion of soluble protein indicates that the protein is folded correctly and likely to possess full biological activity (reviewed by Trombetta & Parodi, 2003). The FGFR4 D2D3$_{144\text{-}354}$ construct comprises 223 amino acid residues including linkers and His-tag from the vector. Secreted, fully glycosylated FGFR4 D2D3 is expected to have a molecular weight (MW) of 29962.9 g/mol and the calculated MW for non-glycosylated protein is 24762.9 g/mol. The bands in **Figure 4.1 B** (see small arrows) confirm these values. In western blot analysis of small-scale expression trials in HEK293S cells (**Section 2.3.2**), the N258Q, N290Q and N311Q constructs showed very low expression level of soluble glycosylated protein in the medium, compared to the non-glycosylated receptor present in the cell lysate (**Figure 4.1 B** lanes 2-7). The N322Q construct seemed to be slightly more soluble. However, the wild-type construct had significantly better solubility and secretion level, suggesting that FGFR4 solubility strongly depends on full glycosylation. Therefore, large-scale expression (**Section 2.3.3**) of FGFR4 D2D3 wild-type in HEK293S cells was chosen. These cells typically produce smaller protein yields than HEK293T cells but introduce shorter and highly homogeneous N-glycosylations, which can be cleaved off enzymatically for crystallisation purposes (Chang et al., 2007). No O-linked glycosylation sites were predicted in the receptor ectodomain sequence that could impede crystallisation.

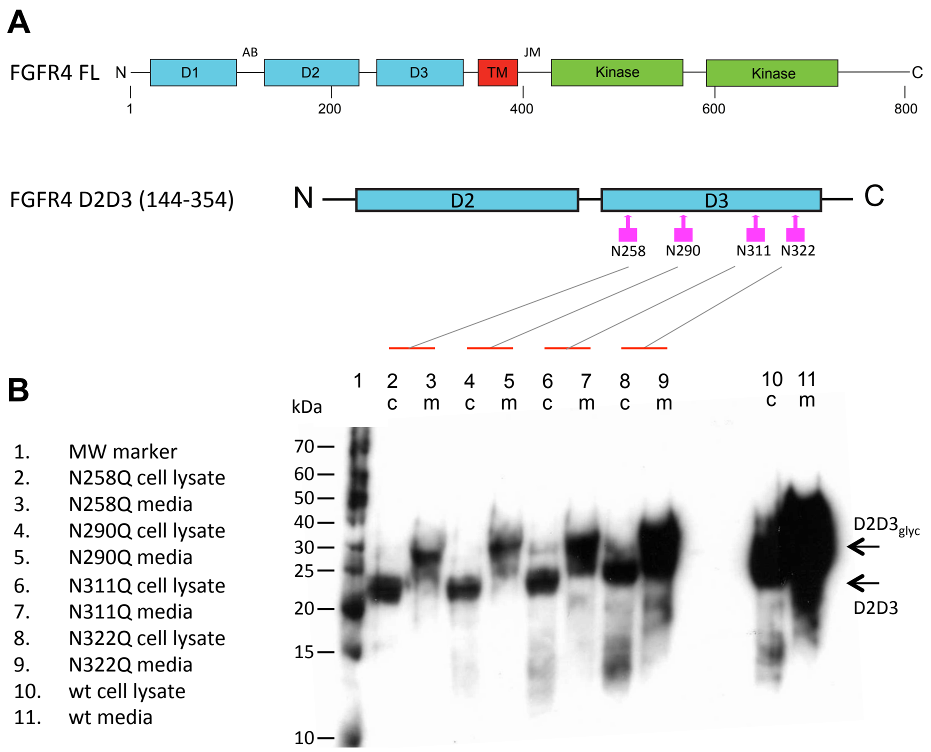


Figure 4.1 FGFR4 domain structure and HEK293S cell small-scale expression test

(**A**) FGFR4 contains three extracellular immunoglobulin like domains, named D1-D3. D1 and D2 are separated by an acidic linker consisting of 27 residues, called the acidic box. The TMD is linked to a split kinase domain by a juxtamembrane domain of 60 unstructured residues. The lower panel shows the basic expression construct. Pink arrows indicate predicted N-glycosylation sites on FGFR4 D2. (**B**) Western blot analysis of expression and secretion of wild-type and various mutant FGFR4 D2D3 constructs by HEK293S cells. All constructs are His-tagged and can be detected to different extents in the conditioned medium using monoclonal anti-Penta-Histidine primary antibody (**Section 2.3.2**). MW standard BenchMarkTM Protein Ladder from Invitrogen was used. Arrows indicate the MW for non-secreted and secreted protein, which match with the respective protein bands. MW values are presented in kDa.

The standard IMAC and SEC purification method was used as described in **Section 2.4.1** and **2.4.3**. Representative elution profile and SDS-PAGE analysis for the IMAC protein elution (see black bracket above elution profile) are shown in **Figure 4.2 A**. As expected, the His-tagged glycosylated protein has a size of around 30 kDa according to the MW marker. Protein fractions eluted at volume 14-24 (see red bracket **Figure 4.2. A**) were pooled and concentrated to 20-30 mg/ml. Typically 400 μl concentrated protein solution were injected to a Superdex S200 16/60 SEC column equilibrated in running buffer (10 mM HEPES (pH 7.5) and 150 mM NaCl).

A representative SEC elution profile and the corresponding SDS-PAGE (**Figure 4.2. A**) show that the high molecular contamination running at 120 kDa (unspecific binding to the IMAC column) could be separated to a satisfying degree from the receptor. Fractions within the red bracket were pooled and concentrated. Interestingly, multiple bands of FGFR4 D2D3 could be observed (see small arrows in **Figure 4.2 B**), each corresponding to a different degree of glycosylation (see also SDS-PAGE analysis of final protein product, lane 1 of **Figure 4.2 C**). For crystallisation purposes, any source of heterogeneity is undesirable and should be removed if possible. As expression of receptor constructs with mutated glycosylation sites failed, deglycosylation by treatment with EndoF1 was tested (**Section 2.3.2**; Chang et al., 2007; Grueninger-Leitch et al., 1996) which also provided information on the accessibility of N-linked glycan moieties at the protein surface. As reference a protein sample was subjected to PNGase F treatment according to the manufacturers protocol (New England Biolabs), which deglycosylates the protein after denaturation and therefore cleaves off all sugars. The first lane in **Figure 4.2 C** shows the concentrated sample after SEC. The second lane shows the same protein sample after EndoF1 (2 h at RT) and the third and fourth lanes show 5 μg and 10 μg protein sample after PNGase F treatment. The main band after EndoF1 treatment evidently corresponds to the fully deglycosylated protein, as it runs at the same level as PNGase F treated samples. Interestingly, EndoF1 is not capable of cleaving off 100 % of the N-linked clycans. Two additional bands at higher MW show that around 30 % of the total protein still possess one or two N-linked glycans. As it is unknown whether the proteins corresponding to the two additionally occurring bands have a homogeneous glycosylation pattern, a direct conclusion on the degree of homogeneity of native and EndoF1 treated samples can not be drawn – an observation that needs to be considered for crystallisation strategies. As a side effect, EndoF1 treatment led to minor protein degradation which could be removed by ultracentrifugation. Short deglycosylation for 5 min at 37 ° led to heavy protein precipitation and did not improve the glycosylation efficiency.

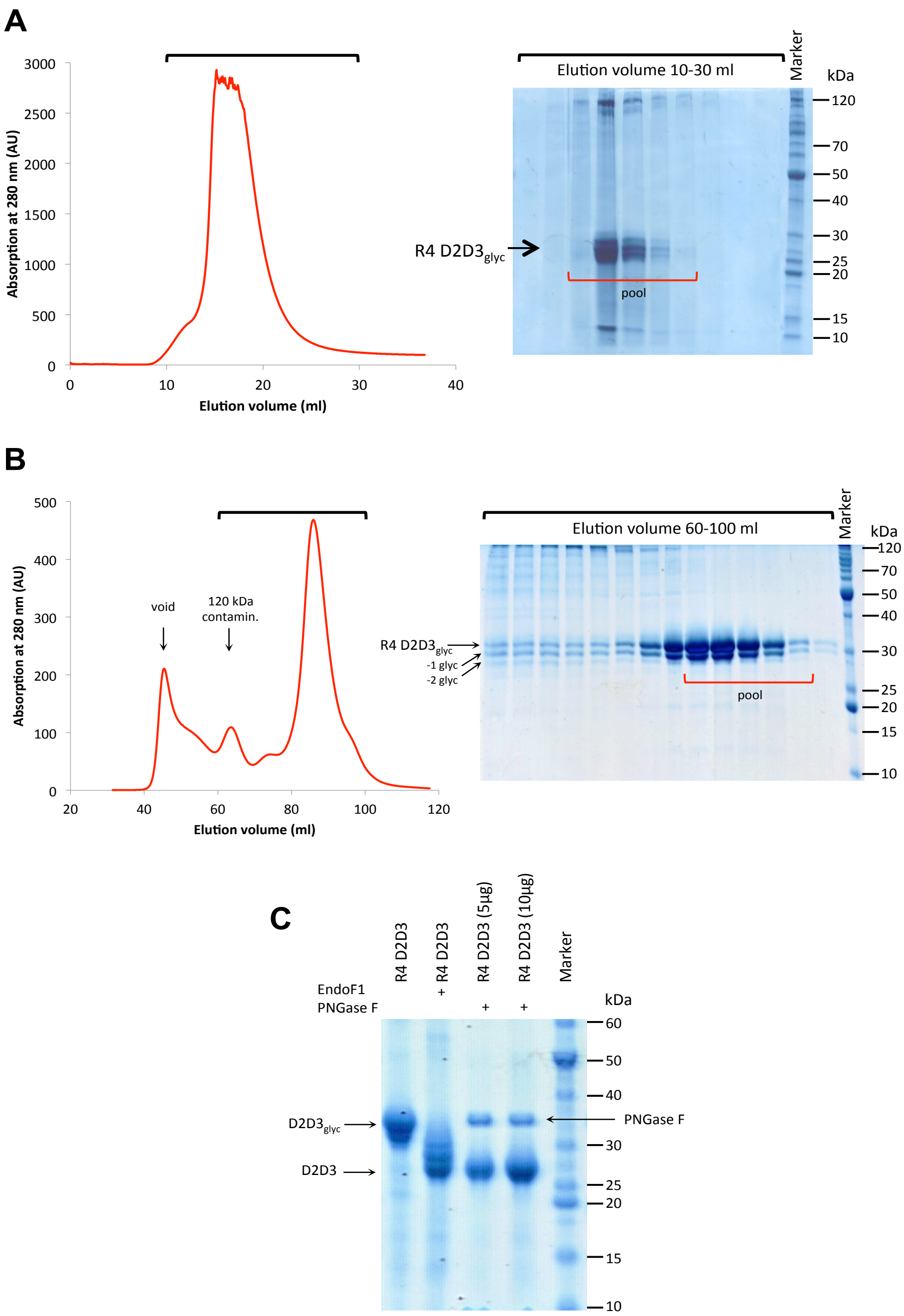


Figure 4.2 IMAC and SEC purification of FGFR4 D2D3

Representative purification of FGFR4 D2D3. (**A**) IMAC elution monitored by absorbance of 280 nm (red curve, left) and SDS-PAGE analysis using a Novex 4-12 %Bis-Tris gradient protein Gel (NuPAGE^{TM}, Invitrogen) and BenchMark Protein Ladder from Invitrogen as molecular weight marker (right). (**B**) SEC of R4 D2D3 using a HighLoad Sperdex S200 16/60 column (GE Healthcare Life Scineces) at a flow speed of 0.5 ml/min in buffer containing 10 mM HEPES (pH 7.5) and 300 mM NaCl. Elution was monitored as in (A). The void and contaminant peaks are labelled individually. The bracket indicates fractions that were subjected to SDS-PAGE analysis. SDS-PAGE was run as in (A). The red bracket indicates on samples that were pure enough for experiments. They were pooled, concentrated and stored. **C**) EndoF1 and PNGase F treatment of glycosylated FGFR4 D2D3 sample. Arrows to the left of the SDS-PAGE indicate fully glycosylated and fully deglycosylated protein. An arrow to the right of the gel indicates the PNGase F band. As EndoF1 has a MW of 32 kDa and is applied in small catalytic amounts, it is hidden by the FGFR4 D2D3 band.

4.4 Expression and purification of FGF1

Multiple crystal structures comprising human FGF1 or FGF2 have been published to date. Based on these structures, FGF1 constructs were designed encompassing amino acid residues Asn17-Asp155. Inserts were cloned as described in **Section 2.2.2** and **2.2.3** using primer o73 and o74 (see sequences in **Appendix C**). In the same way a full-length FGF2 construct was designed encompassing amino acid residues Met1-Ser155. Required restriction sites were designed into the synthetically produced FGF2 gene (Eurofins Genomics, Germany). These FGF1 and FGF2 constructs were cloned into pHLsec, pHLsec-mV and pHLsec-avi (for vectors see **Section 2.2.1**) to test for expression in HEK293S cells. Small-scale tests failed to express any of these constructs. Co-transfection test of FGF1 or FGF2 in pHLsec plasmid with FGFR4 D2D3 for co-expression showed efficient expression and secretion of the receptor, but no secretion of either FGF1 or FGF2 could be observed. As no N-linked glycosylation sites were predicted on FGF1 or FGF2, expression in mammalian cells was not essential for correct protein folding of the ligands. Protocols for production of FGF1 and FGF2 in *E.coli* are well established (Pellegrini et al., 2000; Rassouli et al., 2013). These bacterial cell culture protocols were adapted for production of all FGF ligand constructs in this study (see **Section 2.3.1**). Small-scale bacterial expression tests showed heavy expression for $FGF1_{17\text{-}155}$ as opposed to relatively weak $FGF2_{1\text{-}155}$ expression. **Figure 4.3 A** shows a representative elution profile of His-tagged $FGF_{117\text{-}155}$ from the IMAC column of 5 L bacterial culture. The corresponding SDS-PAGE confirms correct size of the protein (theoretical MW = 17087 Da) and a high degree of purity. The fractions corresponding to elution volume 13-17 ml (see red bracked) were pooled, concentrated to about 30 mg/ml and flash frozen in small aliquots of 150 μl. For further purification, single aliquots were thawed and buffer exchanged by SEC using a Superdex S75 column (see **Section 2.4.3**). The corresponding elution profile in **Figure 4.3 B** shows that FGF1 elutes in a sharp monodisperse peak. Peak fractions showing protein of about 97 % purity were pooled and directly used for experiments. One litre of bacterial culture yielded about 12 mg/L pure protein.

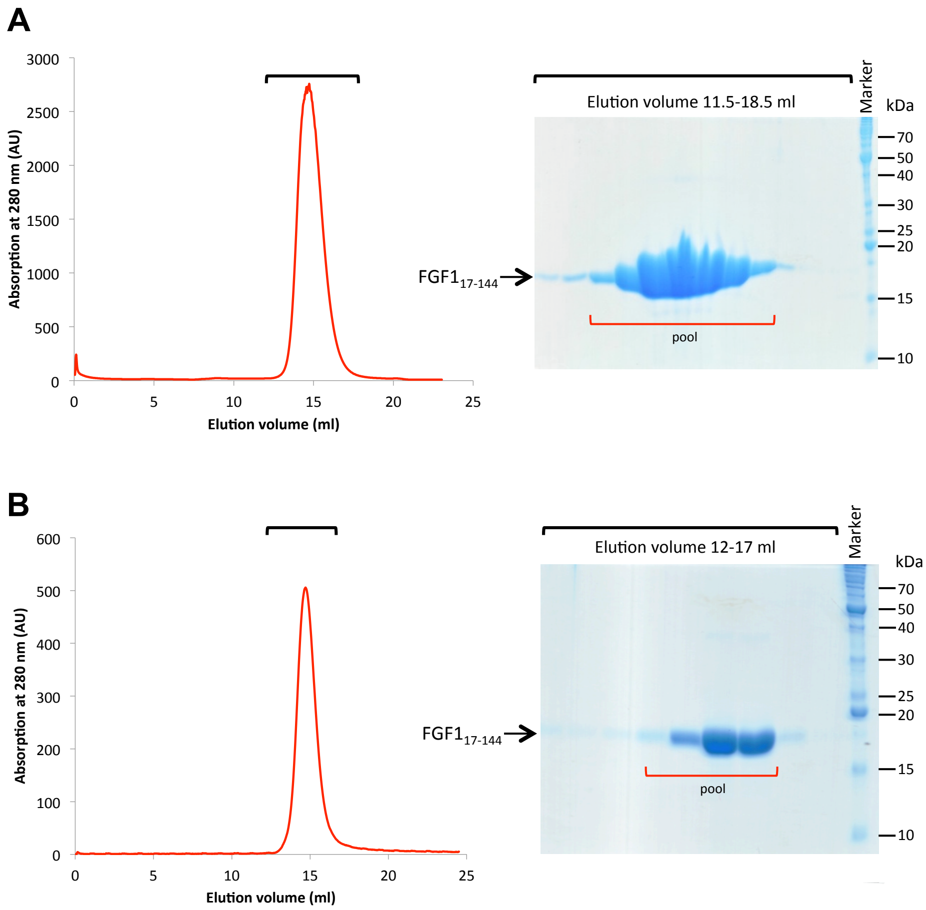


Figure 4.3 IMAC and SEC purification of FGF1

Representative purification of FGF1$_{17\text{-}155}$. (**A**) IMAC elution of FGF1$_{17\text{-}155}$ monitored by absorbance of 280 nm (red curve, left) and SDS-PAGE analysis using a Novex 4-12 %Bis-Tris gradient protein Gel (NuPAGE^{TM}, Invitrogen) and BenchMark Protein Ladder from Invitrogen as molecular weight marker (right). (**B**) SEC of R4 D2D3 using a HighLoad Sperdex S200 16/60 column (GE Healthcare Life Scineces) at a flow speed of 0.5 ml/min in buffer containing 10 mM HEPES (pH 7.5) and 150 mM NaCl. Elution was monitored as in **A**. The void and contaminant peaks are labelled individually. Black brackets indicate fractions that were subjected to SDS-PAGE analysis. Red brackets indicate fractions that were pure enough for experiments. These fractions were pooled, concentrated and stored.

4.5 Attempts at FGFR4-heparin-FGF1 complex formation

4.5.1 Glycosylated FGFR4 and FGF1 do not form binary complex

After obtaining reasonably pure protein material of FGFR4 D2D3$_{glyc}$ and FGF1$_{17-144}$, purification of a stable 1:1 receptor-ligand complex was attempted. This was supposed to form the first building block for purification of a 2:2 complexes by addition of heparin. FGFR4 D2D3 and FGF1 were mixed together in equal molar amounts at 4 °, incubated for 10 minutes, warmed up slowly to RT and injected into a Superdex S200 HR10/30SEC column. As reference, both single components were run on the same SEC column (see **Figure 4.4 A**; FGF1 as black line, FGFR4 as black dotted line). The red curve in **Figure 4.4 A** corresponds to the complex sample and shows that both components are mainly eluted as monomers. Only a vague shoulder at earlier elution volume indicates the existence of a dimer (see arrow in **Figure 4.4 A**). This was surprising as multiple biophysical studies have reported inconsistent, but reasonably high affinities for the interaction between different FGFR-FGF pairs that were measured by either isothermal titration calorimetry (ITC) (A. Brown et al., 2013) or surface plasmon resonance (SPR) (Ibrahimi et al., 2005; Kalinina et al., 2012). Values typically ranged between 300-4000 nm. In addition, multiple studies have reported crystal structures of FGFR-FGF complexes, obtained from preformed and isolated receptor-ligand dimers using the same type of SEC column as used in this study. Importantly, all these FGFR-FGF complex structures were obtained from refolded, non-glycosylated receptor protein, expressed in bacterial cell-cultures.

The crystal structure of the full signalling complex by Schlessinger et al. served as a template for generation of a FGFR4 D2D3 homology model using Modeller 9.15 ((Eswar et al., 2007); sequence alignment as presented in **Appendix B**). **Figure 4.4 B** presents a single FGFR1-FGF2 unit of the template (left) and the same unit of the FGFR4-FGF2 homology model (right). Mapping the predicted N-linked glycosylation sites (red residues) on both models shows that the 4 positions on D3 are conserved. Residues Val316 and Asn317 of FGFR1 form part of the well studied ßC'-ßE loop, which is reported to provide the ligand selectivity filter (Beenken et al., 2012). Whereas Val316 undergoes hydrophobic interactions with multiple ligand residues, Asn317 (predicted to be a target residue for glycosylation) lies in direct proximity to the receptor-ligand interface. The same conclusion seems to hold for the equivalent residues, Ile310 and Asn311, in FGFR4. Glycosylation of the respective residue might alter both position and flexibility of the ßC'-ßE loop and thereby impede heparin independent receptor-ligand interaction. Stauber et al. reported the first FGFR-FGF complex structure in 2000 (**Figure 4.4 C**). In this structure FGF1 residues Leu57 and Tyr64 form direct hydrogen bonds with Asn318 of FGFR2, an interaction of the ßC'-ßE loop with the ligand that would be impossible in presence of an N-linked glycan. In fact, this shows that the receptor-ligand interaction can be stronger in absence of the N-linked glycosylation at position N318. This could be a reason why FGFR4 $D2D3_{glyc}$ expressed in HEK293S cells has different ligand binding properties compared to refolded receptor, expressed in E.coli. Further experiments, particularly glycosylation site mutagenesis studies need to be performed in order to understand better whether N-linked glycans are the reason for the inability of FGFR4 and FGF1 to directly interact in vitro.

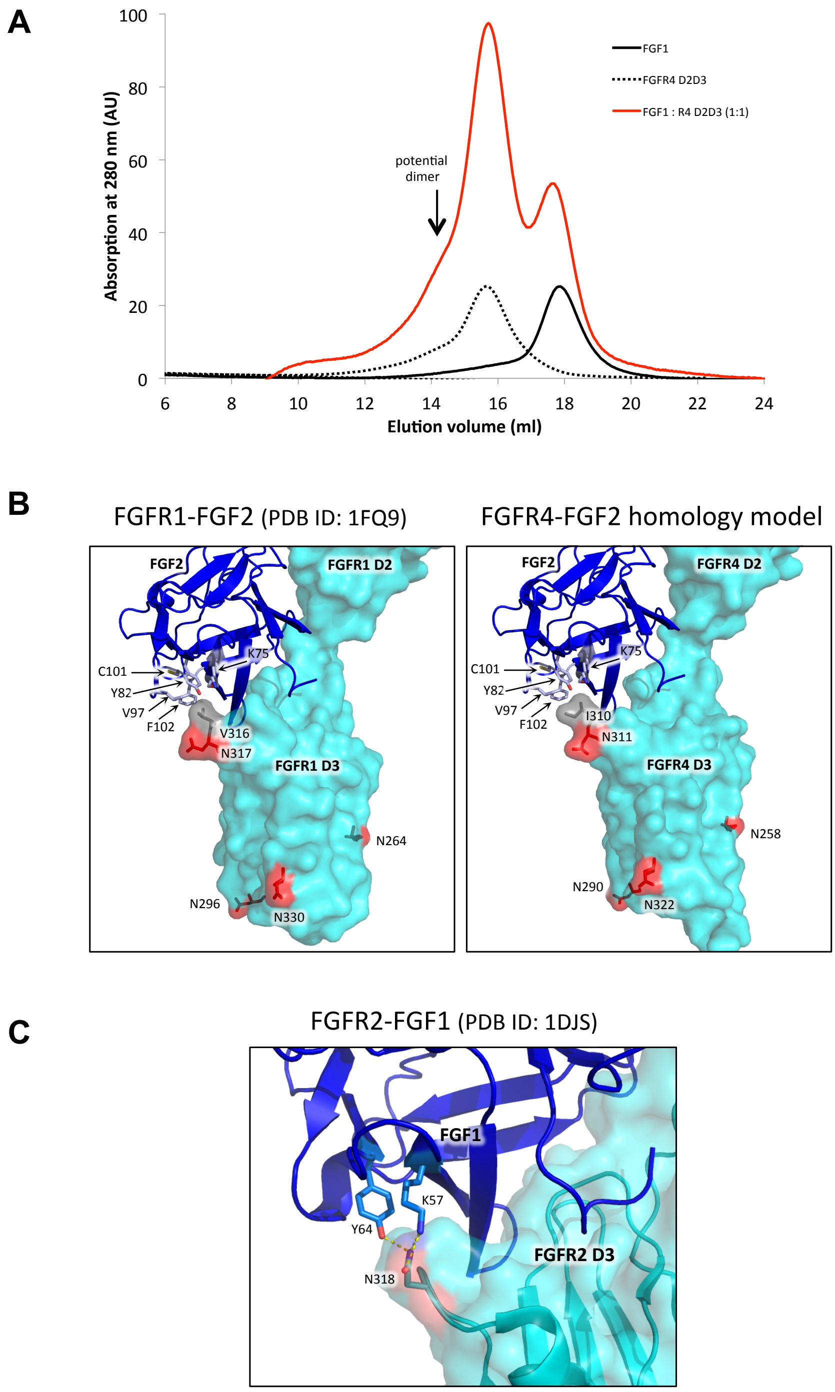


Figure 4.4 Heparin independent receptor ligand interaction

(**A**) Equal molar amounts of glycosylated FGFR4 D2D3 and FGF1 were mixed and incubated at 4 ° for 10 minutes before application to a Superdex S75 10/30 for SEC (buffer containing 10 mM Hepes pH 7.5, 150 mM Nacl). The black curve represents FGF1 alone, the black dotted line represents FGFR4 alone and the red curve represents the mixed sample. (**B**) The left panel shows the complex of FGFR1 (cyan surface) and FGF2 (blue cartoon) by Schlessinger et al., 2000. Residues of the FGFR1 ßC'-ßE loop and corresponding interacting residues on FGF2 are shown as sticks. Predicted sites for N-linked glycosylation are red. The right panel shows a FGFR4 homology model derived from the structure to the left. Colour coding for FGFR4 and FGF2 is the same as the left panel. **C**) Interface of the first FGFR structure bound to its ligand (Stauber et al., 2000). FGFR2 Asn318, which forms multiple hydrogen bonds to FGF1, is a predicted N-linked glycosylation site.

4.5.2 Heparin induced formation of FGFR4 signalling complex

After finding that receptor-ligand formation is not possible with FGFR4 D2D3$_{glyc}$, this chapter describes attempts to induce formation of a FGFR4 signalling complex by addition of heparin. As described in **Section 1.4.2**, both previously published FGFR signalling complexes have different heparin stoichiometry (Pellegrini et al., 2000; Schlessinger et al., 2000). Reports following on these two publications, suggested that the heparin binding canyon of the symmetric Schlessinger model might be occupied with only one heparin chain in physiological conditions (Harmer et al., 2004; Pellegrini, 2001), rather than with two, which would result in a receptor-heparin-ligand ratio of 2:1:2. However, since year 2000 no additional signalling complex structures have been published in either of these two architectures and clear evidence for a 2:1:2 signalling complex in Schlessinger architecture still remains to be shown. Importantly, before soaking heparin into the crystals, Schlessinger et al., grew their crystals from preformed and purified FGFR1-FGF2. It is also important to note that formation of crystal packings found in the Schlessinger and Pellegrini architectures was only possible in presence of high sulphate concentration, which presumably stabilises otherwise repulsive charges of unoccupied heparin binding sites during complex formation and crystallisation.

The aim of this work is to obtain crystals in more physiological, lower salt conditions, in presence of heparin. The stoichiometric quantity of heparin has to be chosen carefully as it has been shown by SEC experiments that excess of heparin can lead to complex dissociation (Harmer et al., 2006). Supposedly this is due to saturation of receptors and ligands with heparin, which disables heparin-mediated multimerisation. Initial complex formation screenings in this study were performed in a FGFR4-HepDP12-FGF1 ratio of 2:1:2. The reasoning is that limited supply of heparinDP12 should be compatible with ternary complex formation in both published architectures, including formation of the Schlessinger architecture for the case that the heparin binding canyon can be occupied by one single heparin molecule.

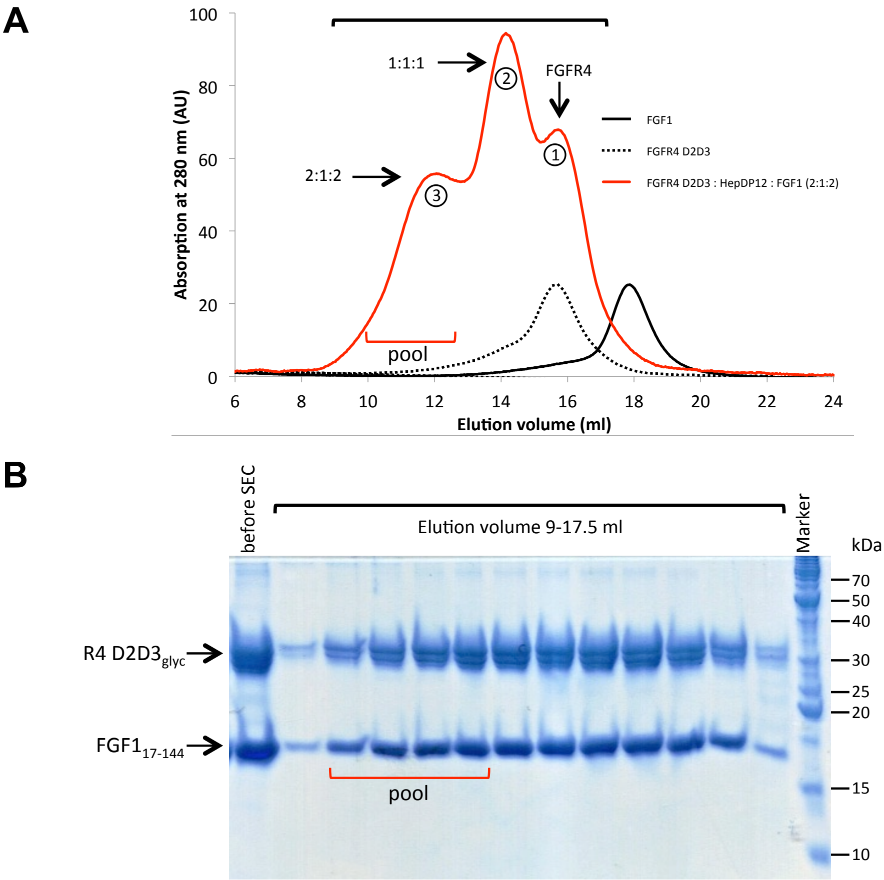


Figure 4.5 SEC analysis of a ternary FGFR4 D2D3-hepDP12-FGF1 complex

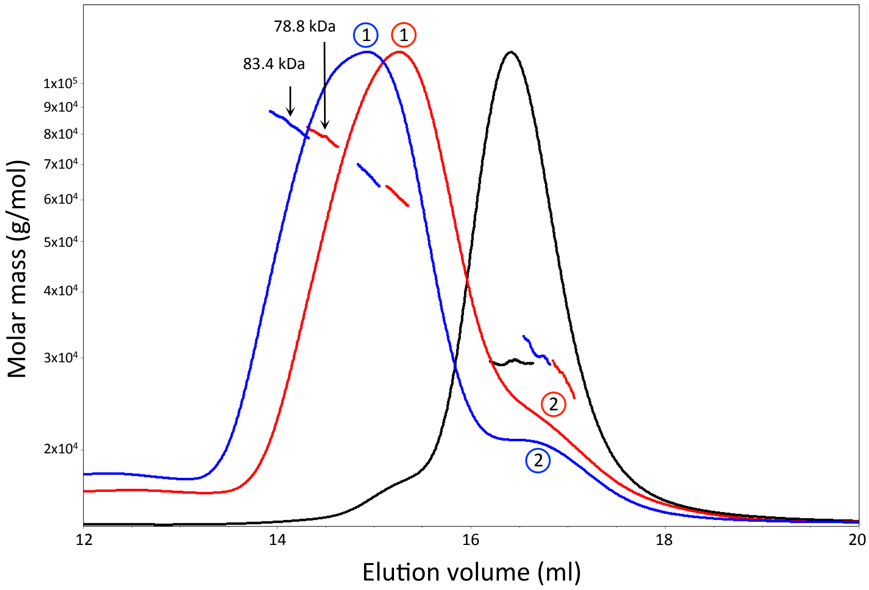
(**A**) SEC of a 2:1:2 complex prepared of FGFR4 D2D3, heparin DP12 and $FGF1_{17\text{-}144}$ using a Superdex S200 HR10/30SEC column, performed at a flow speed of 0.5 ml/min in 10 mM Hepes pH 7.5, 150 mM NaCl at RT (red curve). FGF1 (black curve) and FGFR4 D2D3 (black dotted curve) were run as standards. The elution volume that was subjected to SDS-PAGE analysis is highlighted with a black bracket at the top and the elution volume that was pooled and concentrated for further analysis is highlighted with a red bracket. (**B**) SDS-PAGE analysis of the elution volume 9-17.5 ml from SEC described in (A). The position of receptor and ligand are labelled with arrows at the left hand side. A red bracket highlights the pooled fractions for further MALLS analysis.

One molar quantity of HeparinDP12 (Iduron) (solubilised in water to a concentration of 2 nM) was mixed on ice with two molar quantities of FGFR4 D2D3, incubated at RT for 5 minutes before cooling back to 4 °. Two molar amounts of cooled FGF1 were added and the complete mixture incubated for 5 more minutes at RT before application to a Superdex S200 HR10/30 column, equilibrated in running buffer (10 mM Hepes pH 7.5, 150 mM NaCl). For comparison, both protein components were run as monomers (**Figure 4.5 A**). The elution profile of the complex sample shows three poorly separated peaks. The last peak (encircled ***3***) appears at an elution volume of approximately 16 ml, which corresponds to the elution volume of monomeric receptor with a MW of about 30 kDa. However, separation on this column is not good enough to exclude that a heparin mediated FGF1 dimer (corresponding to MW of about 37 kDa) forms part of this peak. The next peak (encircled ***2***) occurs at elution volume of 14.2 ml and is assumed to corresponds to a 1:1:1 complex of FGFR4 D2D3-hepDP12-FGF1. This suggests, that heparin is required to enable receptor-ligand interactions. The first peak eluting at around 12 ml most likely corresponds to the full FGFR4 D2D3-hepDP12-FGF1 signalling complex with stoichiometry factors 2:1:2. Fractions collected at elution volume 9-17.5 ml (black bracket at top of **Figure 4.5 A**) were analysed by SDS-PAGE. The first lane of the SDS-PAGE (**Figure 4.5 B**) shows the input sample for the SEC run, containing equal amounts of FGFR4 D2D3 and $FGF1_{17\text{-}144}$. The whole elution volume does not give any conclusion on the composition of the distinct peaks. In fact, receptor and ligand seem to be contained in all peaks in the same ratio. However, this SDS-PAGE provides evidence, that heparin mediates formation of different receptor-ligand complex species of different molecular weight.

Fractions collected at elution volume 10-12.5 ml (red bracket **Figure 4.5 B**, corresponding to peak 3, assigned to the 2:1:2 signalling complex) were pooled and concentrated with a 30 kDa MW cut off concentrator (Millipore) in order to get rid of dissociated receptor and ligand. Samples for MALLS analysis were prepared according to the protocol described in **Section 2.5**. The first sample contained monomeric receptor (black curve **Figure 4.6**) and confirm the predicted MW for glycosylated FGFR4 D2D3 of 30.0 kDa (experimental MW showed only a difference of – 0.7 kDa to the calculated value; see table in **Figure 4.6**). The predicted MW for the isolated 2:1:2 complex formed with fully glycosylated receptor is 98 kDa. The main peak of the MALLS analysis for the complex sample (see blue encircled ***1***) gave an average MW of only

66.8 kDa. This is a difference of -31.8 kDa to the expected MW, and actually corresponds to the MW of one single receptor molecule. Thus, the sample which was eluted from the MALLS machine can be assume not to contain the full 2:1:2 complex but rather an incomplete 1:1:2 complex. This is consistent with findings from mass spectrometry acceleration experiments that showed that the release of one FGFR molecule is the most favourable dissociation pathway for a 2:1:2 complex (Harmer et al., 2004). A second peak (see blue encircled ***2***) could be assigned to the dissociated FGFR4 D2D3 with a MW of 30.8. Considering the peak height of the 1:1:2 complex, a higher peak for the monomeric receptor could have been expected. The vast amount of monomeric receptor must have been separated during complex concentration with a 30 kDa MW cut off concentrator and the observed peak might originate from receptor that dissociated through the course of the SEC-MALLS experiment. There is a long slope building up to the first peak, which could not be assigned to a particular MW. However, a short stretch of this slope was selected to attempt to determine an estimated average MW for the biggest complex species. The estimate yielded a MW of 83.4 kDa, which shows that the whole peak must represents a broad mixture of 2:1:2 and 1:1:2 complexes. The small lines below or above the UV 280 absorption curve that indicate the MW, show a strong shift. This also indicates that the sample does not consist of one specific complex species with one specific MW but that the sample represents a mix of complex stoichiometries.

To test whether deglycosylation after complex formation has an effect on complex stability, the same complex sample was treated with EndoF1 as described before and applied to MALLS. The experimental MW was 60.1 kDa, 6.7 kDa less than the glycosylated equivalent. This decrease of MW corresponds to cleavage of 5.5 out of 8 N-linked glycosylations. This is consistent with incomplete deglycosylation as observed in **Figure 4.2 C**. Interestingly, the difference to the expected MW for a full deglycosylated signalling complex was – 27 kDa, which again suggests that the main complex species in this peak is a 1:1:2 complex. Peak 2 (see red encircled ***2***) again corresponds to dissociated receptor.



sample	peak	MW pred. (kDa)	MW exp. (kDa)	Δ MW (kDa)
R4 D2D3$_{glyc}$	-	30.0	**29.3**	- 0.7
R4 D2D3-HepDP12-FGF1 (2:1:2)	1	98.0 (2:1:2)	**66.8**	- 31.8
	2	30.0	**30.8**	+ 0.8
R4 D2D3-HepDP12-FGF1 (2:1:2) + EndoF1	1	87.6 (2:1:2)	**60.1**	- 27.5
	2	24.8	**27.6**	+ 3.2

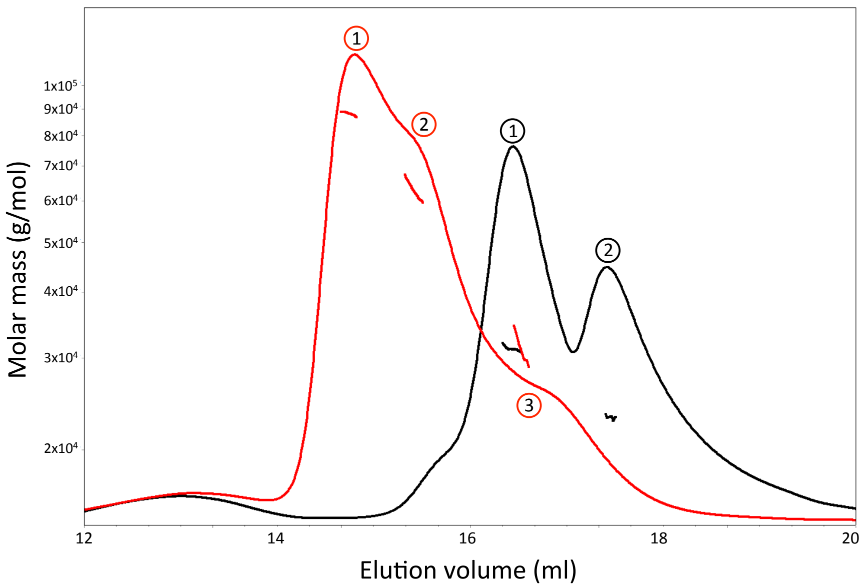
Figure 4.6 SEC-MALLS analysis of a ternary FGFR4 D2D3-hepDP12-FGF1 complex

MALLS analysis was performed to assess MW values for the highest MW complex species after SEC (see Figure 4.5). 3.5 μmol of each protein sample were diluted in 90 μl sample volume and directly injected to SEC-MALLS analysis. Size exclusion chromatography was performed on a 30 ml Superdex S200 HR 10/0 column (GE Healthcare), run by an Agilent chromatography system. The protein solution was directly eluted into an attached Dawn Helios II Multi-Angle Light Scattering detection instrument. The black curve shows UV absorption at 280 nm and the short lines below/above the peaks indicate the respective calculated MW. The curve for the monomeric receptor sample is coloured black, the glycosylated and deglycosylated complex samples are coloured blue and red, respectively. Predicted and experimentally determined values for MWs are summarised in the table. The last column presents the difference between predicted or expected MW and the experimental MW.

To summarise, FGFR4 D2D3$_{glyc}$ and FGF1 do not form binary complex in solution (**Figure 4.4 A**). Addition of heparinDP12 leads to weak complex formation but the main complex species only contains one receptor, in a 1:1:2 stoichiometry. The formation of this complex was performed according to the Pellegrini protocol: no preformation of receptor-ligand complex and limited supply of heparin. Notably, the single heparin chain in the Pellegrini model interacts with both ligands but only with one receptor (**Section 1.4.2**). The second receptor is attached to the complex by receptor-ligand interactions only. It is likely that our 1:1:2 complex is similarly formed as the single-heparin mediated Pellegrini model. As heparin supply was limited, a 2:2:2 complex formation in Schlessinger mode is rather unlikely. However, if formation of the Schlessinger complex was possible with one single heparin chain occupying the full heparin binding canyon, the core element of this complex would be formed of two receptors tightly linked by heparin. Thus, dissociation of one receptor, as observed in the presented MALLS experiment, would be highly unlikely. Notably, as reported for FGFR1 in **Section 5.4.3**, the signalling complex was significantly destabilised when mixing all complex components in a 2:2:2 stoichiometry.

4.6 Formation of more stable signalling complexes

It is important to note, that the protein complex analysed in the previous section was exposed to two Superdex S200 10/30 SEC runs before MALLS data was collected (1st SEC for complex isolation **Figure 4.5 A**, 2nd SEC as part of MALLS analysis). To test whether the ternary 2:1:2 complex is present in solution but pulled apart during the SEC run, fresh batches of receptor and ligand were prepared for analysis of complex composition by SEC-MALLS without an upstream SEC for peak isolation. In order to obtain the best possible signal, samples were ultracentrifuged for 10 minutes at 30k g before injection to the MALLS machine. The first measurement (black curve in **Figure 4.7**) is composed of 3.4 μmol FGFR4 D2D3$_{glyc}$ and 3.4 μmol FGF1$_{17\text{-}144}$. As previously presented by SEC experiments (**Figure 4.4 A**), this measurement confirms that receptor-ligand dimer formation does not occur. Peak one confirms the correct MW of the receptor with high accuracy. FGF1 has been shown in previous experiments to not remain stable as monomer on Superdex S200 columns. Therefore the MW for the FGF1 could not be determined precisely. For the 2:1:2 signalling complex sample (red curve), 1.8 μmol of heparinDP12 were added to 3.5 μmol FGFR4 D2D3 on ice and incubated at RT for 5 minutes. The mixture was cooled back to 4 °C for addition of 3.5 μmol $FGF1_{17\text{-}144}$ on ice, before incubation at RT for 5 minutes and injection to the MALLS machine. Surprisingly, the highest peak is formed by protein complex mixture, which in average has a MW of 88.2 kDa. This is still 9.3 kDa below the MW of a full 2:1:2 signalling complex, but it is evident that the average MW calculated by the MALLS machine is shifted to lower MW by the adjacent shoulder labelled with red encircled ***2***. As expected, with a MW of 63.1 kDa at the centre, this shoulder represents the 1:1:2 complex (only a difference of -4.9 kDa to the predicted MW). Again, the red line below the peak (indicating the MW) forms a strong slope, which suggests that no clear complex species is identifiable. It is rather a mixture of 2:1:2 complex, 1:1:2 complex and monomeric receptor. The bump at the tail of the curve (encircled ***3***) could be assigned to monomeric FGFR4 D2D3 with a MW of 31.3 kDa. These results are very promising at this stage, as they show the first time formation of a soluble 2:1:2 signalling complex involving glycosylated FGFR4.



sample	peak	MW pred. (kDa)	MW exp. (kDa)	Δ MW (kDa)
R4 D2D3$_{glyc}$–FGF1	1	30.0	**31.2**	+ 1.2
	2	17.0	**22.9**	+ 5.9
R4 D2D3$_{glyc}$-HepDP12-FGF1 (2:1:2)	1	98.0 (2:1:2)	**88.2**	- 9.8
	2	68.0 (1:1:2)	**63.1**	- 4.9
	3	30.0 (1:0:0)	**31.3**	+ 1.3

Figure 4.7 MALLS analysis of 2:1:2 FGFR4 D2D3$_{(glyc)}$-hepDP12-FGF1$_{17\text{-}144}$ complex

The black curve shows the MALLS measurement of 100 μl protein sample containing 3.4 μmol FGFR4 D2D3$_{glyc}$ and 3.4 μmol FGF1$_{17\text{-}144}$, performed as described in **Section 2.5**. The red curve represents the measurement of a complex sample composed of 1.8 μmol heparinDP12 , 3.5 μmol FGFR4 D2D3$_{glyc}$ and 3.5 μmol FGF1$_{17\text{-}144}$. The short lines below/over the peaks indicate the corresponding MW that were calculated for the particular selected range of elution volume. The table contains the theoretically predicted molecular weights of complexes or monomeric proteins, the experimentally assessed MW and the difference between these two values.

4.7 Crystallisation trials of various FGFR4-heparin-FGF1 complexes

4.7.1 Complex formation for crystallisation trials

As described in **Section 1.4.2**, the first published ternary FGFR signalling complex crystal structure, composed of FGFR1, FGF2 and heparinDP10 (Schlessinger et al., 2000) was obtained from a preformed, purified, binary FGFR1-FGF2 complex. In contrast, the crystals for the asymmetric structure published in the same year (Pellegrini et al., 2000) were obtained from a mix of all three soluble components. As reported in this chapter, formation of receptor-ligand dimers failed (**Section 4.5.1**) as well as attempts to isolate a sample of the full 2:1:2 signalling complex by SEC (**Section 4.5.2**). Thus protein complex samples for crystal trials were produced the same way as described by Pellegrini - by mixing together all required components in the correct molar amounts without an additional step of purification. Multiple crystal structures have revealed important biological implications of N-linked glycans (Crispin et al., 2013; Zhao et al., 2014) and alterations of glycosylation patterns on Ig domains were found to play a role in diseases (Duchesne et al., 2006; Jefferis, 2012; Winterpacht et al., 2000). However, it is also well known that each additional N-glycan on a protein increases heterogeneity in the protein sample, which can be highly unfavourable for crystallisation. As our FGFR4 D2D3 construct comprises four N-linked glycosylation sites, there would be a total of 8 flexible glycan chains in the context of the signalling complex. Therefore each crystallisation condition was tested with native receptor sample and with EndoF1 treated sample. In order to find the best procedure for deglycosylation, the glycosylation state of native complex sample (**Figure 4.8 A** lane 2) was compared to a complex sample formed with receptor subjected to EndoF1 treatment before complex formation (lane 3) and to a second sample that was deglycosylated after complex formation (lane 4). As control either 5 μg (lane 5) or 10 μg (lane 6) of protein complex were treated with PNGase F. SDS-PAGE analysis shows clearly that it does not make a difference whether N-linked glycans are cleaved off before or after complex formation – the degree of deglycosylation is the same. As deglycosylation of monomeric receptor has a destabilising effect and leads to protein loss through degradation (described in **Section 4.4**) the best way for deglycosylation was to add small catalytic amounts of EndoF1 (2 μl of 1 mg/ml EndoF1 on 100 μl protein solution) to the protein complex solution immediately before setting up crystal trials. The 60 minutes time-lapse experiment (**Figure 4.8 B**) confirmed that the deglycosylation process is terminated after 10 minutes, which therefore should not interfere with possible fast crystal formation. A pre-crystallisation test (PCT^{TM} kit, Hampton Research) was performed for each new protein complex batch in order to identify the most appropriate protein complex concentration for initial crystallisation experiments. The optimal concentration was typically 5 mg/ml or lower.

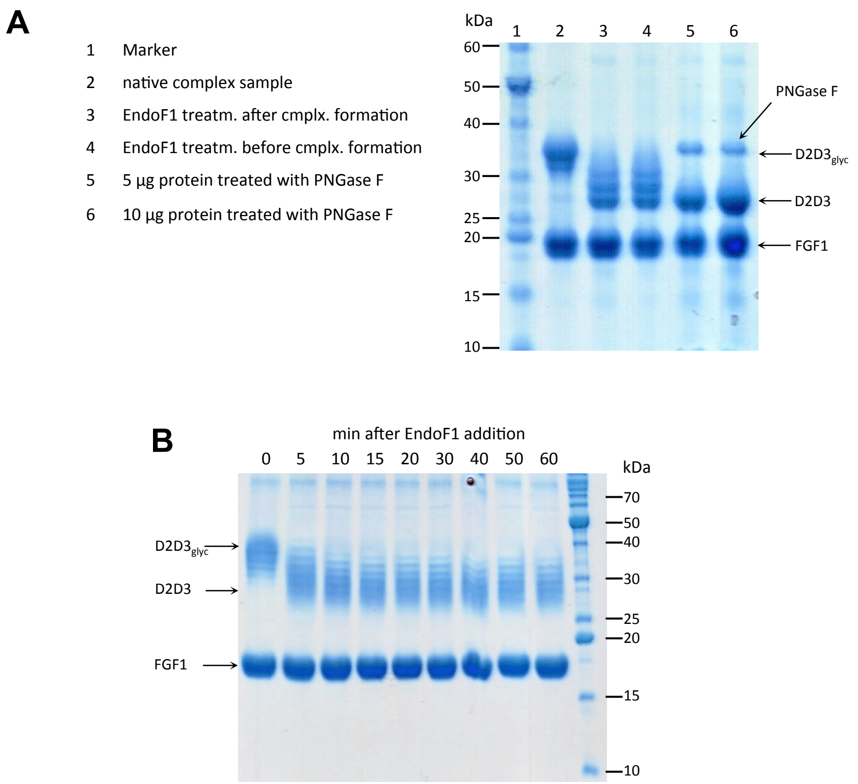


Figure 4.8 FGFR4 signalling complex deglycosylation tests

(**A**) SDS-PAGE analysis of various FGFR4 signalling complex samples before and after EndoF1 (2 μl of 1 mg/ml EndoF1 solution on 100 μl protein solution) or PNGase F treatment (New England Biolabs, according to manufacturer's protocol). See legend for detailed sample information. (**B**) Deglycosylation time-lapse experiment. Protein bands are labelled on figure. Deglycosylation was finished after 10 minutes.

4.7.2 Crystallisation trials of FGFR4 D2D3-heparinDP12-FGF1

FGFR4 D2D3-Heparin-FGF1 complex was prepared as described in **Section 4.6** and **4.7.1**. All 12 screens listed in **Table 4** were tested at 20.5 ˚ C and 6 ˚ C, with and without EndoF1 treatment. This made a total of 4608 screening conditions, each tested in sitting drop vapour diffusion experiments, set up as described in **Section 2.6.1**. Crystals were obtained in various conditions summarised in **Figure 4.9**. All crystals were flash-frozen by immersion in a cryoprotectant containing the mother liquor and then rapidly transferred to a gaseous nitrogen stream. X-ray diffraction data were recorded at the Beamline I02 or I04 at Diamond Light Source, Oxfordshire, England. Unfortunately, none of the crystals gave protein diffraction – most crystals consisted of salt or did not diffract at all.

Two additional FGFR4 constructs were cloned. FGFR4 ABD3 is a construct that encompasses the sequence from the acidic box to the Ig-domain D3. Constructs of varying length were cloned to test the effect of the acidic box length on protein solubility and secretion level in small-scale expression (as described in **Section 4.4**). The longest ABD3 construct encompassing amino acid residues Ser123-Glu354 (cloned with the oligo pair o71/o12) showed best expression and was scaled up to large-scale expression as described in **Section 2.3.3**. Purification and signalling complex formation was performed as described in **Section 4.4** and **Section 4.6** respectively. MALLS analysis did confirm the correct MW for the FGFR4 ABD3 construct and interestingly, the AB improved stability of the 2:1:2 signalling complex (**Figure 4.10 A** blue curve). Compared to the D2D3 construct (see red curve), the ABD3 showed a smaller shoulder of the 1:1:2 dissociated complex species and a smaller peak for monomeric receptor. The same screening was performed with this construct as for the D2D3 construct but to date extensive crystallisation trials have only led to growth of one crystal (**Figure 4.10 B** crystal ***1***) which was tested for diffraction and turned out to be a salt crystal.

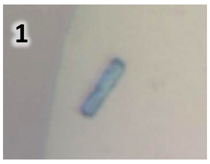
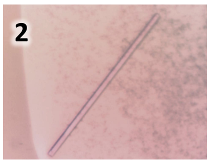
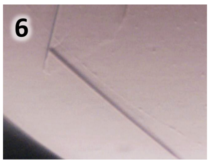
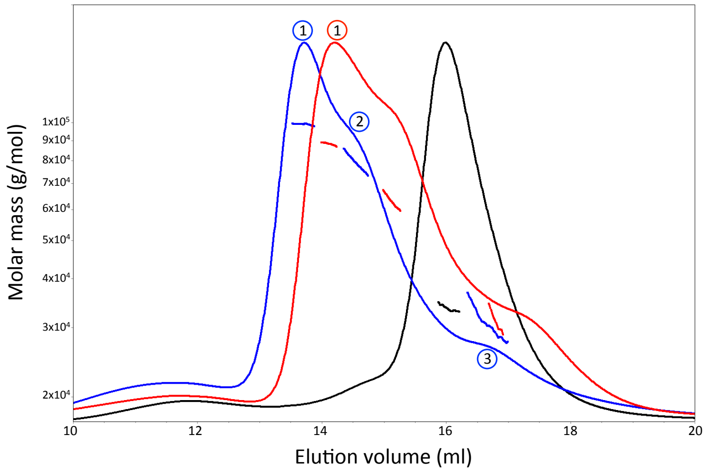
Crystal image	Crystal size (μm)	Crystallisation conditions	Conc. (mg/ml)	EndoF1 treatment
	40	20 %w/v Polyethylene Glycol 6000 0.100 M MES pH 6.0 0.200 M Calcium Chloride	5	No
	270	20 %w/v Polyethylene Glycol 10000 0.100 M HEPES-Na pH 7.5	5	No
	30	20 %w/v Polyethylene Glycol 3350 0.100 M bis-Tris Propane pH 6.5 0.200 M Sodium Fluoride	5	No
	45	25 %w/v Polyethylene Glycol 1500 0.100 M SPG System pH 6.0	3.9	Yes
	70	20 %w/v Polyethylene Glycol 3350 0.200 M Sodium Fluoride	3.9	Yes
	190	30 %v/v 2-Methyl-2,4-Pentanediol 0.040 M Magnesium Acetate 0.050 M Sodium Cacodylate pH 6.0	3.9	Yes

Figure 4.9 Crystals obtained from FGFR4 D2D3-heparin-FGF1 crystal trials

Protein solution concentration was determined by pre crystallisation trials for each complex batch. All crystals were tested for diffraction. No crystal gave protein diffraction.

In order to reduce flexibility, the flexible loop (fl) located on FGFR4 D3 (residues 287-301 comprising one N-linked glycosylation site) was deleted and replaced by linkers of different lengths. The best expressing construct was identified by small-scale expression tests, scaled up and purified for crystallisation trials. This FGFR4 D2D3-Δfl construct was subjected to the same crystallisation screening as and in the context of the signalling complex. **Figure 4.10 B** shows two further crystals (crystal ***2*** and ***3***), which again did not diffract X-rays.

A



sample	peak	MW pred. (kDa)	MW exp. (kDa)	Δ MW (kDa)
R4 $ABD3_{glyc}$	-	31.9	**31.3**	-0.6
R4 $ABD3_{glyc}$-HepDP12-FGF1 (2:1:2)	1	101.6 (2:1:2)	**99.0**	-2.6
	2	69.7 (1:1:2)	**79.4**	-9.7
	3	31.9 (1:0:0)	**29.7**	-2.2
R4 D2D3-HepDP12-FGF1 (2:1:2)	1	98 (2:1:2)	**88.2**	-9.8

B

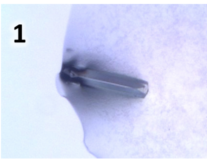
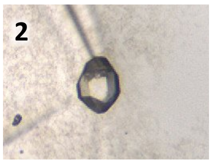
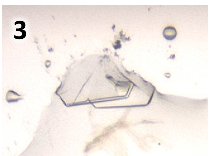
Crystal image	Crystal size (µm)	Crystallisation conditions	Conc. (mg/ml)	EndoF1 treatment
FGFR4 ABD3 complex				
1 	40	0.200 M Lithium Sulphate 0.100 M Tris pH 7.0 1.000 M Potassium Sodium Tartrate 7.600 pH Final pH	7	No
FGFR4 D2D3 Δfl. complex				
2 	270	0.200 M Lithium Sulphate 0.100 M Tris pH 7.0 1.000 M Potassium Sodium Tartrate 7.600 pH Final pH	5.4	Yes
3 	30	20 %v/v 1,4-Butanediol 0.200 M Zinc Acetate 0.100 M Imidazole pH 8.0 7.400 pH Final pH	5.4	Yes

Figure 4.10 MALLS analysis and crystallisation of FGFR4 ABD3 signalling complex

(**A**) SEC MALLS analysis as described in **Section 2.5** and **Section 4.5** of a 2:1:2 complex prepared of FGFR4 ABD3, heparin DP12 and $FGF1_{17\text{-}144}$. The black curve corresponds to monomeric receptor sample and the blue curve corresponds to the FGFR4 ABD3 signalling complex sample. The measurement of the FGFR4 D2D3 signalling complex MALLS curve presented in **Figure 4.7** is copied into this figure as reference. (**B**) Non diffracting crystals obtained from FGFR4 ABD3 signalling complex crystallisation trials (crystal **1**) and from FGFR4 D2D3-Δfl signalling complex crystallisation trials (crystals **2** and **3**).

4.8 SAXS analysis of the FGFR4-heparinDP12-FGF1 complex

As reported in the preceding chapters, the complex targeted for study does not show great stability in solution. This might be the reason why neither the Schlessinger nor the Pellegrini signalling complex structure was crystallised from an isolated signalling complex sample. More importantly, this might be the reason, why to this day no structural data from measurements in solution has been reported for any of the four FGFR homologs. FGFR4 $D2D3_{glyc}$-HepDP12-$FGF1_{17\text{-}144}$ complex samples were prepared in a 2:1:2 stoichiometry as described in **Section 4.7** in different concentrations (0.61, 0.93, 1.55 and 3.34 mg/ml) for collection of SAXS spectra at beamline BM 29 of the ESRF in Grenoble (for method see **Section 2.7**). All samples yielded spectra of good quality. The spectrum collected at 3.34 mg/ml was fitted to theoretical SAXS patterns of the Schlessinger and the Pellegrini complex structures (**Figure 4.11 A**) calculated using CRYSOL (Svergun et al., 1995). The χ^2 value was typically used as a measure for the goodness of the fit between the experimental SAXS curve and the theoretical SAXS pattern of the input model. Both crystal structures had a relatively poor fit with χ^2 = 7.25 for the Schlessinger structure and χ^2 = 10.9 for the Pellegrini structure. Importantly, the SAXS curve integrates the intensities for all molecules that are present in the soluble sample. Considering the complex dissociation that was observed in MALLS experiments, it is reasonable to assume that the SAXS curve does not only represent the 2:1:2 signalling complex, but to a fair extent the various species of dissociation, ranging from a 1:1:2 complex to monomeric receptor, FGF1 and heparin. The maximum particle dimension D_{max} (estimated using the program GNOM (Semenyuk & Svergun, 1991)) defines the longest distance between any two atoms within the measured molecule or complex. Interestingly, D_{max} is estimated at 142 Å from the recorded SAXS data, a distance that is firstly by far not present in any of the dissociation complex species and secondly not present in the Schlessinger (D_{max} = 100.2 Å) or Pellegrini (D_{max} = 106.7 Å) structures. Taking this information together, our SAXS spectrum suggests that FGFR4 $D2D3_{(glyc)}$-HepDP12-$FGF1_{17\text{-}144}$ in solution forms an arrangement that differs from both published structures, probably in a more tight complex than assumed. This new arrangement must form a maximum distance of about 142 Å, a distance that can only be bridged by two fully extended, heparin linked receptor molecules.

We were able to obtain a robust ab initio model using the bead modelling program DAMMIN (Svergun, 1999) without imposed symmetry. Intriguingly, the generated envelope presented in **Figure 4.11 B**, adopted a very symmetric shape, which indeed provided of enough volume to manually fit two units of the previously presented FGFR4-FGF2 homology model for the binary complex in a P2 symmetry-like fashion. In order to improve the fit of the high-resolution model into the low resolution SAXS envelope, CHIMERA's ***fit in map*** function was used (Pettersen et al., 2004). As supposed in the previous paragraph, both receptor molecules are maximally extended. The side view of the ab initio envelope reveals unoccupied space in the envelope, which in the suggested arrangement is the only location that could enable a single heparin molecule to interact with the heparin binding sites of both receptors and both ligands simultaneously. A red rod in the lower panel **Figure 4.11 B** represents the hypothetical position of the single heparin chain. As the ab initio model provided enough volume to fit a full 2:1:2 complex without problems, the SAXS spectrum might actually originate mainly from fully formed signalling complex

Recent studies have shown that N-linked glycosylations can have significant contributions to SAXS intensities (Felix et al., 2013; Guttman et al., 2013). In the presented SAXS experiments each FGFR4 molecule was carrying four N-linked glycans. In addition FGF1 as well as FGFR4 had flexible 6xHis tags at their C-termini. It is important to mention, that the fitted curves presented in **Figure 4.11 A** were calculated by means of unmodified Schlessinger and Pellegrini crystal structures which did not contain any flexible linker or N-linked glycans. However, since our data suggests a plausible, completely distinct arrangement of the signalling complex (**Figure 11 B**), the SAXS curve fittings are adequate to show that none of the published signalling complex architectures is compatible with our FGFR4-hepDP12-FGF1 complex data from solution. Nevertheless, the suggested extended architecture remains speculation at this stage.

4.9 Conclusions and future perspectives

This chapter has described efficient expression and purification of various glycosylated FGFR4 constructs and of FGF1. The same strategies will be applied for expression of FGFR1 constructs. SEC and SEC MALLS experiments have shown that FGFR4 D2D3$_{glyc}$ does not bind efficiently to FGF1. Addition of heparin DP12 to receptor and ligand leads to formation of a 2:1:2 signalling complex of expected MW. However, about 40 % of the protein elutes from the MALLS as partly dissociated complex. Extensive crystallisation trials of monomeric receptor and FGFR4-HeparinDP12-FGF1 mixture have not yielded diffracting crystals. SAXS data recorded from a freshly prepared signalling complex sample could not be fitted to any of the published signalling complexes. Ab initio bead modelling yielded an envelope with a D$_{max}$ value of 142 Å that easily accommodates two receptor-ligand units in a novel extended fashion. Future experiments have to focus on solving the crystal structure of FGFR4 D2D3 as monomer or as signalling complex. Application of different FGF homologs and homologous FGFR4 constructs from other species have to be considered. In addition, production of a FGFR4 D2D3-FGF1 chimera would predefine a 1:1 receptor-ligand stoichiometry, which could favour crystallisation of the 2:1:2 complex. SPR measurements of receptors with mutated glycosylation sites against FGF proteins could provide a better understanding about the role of N-linked glycans in ligand binding affinity

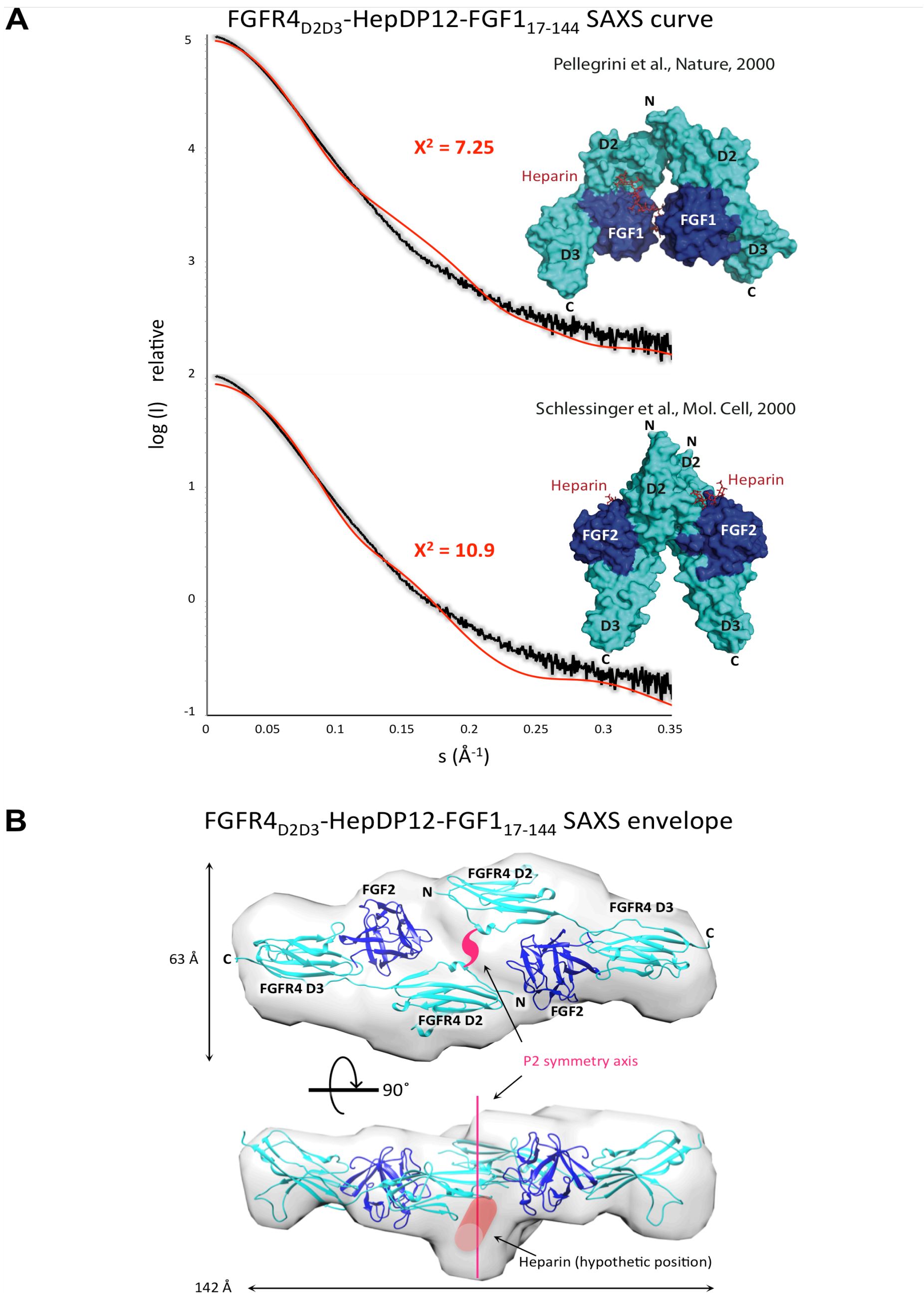


Figure 4.11 FGFR4 D2D3-HepDP12-FGF1$_{17\text{-}144}$ SAXS analysis

(**A**) Complex sample was prepared in a 2:1:2 stoichiometry as described in **Section 4.7** at concentration of 3.34 mg/ml. Scattering data was collected at beamline BM29 of the ESRF in Grenoble as described in **Section 2.7** and . Theoretical SAXS patterns of the Pellegrini (upper panel) and the Schlessinger (lower panel) complex structures were calculated using CRYSOL (**Svergun et al., 1995**) (see red curves) and fitted to the experimental SAXS curve (black curves). The corresponding goodness of fit value χ^2 is annotated above the respective curve. (**B**) As none of the crystal structures fit the experimental SAXS curve, the bead modelling program DAMMIN (**Svergun, 1999**) was used to generate ab initio model without imposed symmetry. Two units of the previously presented FGFR4-FGF2 homology model (see **Figure 4.4 B**) could be fitted into the symmetrically shaped envelope in a P2 symmetry-like fashion using CHIMERA's ***fit in map*** function (Pettersen et al., 2004). The high-resolution homology model is shown as cartoon, the receptor in cyan and FGF in blue. The side view (lower panel) reveals unoccupied space in the envelope, which forms the hypothetic position of the heparin chain, which is represented by a red rod. Dimensions of the envelope are highlighted with labelled distance arrows.

.

Chapter 5

Structural investigation of the

FGFR1$_{glyc}$–heparinDP10–FGF1 signalling complex

Chapter 5 – Structural investigation of the FGFR1-heparinDP10-FGF1 signalling complex

5.1 Summary

The arrangement of an FGFR-heparin-FGF signalling complex on the cell surface has been subject of discussion for 15 years. Biophysical and functional studies were designed to find supporting evidence for the relevance of the pseudo-symmetric, heparin canyon forming two-end architecture by Schlessinger et al., (2000) or the asymmetric architecture by Pellegrini et al., (2000). Here, a new symmetric FGFR1-heparinDP10-FGF1 signalling complex architecture with 2:1:2 stoichiometry is presented, which will be referred to as the ***new structure*** or the ***extended model***. A detailed SEC-MALLS analysis of FGF-heparin interaction and signalling complex formation preceded crystallisation and structural determination. Key features of the new structure, including crystallisation procedure and electrostatic analysis, are discussed in comparison to the Schlessinger and Pellegrini models. These results suggest that the arrangement in the new structure has most compatibility with physiological conditions on the cell surface.

5.2 Aims

The aim of the work presented in this chapter is firstly to analyse the heparin mediated FGF1 dimerisation in more detail. This information will be used to develop a better strategy for FGFR1-heparin-FGF1 signalling complex formation. Determination of this protein complex crystal structure will provide many important insights:

- Localisation of predicted N-linked glycosylation sites in context of full signalling complexes

- The detailed binding mode of heparin to FGFR and FGF
- Order of events on the cell surface leading to receptor activation
- Review of the Schlessinger and Pellegrini structures

5.3 Analysis of the interaction between FGF1 and heparin

Experiments in different cell-lines lacking endogenous heparan sulphate have shown that exogenous heparin is required for FGFR dimerisation (Spivak-Kroizman et al., 1994). Several crystal structures describing FGF molecules in monovalent complexes with heparin-like sugars have been published (Faham et al., 1996; Ornitz et al., 1995; H. Zhu et al., 1997) reporting the same binding site for glycans on FGF. Two further structural studies describe conflicting modes of heparin mediated FGF dimerisation. Herr and colleagues studied FGF2 dimerisation by analytical ultra centrifugation experiments (Herr et al., 1997) and found indications that inactive heparin-like disaccharides induce FGF2 dimers as head to head dimers (both FGF molecules are in *trans*-position to the heparin). In contrast, they found evidence, indicating that heparin octasaccharide-induced dimer complexes adopt a side-by-side architecture (both FGF molecules in *cis*-position to the heparin) under formation of direct protein-protein interactions. Thus, their conclusion was that a specific side-by-side heparin-induced FGF dimer is required for activation of the FGF receptor. Indeed, neither the Schlessinger nor the Pellegrini model is compatible with the side-by-side FGF dimer. There is no heparin-mediated FGF dimer present in the Schlessinger model and the Pellegrini model reveals an asymmetric head to head dimer. On the base of a crystal structure of human FGF1 in a complex with a fully sulphated, homogeneous heparin decasaccharide, DiGabriele and co-workers suggested that the biologically active FGF dimer binds in a symmetric head to head mode (DiGabriele et al., 1998). Furthermore, using purified FGF1–heparin complex they were able to induce promotion of tyrosine auto- phosphorylation, activation of FGFR and neurite outgrowth in living cells. Again, neither the Schlessinger nor the

Pellegrini model is compatible with this symmetric head to head FGF dimer. These structural studies and additional cell-based studies (Ornitz et al., 1992; Spivak-Kroizman et al., 1994) have established that heparin mediated dimerisation of FGF triggers productive interaction with the FGF receptor. The following experiments are aimed to understand the behaviour of FGF1 in interaction with oligosaccharides *in vitro*. Heparin oligosaccharides prepared from high-grade porcine heparin using bacterial heparinase and isolated by high-resolution gel filtration (Iduron) are applied in different lengths and in different concentrations. Thorough understanding of the heparin-binding behaviour of FGF1 and of glycosylated receptor will help to find a more sophisticated way of FGFR1 signalling complex formation for crystallisation purposes.

5.3.1 HeparinDP12 forms stable FGF1 dimer

To carefully dissect the properties of the FGF-heparin interaction with minimal bias from purification tags, a FGF1-Δtag construct was produced (deletion of the C-terminal 6xHis tag) that yielded very pure protein using heparin-affinity chromatography combined with SEC. However, a detailed NMR study on purification of FGF2 has shown that purification with heparin-sepharose chromatography is likely to contain trace contamination of heparin stripped from the resin during the elution step, which is not the case for FGF2 purified by ion-exchange chromatography (Platonova et al., 2014). These trace amounts of oligosaccharide seem to modulate the oligomerisation properties of FGF2 in solution, which was the reason why the following experiments are all performed with 6His-tagged FGF1, expressed and purified as described in **Section 4.4**. Eleven nmol freshly prepared and highly pure $FGF1_{17\text{-}144}$ were mixed with 5.5 nmol heparin oligosaccharides of different length in a 2:1 ratio. SEC analysis using a Superdex S75 column on an Äkta purifier system (see **Section 2.4.3**) was used as a read out for efficiency of complex formation and for an estimation of the oligomeric state of each complex species. A FGF1 monomer standard sample was run for comparison, which shows a monodisperse peak eluting at 14.7 ml (black curve in **Figure 5.1**). The shortest tested heparin is an octasaccharide (heparinDP8), which has been described as the shortest oligosaccharide able to promote FGF dimerisation (A. Brown et al., 2013), able to cross-link ligand and receptor for formation of a signalling complex stable in solution (DiGabriele et al., 1998; Harmer et al., 2004; Pellegrini et al., 2000) and able to induce mitogenic activity (Ornitz et al., 1992). Interestingly, the elution profile of the FGF1-heparinDP8 sample shows that just about 70 % of the total protein sample has formed a dimer (dotted curve). The next tested heparin was a dodecasaccharide (heparinDP12), which was used in **Section 4.6** for formation of FGFR4 signalling complex although both published signalling complex structures contained a heparinDP10. This was due to the reason that a crystal packing as observed by Pellegrini and co-workers could have revealed more heparin-protein interactions in presence of a longer heparin. At the same time heparinDP12 would have been more likely to be able to occupy the whole positively charged heparin canyon of the Schlessinger structure (instead of occupancy with two heparin chains). Dodecasaccharide (red curve in **Figure 5.1**) shows much more efficient dimer formation compared to heparinDP8, which strongly suggests the presence of a stable FGF1-heparin complex in a 2:1 ratio without formation of higher molecular weight complexes. Only10 % of the total FGF1 remain unliganded. Finally, 11 nmol FGF1 were mixed with 2.7 nmol heparinDP20, resulting in a molecular FGF-heparin ratio of 4:1. This ratio was chosen as previous studies reported that two signalling complexes can be formed on a heparin chain with 20 saccharide units (Harmer et al., 2006), which would probably involve formation of two FGF dimers mediated by one single heparin chain. Indeed, around 80 % of the total FGF1 forms a complex of higher MW, which could correspond to a 4:1 FGF-heparin complex, the remaining protein seems to stay monomeric. In conclusion, fully sulphated heparin 12-mer leads to nearly full dimerisation of $FGF1_{17\text{-}144}$. The reason for about 10 % non-dimerised FGF1 could be inaccuracies of the heparin stock concentration, resulting in limited heparin supply. Heparin oligosaccharides are delivered as powder in amounts of 2 mg (weighed by a scale with a typical minimum weighing capacity of 0.1 mg) and there is no assay to assess the accurate concentration of a heparin solution. Thus, it is difficult to prepare heparin stocks with precisely defined concentration.

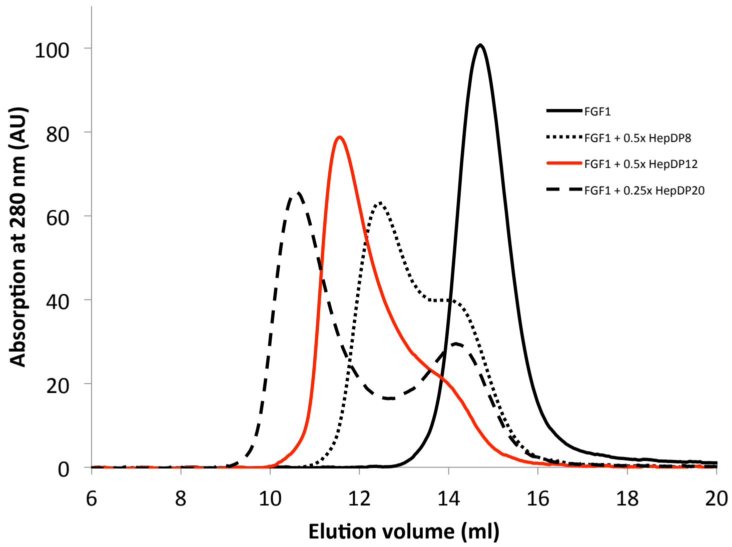


Figure 5.1 SEC screening of heparin length dependent dimerisation of FGF

Analysis of $FGF1_{17\text{-}144}$ dimer formation with heparinDP8 (dotted curve), heparinDP12 (red curve) and heparinDP20 (dashed curve). For each sample, including the monomeric FGF1 sample, 11 nmol FGF1 were mixed in the respective molar ratio of heparin as indicated in the figure legend. SEC analysis was performed using a Superdex S75 column on an Äkta purifier system, tracking protein absorption at 280 nm of eluting material (see **Section 2.4.3**).

5.3.2 HeparinDP12 mediated FGF1 dimer formation is cooperative

The differences in architecture and heparin stoichiometry between the Schlessinger and Pellegrini signalling complex structures have not contributed to a better understanding of the order of events for receptor activation on the cell surface. The Schlessinger structure suggests that heparin crosslinks a receptor-ligand unit and thereby promotes assembly of two 1:1:1 units to the 2:2:2 signalling complex. In contrast, the central heparin mediated FGF1 dimer in the Pellegrini structure strongly suggests that two receptor molecules join a preformed ligand complex. The latter hypothesis is backed up by two studies arguing that FGF1 dimerisation on a single heparin chain (ligand-heparin ratio of 2:1) is energetically favoured over formation of 1:1 ligand-heparin complexes (A. Brown et al., 2013; Robinson et al., 2005), a cooperative FGF1 dimerisation mechanism that could explain how a signalling complex in 2:1:2 ratio forms at the cell surface.

The simple SEC screen in the previous section has indicated that heparinDP12 drives FGF1 dimerisation more efficiently than heparinDP8. Heparin DP12 was therefore used here to screen dimerisation efficiency for a defined amount of FGF1 (8 nmol per experiment) over a wide range of oligosaccharide concentrations. The black curve in **Figure 5.2** shows the peak of pure, monomeric FGF1. As described in the preceding section, the adjustment of heparin stock concentration brings some challenges but plays a major role for all complex formation experiments. In order to test whether the concentration of the heparin stock is accurate, 0.25-fold amounts of heparin were added to a 1-fold amount of FGF1, which should lead to dimerisation of only half of the applied protein and thereby form two peaks of the same height. In fact, two peaks of very similar height are observed (see green curve in **Figure 5.2**). The monomeric FGF1 peak is slightly higher, which suggests that the heparin concentration is marginally lower than expected but accurate enough to go ahead with further experiments. As expected, the sample with 0.5-fold amount of heparin (red curve) mainly yielded FGF dimer. However, the shoulder to the main peak at higher elution volume indicates that 5-10 % of the total protein is not dimerised. As a result, all experiments for signalling complex formation in following sections were performed with a 1.1 fold excess of heparin (corresponding to addition of 10% to the required heparin amount) in order to avoid weakened complex formation due to inaccuracies of heparin stock concentration. Next, to test cooperative dimerisation of FGF1 upon a single heparin saccharide a sample was prepared with a 1-fold amount of heparin. In case of cooperative FGF1 dimerisation, formation of a 2:1 FGF1-heparin complex is preferred to a 1:1 complex. The elution profile (purple curve) shows that 100 % of the protein elutes as a ligand dimer confirming that FGF1 dimerises cooperatively upon addition of heparinDP12. Even 4-fold excess of heparin (blue curve) does not lead to dimer dissociation, suggesting that FGF1 is available as dimer even at heparin enriched environments such as the cell surface.

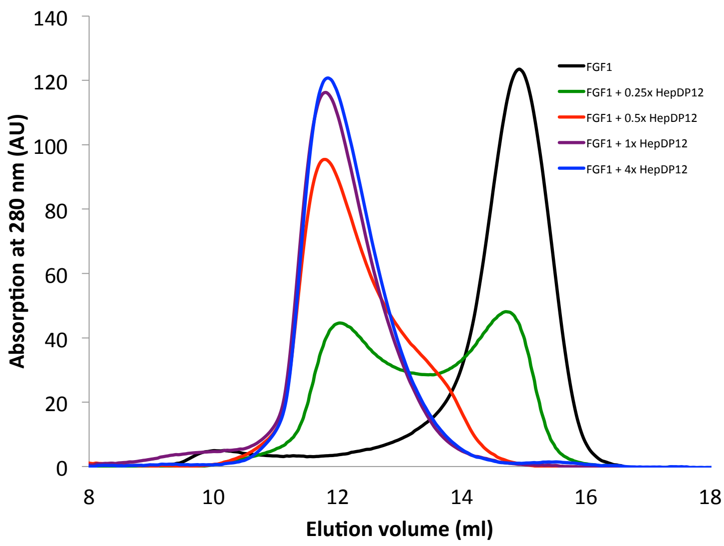


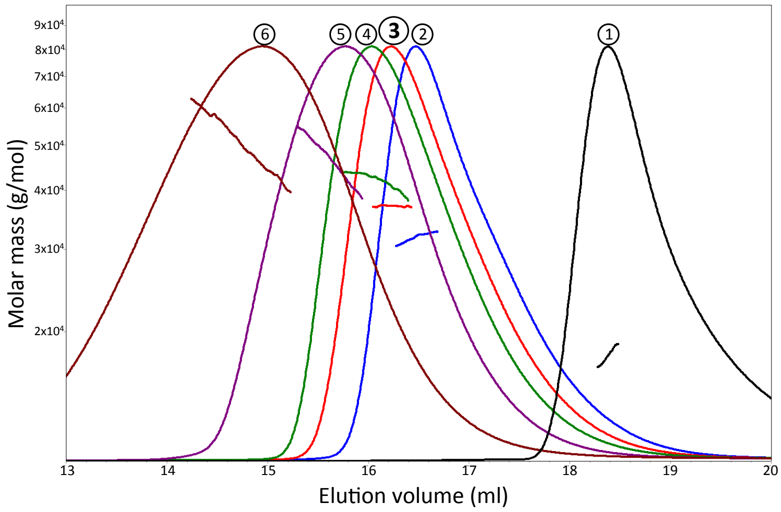
Figure 5.2 SEC screening of heparin concentration dependent dimerisation of FGF

FGF1 samples of 8 nmol were mixed with increasing amounts of heparinDP12 and applied to SEC (Superdex S75 10/30 column) to test for complex formation. The curves are coloured black for pure FGF1, green for 2 nmol heparin, red for 4 nmol heparin, purple for 8 nmol heparin and blue for 32 nmol of heparin.

5.3.3 SEC-MALLS identification of most homogeneous FGF1 dimer

The previous analysis of heparin mediated FGF1 dimerisation by size exclusion chromatography did not provide accurate information on the molecular weight of complex species. SEC-MALLS not only provides accurate estimations of the average MW of selected peak sections, it also indicates whether a peak represents one single complex species or whether it represents a gradient of complexes that differ in their composition. **Figure 5.3** presents the results of SEC-MALLS experiments containing 6 nmol $FGF1_{17\text{-}144}$ supplemented with 6 nmol heparin oligosaccharides with lengths ranging from DP8 to DP30. The FGF1-heparin ratio of 1:1 was chosen to ensure that the calculated molecular weights from MALLS experiments are not influenced by dissociated

FGF. The difference between predicted and experimental MW of a putative heparin mediated FGF1 dimer and the inclination of the MW line on the MALLS graph are used as a measure for the dimer stability. The black curve in **Figure 5.3** (corresponding to peak 1, see first raw values in the table) shows the monomeric FGF1 standard measurement. The predicted MW of 17 kDa (see third column of table) is confirmed with an experimental MW of 18 kDa (see fourth column of table), which makes a difference of +1.0 kDa (see fifth column of table). HeparinDP20 (purple curve, encircled ***5***) and heparinDP30 (brown curve, encircled ***6***) yielded average MWs of 46.9 kDa and 51.3 kDa respectively, which are 6.7 kDa and 8.3 kDa above the expected MWs. More importantly, the strong inclination of the respective MW lines at the selected peak areas indicates that there is no definite complex species present, but rather a broad mixture ranging from FGF dimers to FGF tetramers. Comparing the heparinDP8, heparinDP10 and heparinDP12 samples (blue, red or green curves respectively) the horizontal MW curve clearly indicates that heparinDP10 forms the most homogeneous and stable dimer. Strikingly, the FGF-heparinDP10 complex also has the best agreement between predicted (37.0 kDa) and experimental (37.1 kDa) molecular weight. As a result of this analysis, the subsequent experiments will mainly be performed with heparinDP10.



Heparin-mer added to FGF1 And respective MW (g/mol)	peak	MW pred. for dimer (kDa)	MW exp. (kDa)	Δ MW (kDa)
No heparin, only FGF1	1	17.0 (monomeric)	**18.0**	+ 1.0
DP8 (2400)	2	36.4	**31.8**	**+ 5.9**
DP10 (3000)	3	37.0	**37.1**	**+ 0.1**
DP12 (3550)	4	37.6	**42.1**	**+ 4.5**
DP20 (5750)	5	39.8	**46.9**	**+ 6.7**
DP30 (9000)	6	43.0	**51.3**	**+ 8.3**

Figure 5.3 SEC-MALLS analysis of FGF oligomerisation

SEC-MALLS was used for identification of the ideal heparin length for FGF1 dimerisation. Each sample consists of 100 μl containing 6 nmol of FGF1 mixed with 6 nmol of heparin 8-mer (blue curve), heparin 10-mer (red curve), heparin 12-mer (green curve), heparin 20-mer (purple curve), heparin 30-mer (brown curve) or no heparin (black curve). 90 μl solution were injected to a Superdex S75 10/30 column previous to MALLS analysis. The results are listed in the table according to heparin length (see increasing degree of polymerisation in first column). The second column lists the sample indexes, also indicated on top of each peak. The third and fourth columns list the predicted MW and the experimental MW. The difference between these two values is listed in the fifth column and is used as an indicator for FGF1 dimer stability.

5.4 FGFR1 signalling complex formation

5.4.1 Expression and purification of FGFR1

According to the NetNGlyc 1.0 Server the ectodomain of FGFR1 has eight N-linked glycosylation sites and is thereby the hFGFR homolog with most glycosylation sites. For the purpose of complex formation studies and crystallographic studies, construct boundaries were chosen to produce proteins comprising Ig domain 2 (D2) and Ig domain 3 (D3). Glycosylation sites N227 and N240 are located on D2 and glycosylation sites N264, N297, N317, N330 are located on D3 (**Figure 5.4** A). Just like for the design of FGFR4 constructs (**Section 4.3**), boundaries for the FGFR1 constructs were chosen based on sequence alignments with human homologous receptors (see sequence alignment in **Appendix B**), on secondary structure prediction and on published structures of FGFR1 and FGFR2 (Beenken et al., 2012; Plotnikov et al., 2000; Schlessinger et al., 2000). Six constructs (FGFR1 D2D3-01 to -06) with slightly varying N-terminal and C-terminal boundaries were generated (see **Figure 5.4 B** for boundaries), in order to identify the best expressing construct in small-scale expression tests. The detailed construct design strategy is described in **Section 2.2** including choice of expression vector and cloning sites. The following primer pairs were used to generate the respective FGFR1 D2D3 constructs: D2D3-01 with o65/o66, D2D3-02 with o65/o67, D2D3-03 with o65/o68, D2D3-04 with o64/o66, D2D3-05 with o64/o67, D2D3-06 with o64/o68 (see primer sequences in **Appendix C**). Western blot analysis of small-scale expression tests (**Section 2.3.2**) of these constructs identified construct D2D3-04$_{141\text{-}365}$ as the best expressing construct (**Figure 5.4 B**). This construct was expressed in large scale and purified the same way as described for FGFR4 D2D3 constructs in **Section 4.3**. A representative elution profile and the corresponding SDS-PAGE analysis are shown in **Figure 5.4 C**. SEC could remove the impurity at 120 kDa and only fractions containing reasonably pure FGFR1 protein were pooled (see red bracket on **Figure 5.4 C**). The first lane in **Figure 5.4 D** shows the final product. The following lanes show a time lapse of the same sample after addition of EndoF1 (as described in **Section 4.3**). Deglycosylation of FGFR1 D2D3 is terminated after 10 minutes. Only a very weak band of single glycosylated receptor remains, which makes an estimated amount of 5 % of the total protein. The predicted MW for fully glycosylated D2D3$_{141\text{-}365}$ is 34502 Da and 26464 Da for fully deglycosylated receptor. SDS-PAGE analysis can confirm correct MW for glycosylated and deglycosylated receptor within the typical range of accuracy (SEC-MALLS analysis will allow determination of MW with higher accuracy). Just like FGFR4 D2D3 constructs, FGFR1 D2D3$_{141\text{-}365}$ did also precipitate slightly through the process of deglycosylation. Ultracentrifugation was used to remove the precipitated protein that reoccurred after storage over night at RT or 4 °C. Concentration of deglycosylated receptor in 15 or 30 kDa MW cut off concentrators was also always accompanied with loss of about 20 % of the total protein.

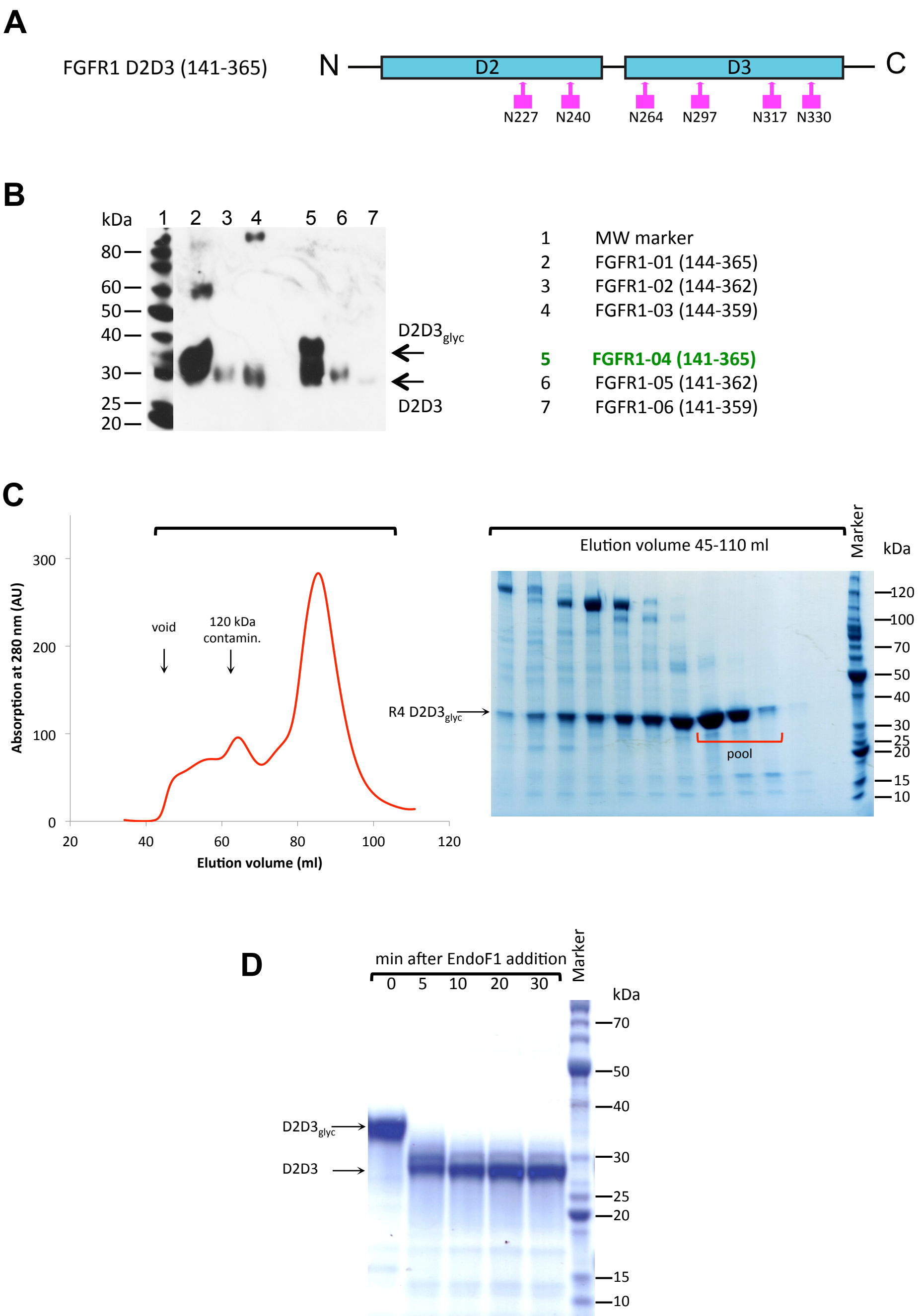
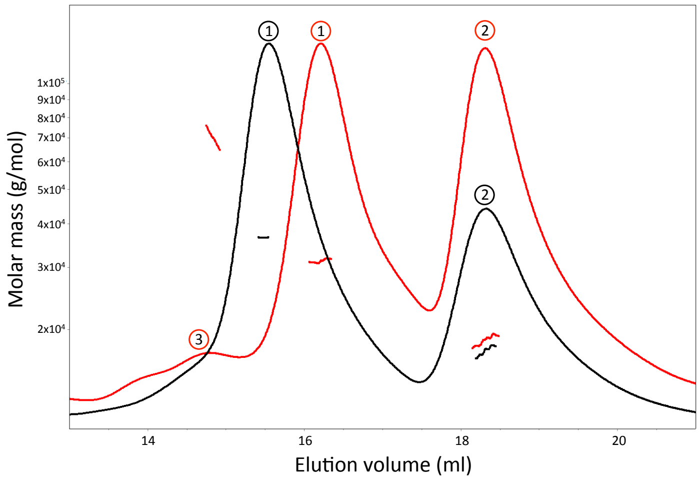


Figure 5.4 Design and purification of FGFR1 D2D3 constructs

(**A**) Domain structure of a FGFR1 D2D3 construct, indicating positions of N-linked glycosylation sites. (**B**) Western blot analysis of small-scale expression tests for FGFR1 D2D3-01 to -06. The labelling describes the exact domain boundaries of each construct. (**C**) Elution profile of a SEC run using a Superdex S200 16/60 column on an Äkta purifier system. Void peak and typical contamination of 120 kDa are highlighted with small arrows. SDS-PAGE analysis of the fractions corresponding to elution volume 45-110 shows that the main peak mainly consists of FGFR1. Fractions within the red bracket were pooled and concentrated using a 30 kDa MW cut off concentrator (Millipore). (**D**) EndoF1 (2 μl of 1 mg/ml EndoF1 solution on 100-μl protein solution) was added to pure protein sample and the degree of deglycosylation was analysed by means of SDS-PAGE in specific time steps. The first lane shows the D2D3 construct before addition of EndoF1, which corresponds to sample at time 0. The respective sampling time for each sample is indicated above each lane. The deglycosylation process is terminated after 10 minutes.

5.4.2 EndoF1 Deglycosylated FGFR1 does not bind to FGF1

In **Section 4.5.1** it was shown by size exclusion chromatography that glycosylated FGFR4 D2D3 expressed in HEK293S cells does not form a binary complex with FGF1. Thus, it was not possible to set up crystal trials according to the protocol described by Schlessinger and co-workers (2000) by preformation of a 1:1 receptor-ligand complex. In this work it has not succeeded to find out whether this is a special characteristic of FGFR4, whether it is a result of the presence of N-linked glycans, or whether both or none of those points play a role. **Figure 4.4 B** presents the mapping of glycosylation sites on domain D3 of FGFR1 (and on the homology model of FGFR4), which only comprises four of six total glycosylation sites predicted for the D2D3$_{141\text{-}365}$ construct (see **Figure 5.1 A**). This mapping revealed that the glycosylation sites N227 on Ig domain 2 and N317 on Ig domain 3 are in direct proximity to the ligand-binding site and could potentially cause sterical hindrances for FGF1 binding. Here SEC-MALLS is used to assess whether FGFR1 D2D3 is able to form a binary receptor-ligand complex. 3.5 nmol of fully glycosylated FGFR1 D2D3$_{141\text{-}365}$ were mixed with the same molar amount of pure FGF1$_{17\text{-}144}$ for injection to the SEC-MALLS machine. This sample (**Figure 5.5** black curve) formed two peaks that correspond to the MWs of fully glycosylated monomeric receptor (peak 1) and to monomeric FGF1 (peak 2). A dimer peak could not be observed. In order to assess the effect of N-linked glycans on the affinity to FGF1 in absence of heparin, binary complex formation was also tested using EndoF1 treated receptor sample (red curve; 2-fold excess of FGF1 with 7 nmol of purified ligand). Again the sample was eluted in two peaks that could be assigned with acceptable accuracy to deglycosylated FGFR1 D2D3 (ΔMW between predicted and experimental values adds up to 4.8 kDa; SDS-PAGE analysis of FGFR1 deglycosylation in **Figure 5.4 D** has shown that EndoF1 does not lead to absolute deglycosylation) and to FGF1. A dimer of EndoF1 deglycosylated receptor and FGF1 has a predicted MW of 51.5 kDa. A very small peak labelled as peak 3 could be found with a MW of 70.1 kDa. This is 18.6 kDa heavier than the expected dimer and the equivalent of a putative 1:2 FGFR1-FGF1 complex. Formation of this complex in absence of heparin is highly unlikely. In addition, the line that indicates the MW at the selected peak region shows a very strong slope suggesting that peak 3 is rather formed by degraded receptor or other contaminations. In conclusion this experiment suggests that N-linked glycans do not have a direct influence on FGFR1-FGF1 dimer formation. This has to be taken into consideration for design of crystallisation experiments.



sample	peak	MW pred. (kDa)	MW exp. (kDa)	Δ MW (kDa)
R1 D2D3$_{glyc}$-FGF1 (1:1)	1	34.5	**36.4**	+ 1.9
	2	17.0	**17.4**	+ 0.4
R1 D2D3-FGF1 (1:1) + EndoF1	1	26.5	**31.3**	+ 4.8
	2	17.0	**18.7**	+ 1.7
	3	51.5	**70.1**	+ 18.6

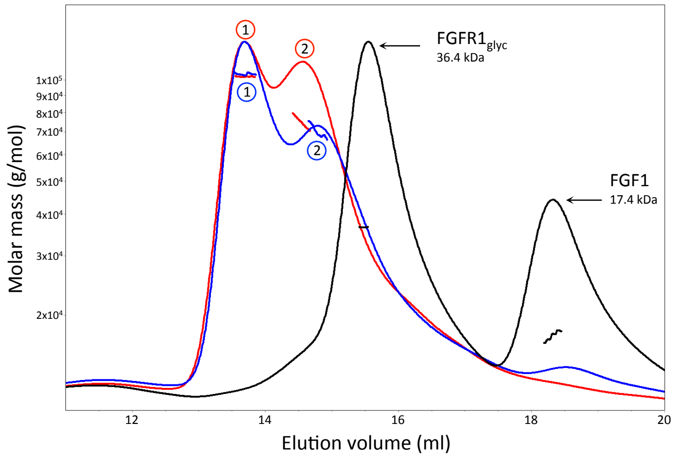
Figure 5.5 SEC-MALLS analysis of FGFR1–FGF1 dimer formation

Two FGFR1 D2D3 samples were tested for dimer formation. 3.5 nmol of glycosylated FGFR1 D2D3$_{141\text{-}365}$ were mixed with 3.5 nmol of FGF1$_{17\text{-}144}$ and injected to a Superdex S200 16/60 for MALLS analysis (black curve). Peaks are numbered on the elution profile and respective predicted and experimental MWs are listed in the table below. The last column lists the difference between these two values. The second sample contained 3.5 nmol of FGFR1 D2D3$_{141\text{-}365}$ (treated with EndoF1 for deglycosylation) and 7 nmol of FGF1 (red curve). Peaks are labelled 1-3 and respective MW values are listed in the red area of the table. For detailed description of the method see **Section 2.5**.

5.4.3 HeparinDP10 in 2:1:2 stoichiometry forms most stable signalling complex

As described in **Sections 5.3.1** to **5.3.3** the choice of heparin saccharide length determines complex stability and the formed FGF-heparin complex species. Experiments presented in **Section 5.3.3** clearly identified heparinDP10 as the favoured length for stable FGF1 dimerisation. Whether heparinDP10 is the ideal saccharide length for formation of a full signalling complex needs to be determined experimentally. Our studies on FGFR4 signalling complex formation (described in **Section 4.6**) have shown that it is not possible to isolate stable complex by size exclusion chromatography, suggesting that the FGFR4 signalling complex is rather unstable. This needs to be tested for FGFR1. According to the Schlessinger model, two heparinDP6 oligosaccharides in a 2:2:2 stoichiometry (or a single heparinDP12 oligosaccharide in a theoretical 2:1:2 stoichiometry with one heparin occupying the complete heparin binding cleft) should be sufficient for formation of a stable signalling complex. According to the Pellegrini model one heparinDP10 is sufficient for formation of the full complex. The following SEC-MALLS experiments screen complex formation under three different conditions that test potential formation of any of these two published signalling complex architectures. By comparison of the peak sizes, the fraction of formed signalling complex is tested with heparinDP10 versus heparin DP12 (**Figure 5.6 A**) and in a second step, the favourable heparin is used to test signalling complex fraction in 2:1:2 stoichiometry (Pellegrini) versus 2:2:2 stoichiometry (Schlessinger) (**Figure 5.6 C**). Predicted and experimentally determined MWs are listed in the table (**Figure 5.6 B**). All complex samples were formed as described for FGFR4 (**Section 4.6**) with the difference that molar amounts of heparin were applied in 1.1-fold (adding 10 %) in order to avoid deficits of complex formation due to inaccuracies of heparin stock concentrations (according to observations from experiments in **Section 5.2**).

A



B

sample	peak	MW pred. (kDa)	MW exp. (kDa)	Δ MW (kDa)
R1 D2D3$_{glyc}$-HepDP12-FGF1 (2:1:2)	1	106.6 (2:1:2)	**102.0**	- 4.6
	2	72.1 (1:1:2)	**74.8**	+ 2.7
R1 D2D3$_{glyc}$-HepDP10-FGF1 (2:1:2)	1	106.0 (2:1:2)	**103.9**	- 2.1
	2	71.5 (1:1:2)	**70.4**	- 1.1
R1 D2D3$_{glyc}$-HepDP10-FGF1 (2:2:2)	1	109.0 (2:2:2)	**99.0**	- 10
	2	71.5 (1:1:2)	**67.0**	- 4.5
	3	34.5 (1:0:0)	**34.4**	- 0.1

C

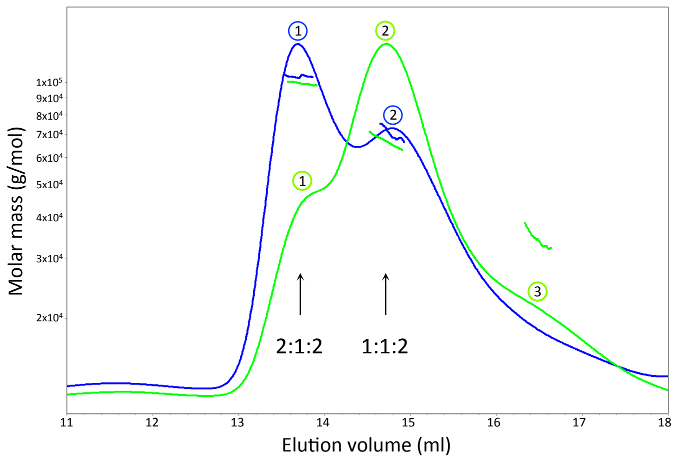


Figure 5.6 SEC-MALLS screening for signalling complex formation

Heparin length and concentration were varied to identify the best conditions for FGFR1-FGF1 signalling complex formation. In order to balance out potential inaccuracies of heparin stock concentrations, the molar amounts of heparin, calculated according to the specific stoichiometric ratios, were applied in 1.1-fold amount. SEC-MALLS analysis was performed as described in **Section 2.5**. (**A**) Two signalling complex samples were formed of FGFR1 D2D3-heparin-FGF1 in 2:1:2 ratio (3.5nmol : 1.93nmol : 3.5nmol) using heparinDP10 (blue curve) or heparinDP12 (red curve). Both main peaks are numbered in the colour of the corresponding curve. The elution profile of the FGFR1 D2D3-FGF1 sample (black curve) from **Section 5.4.2** is added as reference, labelled with the corresponding MW of both peaks. (**B**) List of sample contents, predicted MWs, experimental MWs and the difference between predicted and experimental MWs of signalling complex samples. (**C**) Comparison of the FGFR1 D2D3-heparinDP10-FGF1 2:1:2 sample (blue curve) with a sample formed of the same components in 2:2:2 Schlessinger stoichiometry (3.5nmol : 3.85nmol : 3.5nmol) (green curve). Three peaks are highlighted in green circles. Two arrows highlight the expected peak positions of a 2:1:2 and a 1:1:2 complex.

FGFR1-heparin-FGF1 2:1:2 (3.5nmol : 1.93nmol : 3.5nmol) signalling complex SEC-MALLS analysis are presented for heparinDP10 (blue curve) and heparinDP12 (red curve). Both samples formed a main peak (labelled with red or blue encircled ***1***) that could be assigned to a full signalling complex. In fact, the complex formed with heparinDP10 has the more accurate MW with a difference between predicted MW (106.0 kDa) and experimentally determined MW (103.9 kDa) of only - 2.1 kDa. But more importantly the secondary peak (labelled with blue encircled ***2***), which could be assigned to a 1:1:2 complex (one dissociated receptor) with even higher accuracy, is much less prominent compared to the secondary peak formed by the heparinDP12 sample (labelled with red encircled ***2***). This clearly indicates that signalling complex is more stable when formed with heparinDP10. Thus, for comparison of the Pellegrini stoichiometry with the Schlessinger stoichiometry a FGFR1-heparinDP10-FGF1 2:2:2 (3.5nmol : 3.85nmol : 3.5nmol) signalling complex sample was formed and applied to SEC-MALLS analysis (**Figure 5.6 C**). The direct comparison clearly shows that the stoichiometry found in the crystal structure by Schlessinger and co-workers leads to dissociation of the complex in solution (green curve). The detected MW for the signalling complex peak is 10 kDa below the expected MW (see **Figure 5.6 B**, green encircled ***1***) which must be the result of the adjacent lower MW peak bleeding into the signalling complex peak and thereby lowering the average MW. The main peak (green encircled ***2***) can clearly be assigned to the dissociated 1:1:2 complex with a difference of -4.5 kDa to the expected MW. Interestingly, a third peak (green encircled ***3***) could be assigned to the dissociated monomeric receptor with correct molecular weight.

These results not only give information on how to best form the complex for crystallisation purposes, but also give insights into possible ways of complex formation at the cell surface. Heparan sulphate proteoglycans (HSPG) are known to be ubiquitous and more abundant than most receptors at mammalian cell surfaces (reviewed in (Bernfield et al., 1999)). Experiments in **Section 5.3.2** confirm that heparin mediated dimerisation of FGF1 is cooperative. HS chains move freely in all dimensions and are likely to trap ligands when approaching the cell surface before they get closer to the membrane. This could result in FGF dimerisation and thereby activation for interaction with two membrane anchored FGFRs. This is further supported by the observation that FGF1 only interacts with FGFR in the presence of heparin (see **Section 5.4.2**).

5.5 Crystallisation of FGFR1-heparinDP10-FGF1

5.5.1 Initial crystallisation screening

Fresh samples of FGFR1 D2D3$_{141\text{-}365}$ and FGF1$_{17\text{-}144}$ were prepared by size exclusion chromatography in SEC buffer (150 mM NaCl, 10 mM Hepes pH 7.5), concentrated by ultrafiltration and mixed with heparinDP10 to form a ternary FGFR1-heparin-FGF1 signalling complex with a stoichiometry of 2:1.1:2 (the procedure of complex preparation is described in **Section 4.5.2**) and a concentration of about 10 mg/ml. Half of the prepared complex was subjected to enzymatic deglycosylation with EndoF1. Previous deglycosylation tests have shown that the FGFR1 D2D3 is protected from precipitation within the signalling complex. Only minimal protein precipitation was observed even after 30 minutes of enzymatic deglycosylation (2 μl EndoF1 on 100 μl) at 37 ˚C. Ultracentrifugation was used to remove precipitation. Glycosylated and deglycosylated protein pools were screened with the pre crystallisation tests (Hampton Research) and diluted to the identified optimal concentrations of 4 mg/ml and 6 mg/ml respectively. These two complex samples were subjected to sitting drop vapour diffusion crystallisation experiments for initial crystallisation screening (set up as described in **Section 2.6.1**). Just like the various FGFR4 constructs, FGFR1 was screened against 12 commercially available crystallisation screens (see **Table 4**) at 4 ˚C and 20.5 ˚C in two different sets for glycosylated and deglycosylated complex sample. None of the crystallisation trials set up with glycosylated sample yielded crystals. However, the deglycosylated complex sample crystallised in a range of rather unsimilar conditions and in generally anisotropic shapes. A selection of the best diffracting crystals obtained from initial screening is displayed in **Figure 5.7**. Most crystals grew at 20.5 ˚C in the same space group C2 and had the same unit cell dimensions (crystals numbered 1-5). All presented crystals grown at 20.5 ˚C diffracted X-rays to resolutions between 2.2-3 Å. Interestingly, the only diffracting crystal obtained at 4 ˚C (crystal number 6) grew in a space group $P2_12_12_1$ but did diffract X-rays only to 3.6 Å resolution.

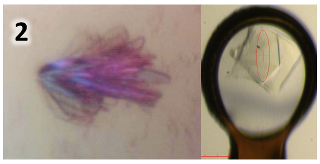
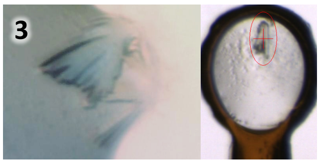
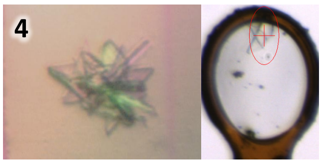
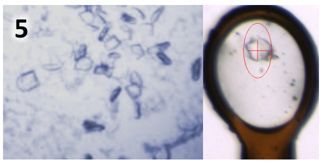
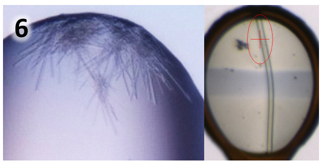
Crystal image	Size longest dimension (µm)	Crystallisation conditions	Space group	Resolution (Å)
20.5 ˚C				
	85	0.1 M carboxylic acids 0.1M Imidazole/MES pH 6.5 37.5 % MPD, PEG 1000, PEG 3350	C2	2.6
	100	0.1 M NaCl 0.100 M bis-Tris propane pH 9.0 25 % PEG 1500	C2	**2.2**
	70	0.12 M alcohols 0.1 M imidazole/MES pH 6.5 30.0 % PEG monomethyl ether 550, PEG 20k	C2	3.0
	45	0.1 M carboxulic acids 0.1M imidazole/MES pH 7.5 37.5 % MPD, PEG 1000, PEG 3350	C2	3.0
	50	0.1 M amino acids 0.1M imidazole/MES pH 6.5 30 % PEG monomethyl ether 550, PEG 20k	Indexing failed	~ 4.7
4 ˚C				
	200	4.000 M Sodium Chloride 0.100 M HEPES pH 7.0	$P2_12_12_1$	3.6

Figure 5.7 Initial crystallisation hits for FGFR1-heparinDP10-FGF1

The best diffracting crystals obtained from initial crystallisation screening of the FGFR1-heparinDP10-FGF1 complex. Crystals 1-5 grew at 20.5 ˚C and crystal 6 grew at 4 ˚C. The first image of each crystal to the left shows the crystal in the 96-well plate before manipulation. The second image of each crystal was taken at the beamline before data collection. Size of the crystal for data collection, crystallisation conditions, preliminary identified space group and the maximum resolution are listed for every crystal from left to right.

5.5.2 Crystal optimisation and structure determination

All initial hit conditions, including crystals with poor X-ray diffraction properties such as crystal ***5*** (**Figure 5.7**), were used as starting point for crystal optimisation experiments. Hampton additive screens optimisation as well as manual three-row optimisation (Walter et al., 2005) was applied. Optimisation could not be achieved for crystal number ***6*** (with space group $P2_12_12_1$).

Optimisation of crystals in space group C2 was more successful. Most conditions could be slightly improved. Interestingly, the best diffracting crystals were obtained from the initial crystallisation condition that yielded worst diffracting, non-indexable crystals (crystal ***5*** **Figure 5.7**). Three-row optimisation showed that crystals grow bigger and with sharp edges by changing the protein/reservoir ratio from 100nl+100nl to 200nl+100nl (see crystal ***7*** in **Figure 5.8 A**). The crystal grew to a maximum of 350 μm within 16 days. A second crystal was obtained by using the same protein/reservoir ratio of 200nl+100nl but with slightly diluted reservoir solution (1:25 water:reservoir). This crystal grew to a maximum dimension of 500 μm in 16 days (**Figure 5.8 D**). These were the only crystals that grew in the respective wells but they could be split into multiple smaller pieces for collection of many data sets. Crystals were cryoprotected in 25 % glycerol as described in **Section 2.6.3** and data sets of three single crystals were collected at 100 K on beamline I04 Diamond Light Source (DLS, Oxford, UK), which diffracted X-rays beyond 2 Å. A single sweep typically contained 1800 images taken at oscillations of 0.10 °, 70 % transmission, exposure of 0.100 s and using a beam size of 90x45 μm. The crystals and representative diffraction patterns are shown in the left and middle panel of **Figure 5.8 B, C and D** with zoomed-in views of the highest diffracting visible spots. Data collection statistics of datasets 1-3 are summarised in **Table 8**. Due to radiation damage, the data quality usually drastically decreased after about 600 recorded images. In order to generate one complete dataset with good statistics, these three datasets were merged with xia2 (Winter, 2010) using images 2-800 of dataset 1, images 2-450 of dataset 2 and images 2-800 of dataset 3. According to the cell dimensions, a 1:1:1 complex could be expected, but not a 2:1:2 complex. The crystal structure by Plotnikov et al. (2000) of a 1:1 FGF1 D2D3-FGF1 complex (pdb ID: 1EVT) was fragmented into its domains (D2, D3 and FGF1) and used as a high resolution search model for molecular replacement using Phaser (McCoy et al., 2007). Heparin was ignored during molecular replacement. Matthews coefficients (Matthews, 1968) confirmed that one FGFR1 D2D3 and one FGF1 molecule are present in the asymmetric unit.

Crystal	**Crystal 7 (dataset 1)**	**Crystal 7 (dataset 2)**	**Crystal 8 (dataset 3)**
Data collection			
Beamline	DLS I04	DLS I04	DLS I04
Resolution range (Å)	23.15 – 1.97 (2.02 – 1.97)	22.82 – 2.24 (2.30 – 2.24)	24.91 – 2.07 (2.12 – 2.07)
Spacegroup	C121	C121	C121
Cell dimensions: a, b, c (Å) α, β, γ (°)	72.86, 77.42, 81.11 90.00, 114.37, 90.00	72.96, 77.88, 81.39 90.00, 114.90, 90.00	73.13, 77.90, 81.88 90.00, 114.75, 90.00j
Wavelength (Å)	0.9795	0.9795	0.9795
Total number of reflections	63487 (4386)	60653 (4337)	76764 (5886)
Number of unique reflections	27174 (1970)	19192 (1400)	24671 (1828)
Completeness (%)	93.7 (94.0)	96.2 (95.2)	81.2 (85.5)
R_{merge} (%)	8.6 (93.6)	8.9 (95.8)	12.8 (84.1)
I/σI	7.7 (1.7)	9.6 (1.6)	8.5 (2.1)
Redundancy	2.3 (2.2)	3.2 (3.1)	3.1 (3.2)

Table 8 – Data collection statistics

Data collection statistics for three datasets collected from crystals FGFR1-heparinDP10-FGF1 crystals grown in similar conditions. Numbers in brackets refer to the high-resolution shell.

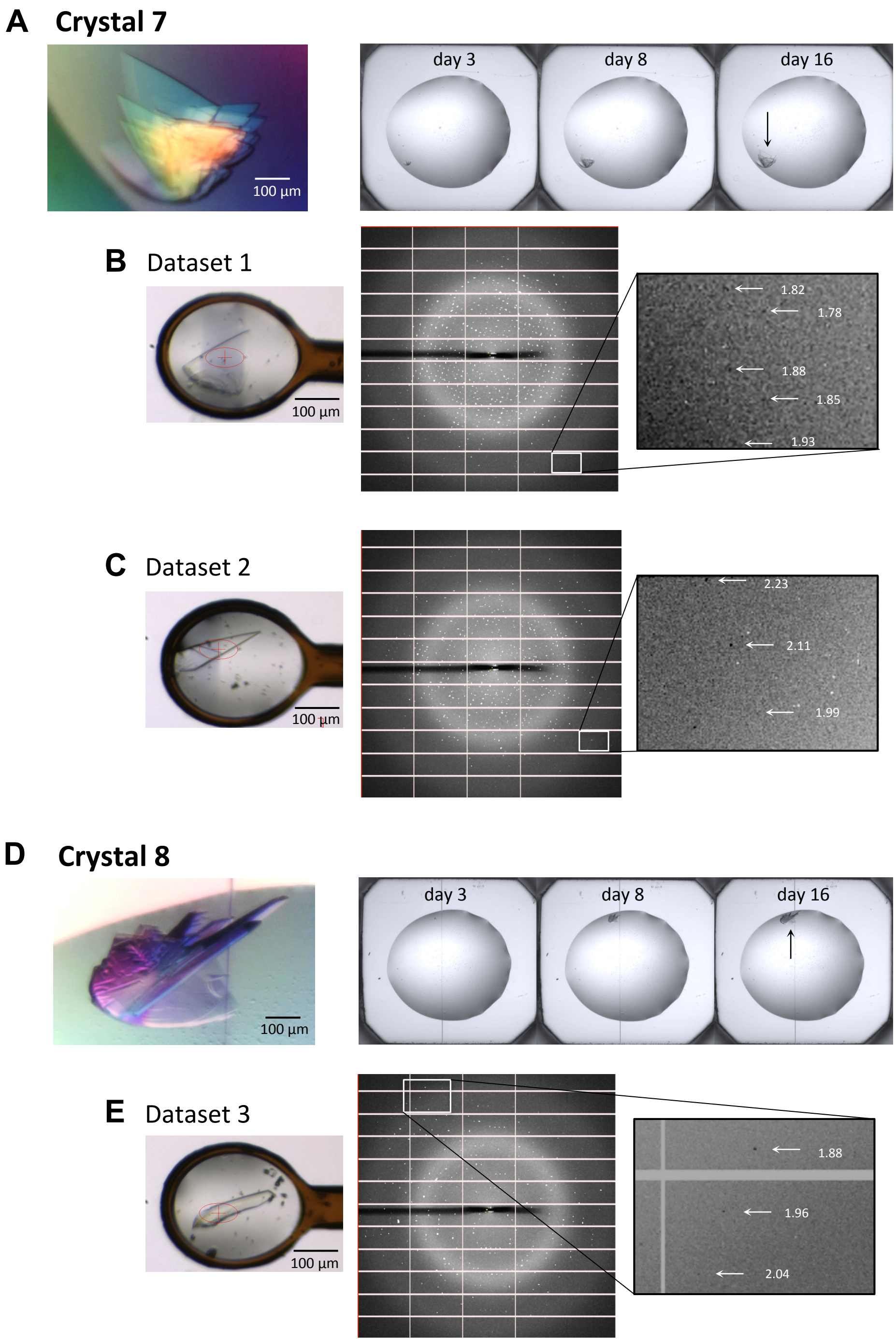


Figure 5.8 Diffraction of FGFR1-heparin-FGF1 crystalls

(**A**) Single crystal after three-raw optimisation of the Morpheus screen condition containing 0.1 M amino acids, 0.1 M imidazole/MES pH 6.5, 30 % PEG monomethyl ether 550/PEG 20k (initial crystals presented in **Figure 5.7** (**Crystal 5**). Images of the 200nl drop after 3, 8 and 16 days show the crystal growth. (**B**) and (**C**) High resolution diffraction data up to 1.76 Å were collected from two smaller pieces of the original crystal (***Dataset 1*** and ***Dataset2***). (**D**) A second chunk of multiple grown-together crystals from the same three-raw optimisation screen was split into small pieces (see crystal growth through the course of 16 days in three small drop images to the right). **E**) One of those pieces allowed collection of the best diffracting data set (***Dataset 3***) with diffraction out to 1.8 Å. Dataset 1-3 were merged to a single dataset.

5.5.3 Refinement, crystal packing and heparin solution

Initial rounds of refinement in Refmac (Murshudov et al., 1997) showed the same FGFR1-FGF1 interface as in previously published structures (Plotnikov et al., 2000; Schlessinger et al., 2000). Due to tight crystal packing (solvent content = 47.01 %) and good diffraction properties of the crystals, many regions of unoccupied positive density occurred which allowed building of flexible regions and of single amino acid side chains that were missing in the molecular replacement input structure by Plotnikov et al. After several rounds of iterative model building and refinement of the FGFR1 and FGF1 molecules, vacant electron density in both the 2Fo-Fc and the Fo-Fc maps occurred between symmetry related FGF1 molecules within the unit cell (**Figure 5.9 C**). These peaks of highest density were considered to be sulphate groups for several reasons; firstly, sulphur (when ordered) scatters X-rays more strongly than carbon or oxygen therefore gives higher signal in the maps, secondly, these peaks lie adjacent to multiple basic groups of Lys and Arg, which are likely to form electrostatic interactions with the sulphate groups of the glucosaminoglycan (GAG) ligand and lastly, those positive density peaks feature pyramidal shapes which is characteristic for SO_4 density (**Figure 5.9 B**). All 10 GAG units of heparinDP10 were modelled manually into this density. This protein was predicted to have 6 N-linked glycosylation sites and GlcNac residues could be modelled into additional positive electron density occurring adjacent to Asn227, Asn240, Asn264, Asn296, Asn330 (**Figure 5.6 B**). Finally 74 water molecules were added, through the final refinement iterations. No further ions could be identified in the electron density map. Model building and refinement was performed as described in **Section 2.6.6**. Temperature factors were refined using TLS group refinement combined with isotropic B factor refinement. Final structural validation was performed as described in **Section 2.6.7**.

Figure 5.9 A presents three different views of the C2 asymmetric unit, in the typical colour code (receptor in cyan, FGF1 in blue and heparin in red). The view in y-z-plane reveals direct heparin-receptor interactions and the view in x-z-plane reveals direct heparin-FGF1 interactions. However, the view in x-y-plane (left panel) shows that only half of the heparinDP10 chain interacts with proteins of the same asymmetric unit in the 1:1:1 complex. A search for biologically relevant higher order complexes revealed a 2:1:2 complex in potential P1 space group (**Figure 5.9 C**). In this complex a single heparin chain connects two receptor-ligand units. However, attempts to solve the structure in P1 space group resulted in worse defined heparin density and generally in worse refinement statistics. This must be due to the anisotropic character of heparin, which has two different options to occupy the heparin-binding site. The identified 2:1:2 complex only allows occupancy of the heparin-binding site with one heparin in one orientation. The best solution was obtained with a 1:1:1 complex in C2 space group with the heparin at half occupancy, allowing the formation of a potentially biologically relevant 2:1:2 complex, given rise by a 2-fold symmetry. The statistics of the finally refined structure with a resolution of 1.8 Å are presented in **Table 9**.

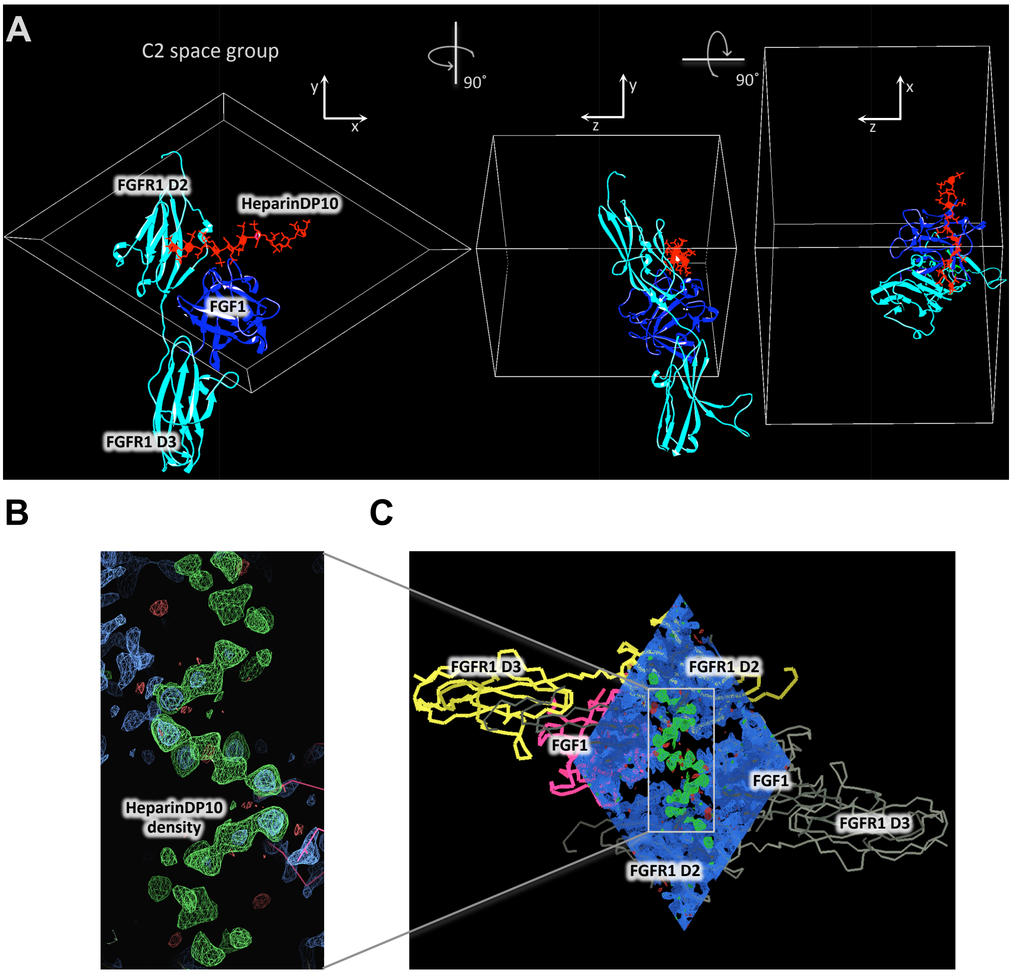


Figure 5.9 The monoclinic unit cell of the FGFR1-heparin-FGF1 complex

(**A**) Presentation of the FGFR1-heparin-FGF1 unit cell in all orientations. The asymmetric unit forms a 1:1:1 complex in C2 space group. (**B**) Heparin density is clearly defined and reaches over to a symmetry related receptor-ligand unit. (**C**) The two receptor-ligand units that are connected by the heparin density form the new, potentially biologically relevant 2:1:2 signalling complex architecture. As both 1:1:1 units add one heparin molecule, heparin had to be reduced to half occupancy to allow formation of a 2:1:2 complex.

Data collection	
Resolution range (Å)	38.80 – 1.80 (1.85 – 1.80)
Space group	C2
Cell dimensions: a, b, c (Å) α, β, γ (°)	72.82, 77.60, 81.16 90.00, 114.51, 90.00
Wavelength (Å)	0.9795
Total number of reflections	131464 (7556)
Number of unique reflections	35332 (2544)
Completeness (%)	99.1 (97.5)
R_{merge} (%)	5.2 (64.5)
I/σI	13.6 (1.3)
Redundancy	3.7 (3.0)
Refinement	
Resolution range (Å)	38.80 – 1.8
No. reflections	
R_{work} %	21.0
R_{free} %	25.9
No. atoms (non-hydrogen)	
Protein	3107
NAG	70
GAG	175
Water	74
B factor (Å^2)	
Protein	28.27
NAG	68.88
GAG	59.68
Water	42.42
RMSD bonds (Å)	0.0181
RMSD bond angles (°)	2.0952
Ramachandran statistics	
Favoured %	95.93
Disallowed	0
MolProbity score	2.31
MolProbity clash score	12.5

Table 9 – Data collection and refinement statistics

Data collection and structure refinement statistics for the FGFR1-heparinDP10-FGF1 1:1:1 complex. Numbers in brackets refer to the high resolution shell.

Figure 5.10 presents the crystal packing in all three orientations. The unit cell is highlighted with white lines and the putative biological unit is coloured according to the typical colour code. The top view in **Figure 5.10 C** looks directly onto the heparin chain that seems to hold both receptors and both ligands together. Interestingly, in that orientation, the crystal packing generates a 2-dimensional layer with all receptor N-termini pointing out of the image plane and all receptor C-termini pointing into the image plane. Although there are no strong crystal contacts between individual 2:1:2 complex units in that planar arrangement, this could indicate on a possible complex assembly for receptor clustering on the cell surface.

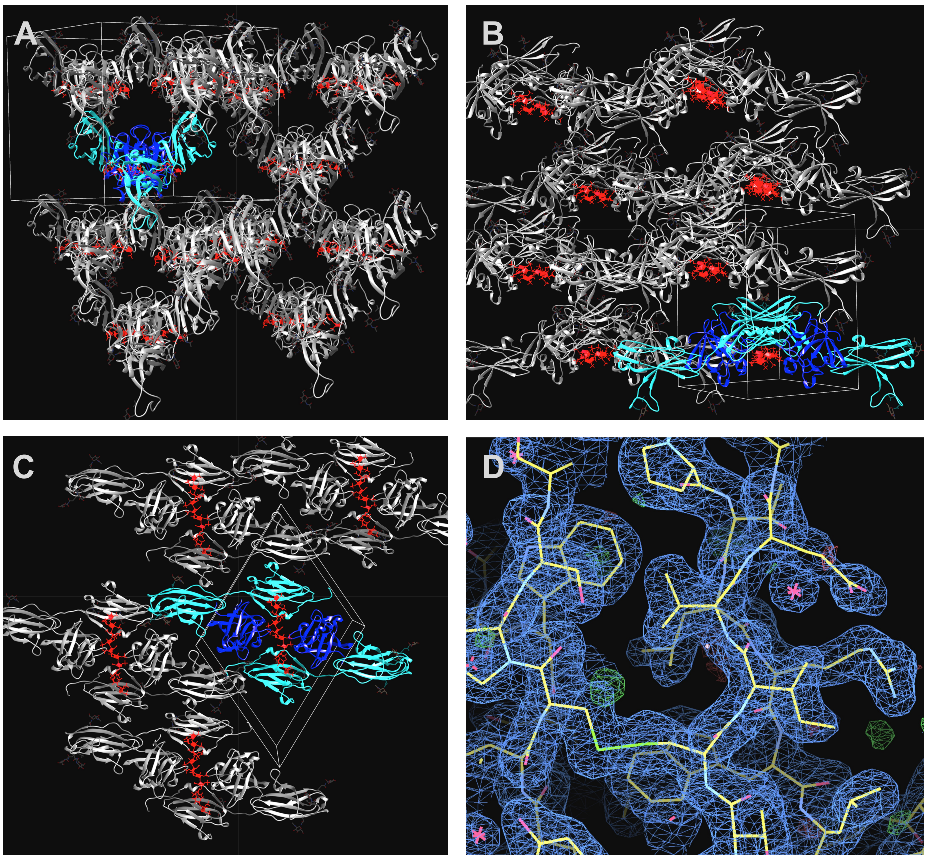


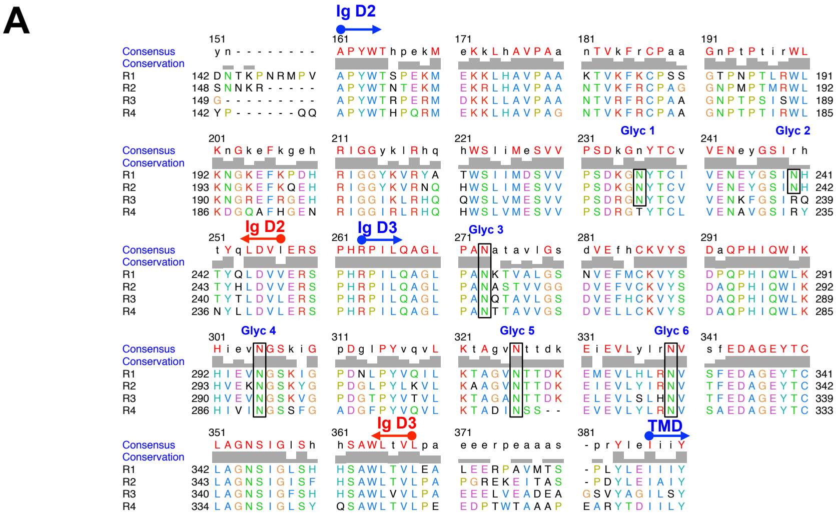
Figure 5.10 Crystal packing and final electron density map

(**A**) Packing of the FGFR1 D2D3-heparinDP10-FGF1 complex in the crystal. The asymmetric unit is comprised of a 1:1:1 complex and the putative biological signalling complex unit is comprised of a 2:1:2 complex. One 2:1:2 unit is depicted in the typical colours (receptor in cyan, FGF1 in blue and heparin in red). (**B**) 90° horizontal rotation of panel A. (**C**) 90° vertical rotation of panel B. (**D**) Electron density map of two β-sheets connected by a disulphide bridge at the end of the refinement.

5.6 N-linked glycosylation of FGFR1

5.6.1 Solving of N-linked glycans

Sequence alignment in **Figure 5.11 A** encompasses Ig domains D2 and D3 of FGFR1-4 and shows that not all N-linked glycosylation sites are conserved across all human homologs. FGFR1 and FGFR2 share number and position of all predicted glycosylation sites, including two additional predicted glycosylation sites on D1 (not shown in this alignment) and are thereby the family members with highest degree of glycosylation. As the FGFR1-heparin-FGF1 complex was crystallised in presence of EndoF1, only single GlcNAc residues were expected at each glycosylation site (see **Figure 5.11 C**). In fact, as the green boxes in **Figure 5.11 B** show, five out of the six GlcNAc residues could be resolved in the presented structure. This is the first crystal structure of a FGF receptor expressed in mammalian cells, showing glycosylations at Asn residues. Representatively, **Figure 5.11 D** shows the well-defined electron density of the N240 linked GlcNAc. Interestingly this GlcNAc residue formed crystal contacts with Ig D3 of a symmetry related receptor, showing that deglycosylation of all receptors at position N240 was essential for crystal formation. The only N-linked sugar that could not be found is located at the ßC'-ßE loop, the loop involved in receptor-ligand specificity (Beenken et al., 2012). The loop could be built completely, however, with relatively high B-factors assigned into residues N317-D320. For reason of completeness, the missing glycan at residue N317 is modelled to the structure in all following analysis.



B

homolog	N_{glyc} 1	N_{glyc} 2	N_{glyc} 3	N_{glyc} 4	N_{glyc} 5	N_{glyc} 6
FGFR1	N227	N240	N264	N296	N317	N330
FGFR2	N228	N241	N265	N297	N318	N331
FGFR3	N225		N265	N294	N315	N328
FGFR4			N258	N290	N311	N322

Ig D2 Ig D3

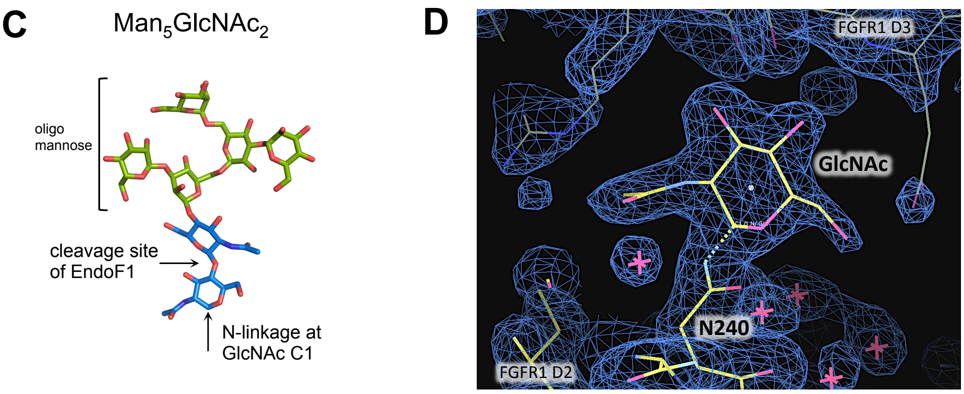


Figure 5.11 Comparison of FGFR D2D3 glycosylation patterns

(**A**) Sequence alignment of human FGFR family members. Predicted glycosylation sites are highlighted with blue label and black boxes. Start and end of domains are indicated with blue and red arrows, respectively. (**B**) Summary of all glycosylation sites with corresponding residue number in each homolog. FGFR1 glycosylation sites in green boxes could be resolved whereas the one in a red box could not be found in the electron density of the presented structure. (**C**) Structure of the oligomannose $Man_5GlcNAc_2$ (mannose in green, GlcNAc in blue) which represents the final N-linked sugar species for protein expressed in HEK293S cells. The position of Asn-linkage and the cleavage site of EndoF1 are highlighted with arrows. EndoF1 treatment leaves a single GlcNAc residue at each glycosylation site. (**D**) Representative GlcNAc residue linked to Asn240, with corresponding electron density of the presented structure.

5.6.2 Localisation of FGFR1 N-linked glycosylation in signalling complexes

In **Section 4.5.1** the potential involvement of N-linked glycans in regulation of the FGFR-FGF dimer formation was discussed. MALLS experiments with deglycosylated FGFR1 sample (**Section 5.4.2**), suggested that N-linked glycans do not have a direct influence on receptor-ligand interaction. These experiments were not capable of monitoring possible interference of N-linked glycosylations with formation of higher order signalling complexes. Various studies have shown that some activating mutations of human FGF receptors 1–3 abolish asparagine residues in consensus N-glycosylation sites, which lead to a diverse group of skeletal disorders. Asparagine 330 to isoleucine (N330I) in FGFR1 causes Osteoglophonic Dysplasia (K. E. White et al., 2005), asparagine 331 to isoleucine (N331I) in FGFR2 leads to Crouzon Syndrome (Steinberger, Mulliken, & Müller, 1996) and substitution of asparagine 328 with isoleucine (N328I) in FGFR3 causes Hypochondroplasia (Winterpacht et al., 2000). These three mutated asparagines correspond to $\boldsymbol{N_{glyc}}$ **6** in **Figure 5.11** and are labelled with red ***Glyc 6*** in **Figures 5.12** and **5.13**. A comprehensive genetic study in C. *elegans* has shown that this *N*-glycan negatively regulates FGFR activity *in vivo* (Polanska, Duchesne, Harries, Fernig, & Kinnunen, 2009), supporting the notion that mutation of N-glycosylation sites in human FGFR may lead to aberrant activation of the receptor.

Here, the aim is to study the localisation of N-linked glycans in the context of the three currently available signalling complexes. As the Schlessinger and Pellegrini structures were obtained from different receptor-ligand combinations, and in addition miss important loops that comprise N-glycosylation sites, some homology modelling was required. The GlcNAc-glycosylated FGFR1-FGF1 unit of the new structure was used to generate models representing the Schlessinger and Pellegrini complex architecture. This resulted in signalling complexes built of identical components (not identical heparin stoichiometry) in three different architectures. The 12 single GlcNAc residues present in each of these models were extended to $Man_5GlcNAc_2$ (using the AllosMod server, (Guttman et al., 2013), which represents the full oligosaccharide chain attached to proteins expressed in HEK293S cells.

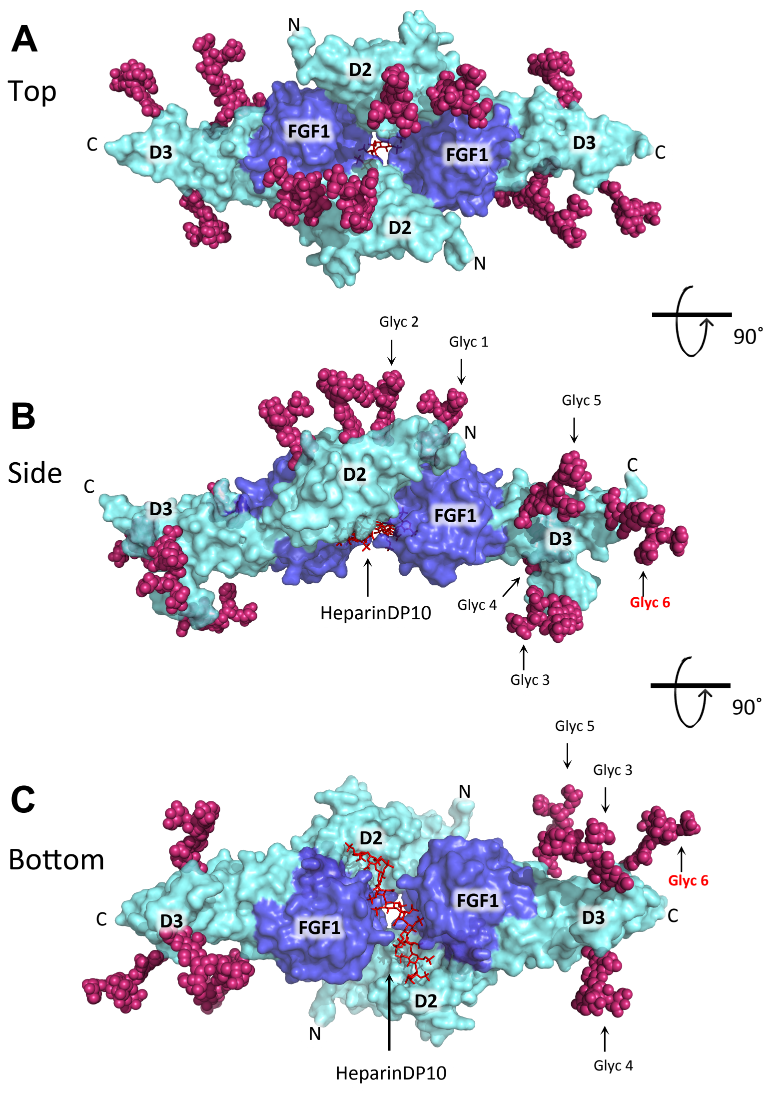


Figure 5.12 The new FGFR 2:1:2 signalling complex in presence of N-linked glycans

(**A**) Top view of the FGFR1-heparinDP10-FGF1 2:1:2 signalling complex. Single GlcNAc moieties that were found attached to asparagine residues in the crystal structure were extended to more bulky $Man_5GlcNAc_2$ saccharides (coloured warmpink). FGF receptors are coloured cyan, FGF1 molecules are coloured blue and the heparin chain is coloured red. (**B**) 90˚ horizontal rotation of panel A. N-linked glycans are labeled ***Glyc 1*** to ***Glyc 6***. ***Glyc 6***, which has been found to cause skeletal disorder upon deletion, is labelled red (see **Figure 5.11.B** for corresponding Asn positions) (**C**) 90˚ horizontal rotation of panel B.

Top-, side- and bottom-views of the new putative signalling complex models are presented in **Figure 5.12 A to C**. Interestingly, the side view shows that all four glycans linked to the D2 domains (***Glyc 1*** and ***Glyc 2***) are located on the exact opposite side to the heparin binding site, pointing into the same direction. This position seems to be the most probable potential orientation relative to the cellular membrane. All three views show that the N-linked glycans heavily hinder accessibility of certain receptor sites for potential binding partners. However, direct sterical hindrances interfering with formation of the 2:1:2 signalling complex cannot be found. Extension of all N-linked glycans to saccharides with more biological length would be fully compatible with the new complex architecture.

The front-view of the FGFR1 D2D3$_{glyc}$-heparin-FGF1 2:2:2 homology model in Schlessinger architecture (**Figure 5.13 B**) shows that most glycans point away from the prominent receptor-receptor interface suggested by (Plotnikov et al., 1999; Schlessinger et al., 2000; Stauber et al., 2000). However, the side view in **Figure 5.13 A** and the zoomed in subsection (**Figure 5.13 C**) indicate that the sugars corresponding to ***Glyc 3*** on Ig D3 (glycosylations at position N264) directly point towards each other. In fact, modelling of this glycan pair without spatial clashes required multiple attempts. Both neighbouring N264 residues and the first GlcNAc residues, which are part of the crystal structure, are shown as sticks. These two residues on each side point directly towards each other and imply that attached sugars are likely to generally point in the directions of the red arrows. The very well defined electron density for N264-GlcNAc confirms that there is no doubt about the correct conformation of the asparagine and the GlcNAc residues. In fact, N264-GlcNAc did not involve additional crystal contacts, which means that its orientation is a pure consequence of the saccharide stereochemistry. Duchesne et al. (2006) analysed the composition of N-glycans attached to FGFR1 (expressed in CHO cells) by mass spectroscopy. The most frequently detected oligosaccharide species were built of 13 sugar moieties and the largest were built of up to 18 sugar moieties spread in 4 branches. For the case that the Schlessinger architecture represents a biologically relevant complex, the presented modelling of

N-glycans suggests that the whole arrangement would require a huge degree of flexibility in order to allow accommodation of two bulky glycans between both Ig domains 3.

Analysing the same glycan modelling for the Pellegrini architecture, again, the only N-linked sugar that could potentially cause sterical clashes is $N264_{glyc}$. Due to the atypical positioning of Ig D3 (caused by a *cis*-proline in the D2-D3 linker), N264 is located directly adjacent to the FGF binding site (***Glyc 3*** in **Figure 5.13 E**). In this asymmetric model, heparin forms very different interactions to each receptor-ligand unit. The zoomed-in window presents ***Glyc 3*** on the opposite side after 180° rotation (**Figure 5.13 G**). This view suggests that the N264 linked glycosylation would theoretically hinder the continuation of the reducing end of the heparin chain. Whereas nonreducing ends are the natural ends of heparan sulphate chains, reducing ends do not exist on the cell surface as they are attached to the proteoglycan. Thus, the reducing heparin end would have to continue its path directly through the space that is occupied by the comparatively small, modelled N264-linked Glc_5NAc_2. This makes the formation of the Pellegrini signalling complex appear rather unlikely. To date there is no clear evidence whether the Schlessinger and Pellegrini signalling complexes are formed on the cell surface for cellular activation. The location of ***Glyc 6*** (that has been linked to skeletal disorders) does not seem to be able to interfere with signalling complex formation of any type. However, as it is positioned at the very end of D3 it is possible that its regulatory effect relies on interactions with the cellular membrane. Another possible regulatory mechanism could be that ***Glyc 6*** inhibits formation of clusters of activated receptors. Further cellular studies will be required to decipher the role of individual, highly conserved N-glycosylations.

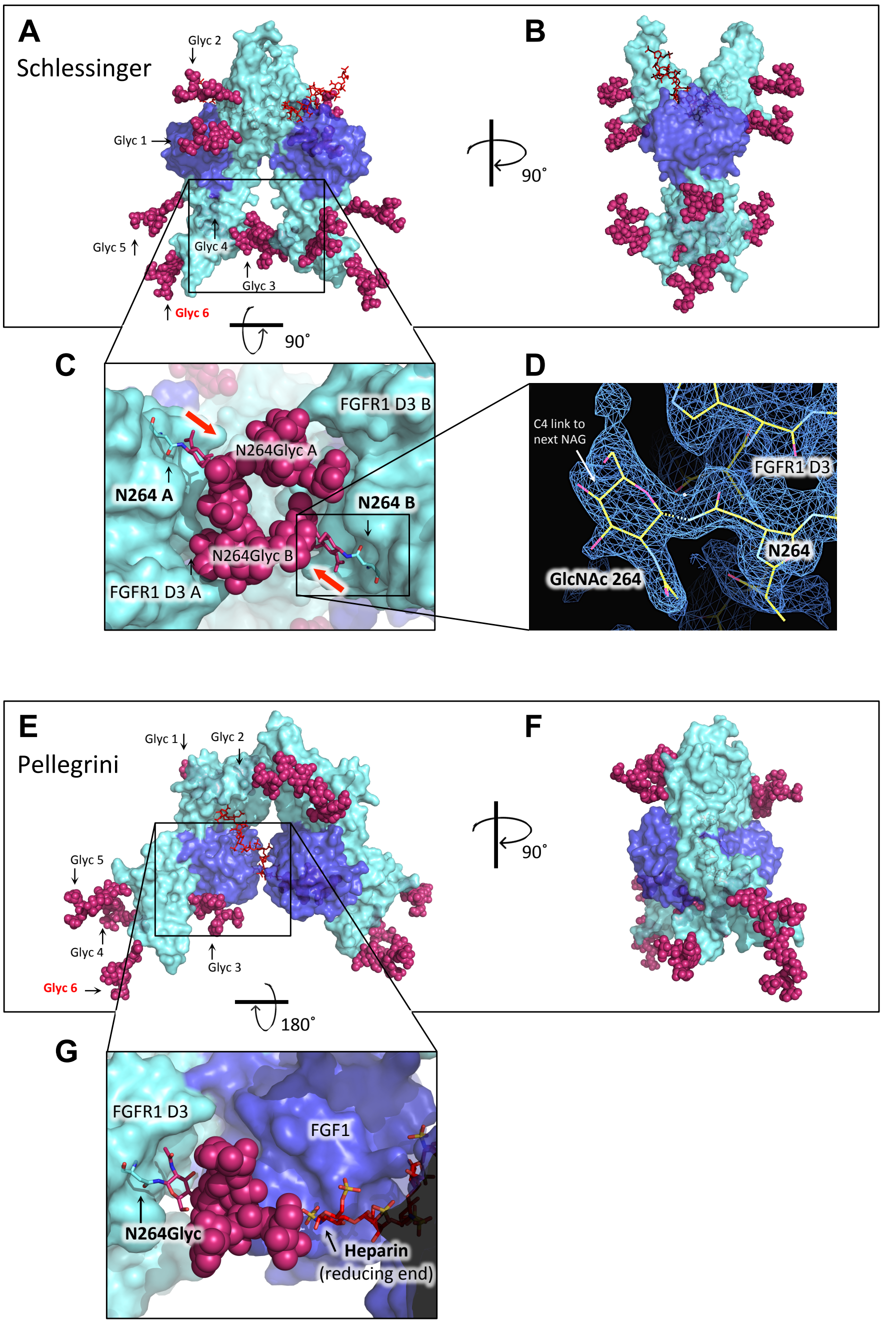


Figure 5.13 The Schlessinger and Pellegrini structures in presence of N-linked glycans

The FGFR1$_{glyc}$-FGF1 unit of the new structure was used to generate models in the Schlessinger and Pellegrini architectures. All GlcNAc residues were extended to Man$_5$GlcNAc$_2$ (using the AllosMod server, **Guttman et al., 2013**), which represents the full oligosaccharide chain attached to proteins that are expressed in HEK293S cells. Colour coding and labels as in **Figure 5.12**. (**A**) Front view on the FGFR1-FGF1 Schlessinger model. (**B**) 90˚ vertical rotation of panel A. (**C**) Zoomed in view of highlighted window in panel A after 90˚ horizontal rotation. Labels highlight glycosylated asparagine residues 264 from receptor A and receptor B. Red arrows indicate that both the rigid Asn side chains as well as the first GlcNAc residues point directly towards each other, indicating an approximate positioning of the more bulky glycosylation in biology. (**D**) Sticks model of the glycosylated FGFR1 Asn264 from the new structure with corresponding 2Fc-2Fo electron density map (blue) at 1 σ. (**E**) and (**F**) Same modelling as in A and B for the Pellegrini architecture. **G**) Zoomed in view of highlighted window in panel E) after 180˚ horizontal rotation.

5.7 The role of heparin in the FGFR1 signalling complex formation

5.7.1 Nonreducing heparin ends form same interactions to receptor-ligand Interface in all three models

After comparison of the N-glycan locations in the preceding section, the following section deals with the heparin-binding mode in all available signalling complex structures. **Figure 5.14 A** shows the asymmetric FGFR1-heparinDP10-FGF1 unit with the helical heparin axis in the image plane (upper panel) and with the heparin axis perpendicular to the image plane. (lower panel). The equivalent views are presented for the 1:1:1 receptor-heparin-ligand unit of the Schlessinger (**Figure 5.14 B**) and Pellegrini architectures (**Figure 5.14 C**) (as both structures are not symmetric, the complex unit with most HS interactions was chosen). Interestingly, although the three signalling complex architectures differ strongly, all FGFR D2-heparin-FGF units adopt the same arrangement (with exception of the second receptor-heparin-ligand unit of the asymmetric Pellegrini model, not shown). The βC'-βE loop on Ig domain D3 (highlighted for all structures in **Figure 5.14**) has been established to be responsible for ligand binding specificity by direct interactions with FGF (Beenken et al., 2012; Olsen et al., 2004). Importantly, these interactions are observed in the new structure. In contrast, structural and functional studies have shown that the Ig D3 orientation in the Pellegrini structure, imposed by the *cis*-conformation of P235, is not compatible with the ligand specificity filter as the βC'-βE loop can not form interactions with the ligand (Ibrahimi et al., 2005). The large insertion of 14 residues, forming the extended βC-βC' loop in D3, is fully ordered in the new model. In the Schlessinger structure this loop is ordered in only one of both D2D3 molecules in the asymmetric unit and it is completely disordered in the Pellegrini structure (see arrows indicating the βC-βC' loop). This loop comprises a highly conserved glycosylation site but its function remains elusive.

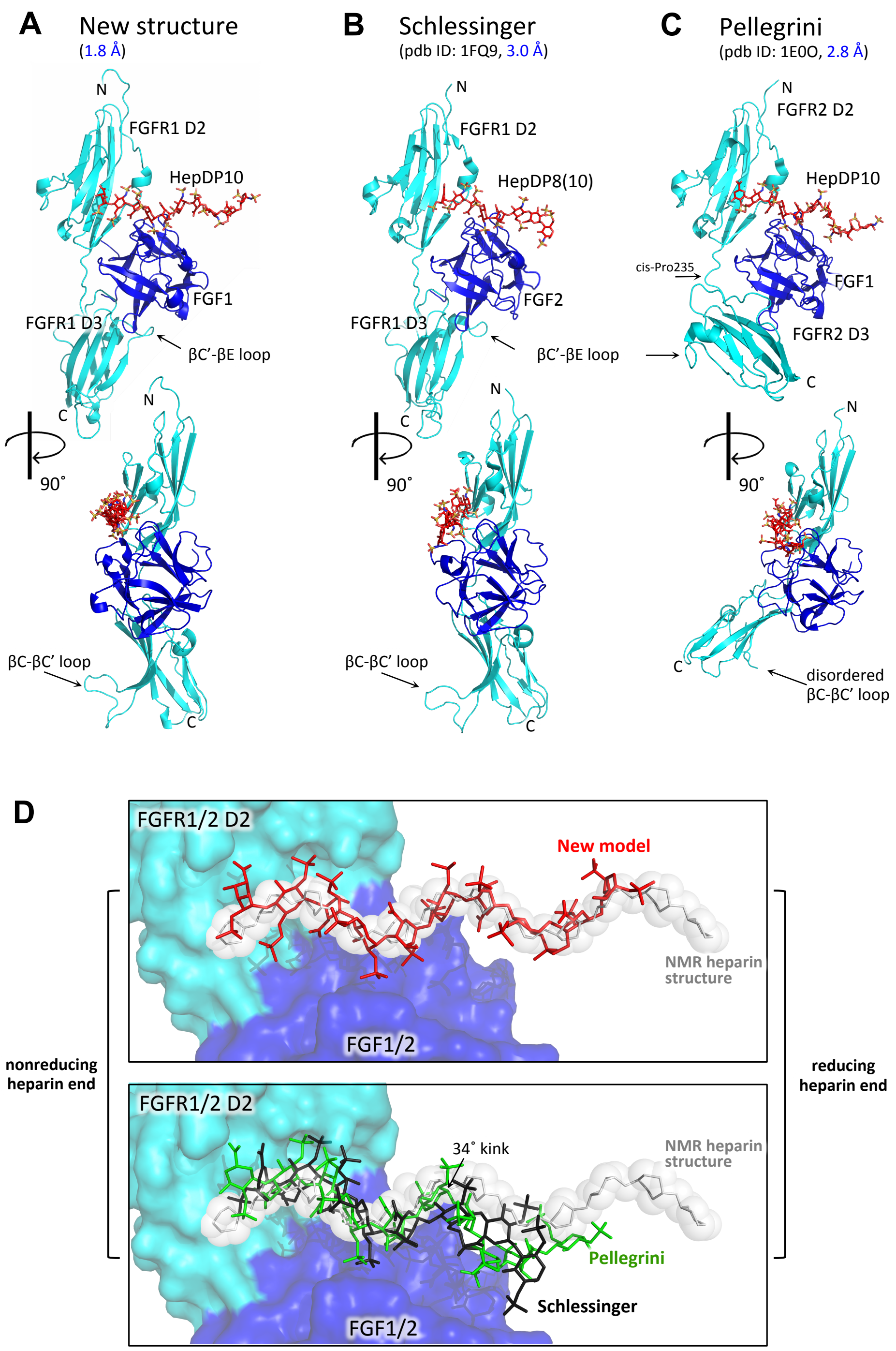


Figure 5.14 – Heparin binding mode in different FGFR-heparin-FGF 1:1:1 units

The single 1:1:1 units of 2:1:2 (or 2:2:2) signalling complexes are presented in **A**, **B** and **C** from the same front view (upper panel) and side view (lower panel) in cartoon representation. FGFR molecules are coloured in cyan, FGF molecules in blue and the sticks models of heparin chains are coloured in red with additional atom specific colouring. The βC'-βE loop on Ig domain D3 is labelled for all three structures. (**A**) The 1.8 Å FGFR1-heparinDP10-FGF1 structure resolved all 10 saccharides with half occupancy in each orientation. Only one of both possible heparin orientations is shown. 90° horizontal rotation shows the linear helical axis of the heparin molecule and reveals the fully resolved βC-βC'. (**B**) Only 8 out of 10 heparin saccharide units are ordered in the Schlessinger structure with 3.0 Å resolution (pdb ID: 1FQ9). The side view in the lower panel shows that the helical heparin axis is slightly bent. (**C**) In the Pellegrini structure with 2.8 Å resolution (pdb ID: 1FQ9), the proline 235 adopts a *cis*-conformation, which causes a rotation of the whole Ig D3 (compare new position of βC'-βE loop). All heparin units are ordered. The βC-βC' is disordered. (**D**) Comparison of orientation and position of the heparin chains in all signalling complex structures. In order to represent the heparin binding cleft the FGFR1-FGF1 dimer of the new structure is shown in surface representation in the background of both panels. In addition, the saccharide backbone of the heparinDP12 NMR structure (**Forster et al., 1993**; pdb ID: 1HPN) is shown in sticks and spheres representation (white), representing the ideal heparin conformation in solution. The top panel shows the alignment of the NMR structure to the heparinDP10 from the new model (red). The lower panel compares the helical axis of the NMR structure with one Schlessinger heparin (black) and the Pellegrini heparin (green). All heparin chains point their nonreducing end towards the receptor-ligand interface.

In the new structure and in the Pellegrini model all saccharide units form hydrogen bonds with receptor and/or ligand, which caused the complete decasaccharide to be ordered in the crystal. In contrast, in the Schlessinger structure only the first 6 saccharide units form interactions with receptor or ligand. Thus only 8 sugars could be resolved on the 1:1:1 unit presented in **Figure 5.14 B** (or 6 on the corresponding pseudo-symmetric unit, not shown). In solution heparin adopts a rigid helical conformation (Sommer & Rifkin, 1989). In **Figure 5.14 D** the sugar backbone of the linear NMR solution structure of heparinDP12 (B Mulloy, 1993); pdb ID: 1HPN; see white spheres) is manually fitted into the heparin binding site of a receptor-ligand dimer, enabling conformational comparison to the heparin observed in the new structure (upper panel) and to the heparin in the Schlessinger and Pellegrini structures (lower panel). All heparin molecules adopt the same orientation, in which the nonreducing end reaches into the same binding pocket of FGFR D2. The general spiral track of the first six saccharides is in good agreement across all three structures. Sulphate moieties of the different heparin chains are clustered at the same positions, where they interact with positively charged arginine or lysine residues on FGFR and FGF. Interestingly, only the conformation of heparinDP10 in the new model is in good agreement with the conformation adopted by heparin in solution (which corresponds to the sterically and energetically most favoured conformation). In contrast, the Pellegrini heparin chain adopts a prominent 34° distortion of the helical axis between the 6^{th} and the 7^{th} saccharide (see green heparin chain), which has been assumed to cause higher affinity for a second FGF molecule by realignment of free sulphate groups (Robinson et al., 2005). However there is no experimental evidence for such a mechanism, which moreover would involve an FGF dimer with untypical secondary FGF-heparin interface, as will be discussed in **Section 5.7.3**.

5.7.2 C2-symmetry axis in new model abolishes orientational restraints of heparin

In context of the 1:1:1 complex, heparin chains appear to have a determined orientation. However, in context of the full signalling complex, the new model reveals for the first time directional independency of the heparin chain. **Figure 5.15 A** illustrates that the 2:1:2 complex has a C2 symmetry axis in the centre of the heparin chain. Both perfectly symmetric 1:1:1 units contribute a heparin chain with half occupancy that are merged to a 2:1:2 complex with two heparin chains of half occupancy. The green arrows indicate that heparin is observed in both orientations in the crystal (with a 50 % probability for either orientations). In contrast the Schlessinger (2:2:2) and Pellegrini (2:1:2) complexes were determined as crystallographically asymmetric complexes. Schlessinger et al. solved their crystal structure in space group $P4_12_12_1$. Both heparin chains point their nonreducing ends (coloured red) directly towards the central pseudo C2-symmetry axis (see green arrows in **Figure 5.15 B**), which makes them directionally identical in each 1:1:1 complex subunit. Importantly, as the authors of this crystal structure report, the formation of this complex on the cell surface depends on two free nonreducing heparin ends. Besides their observation that terminal 2-O-sulfation (which is only present at the nonreducing end) is essential for complex formation, it is important to mention that free reducing heparin ends (coloured blue) do not exist on the cell surface. Thus the opposite orientation would not be compatible with the biochemistry at the cell surface (see red arrows in **Figure 5.15 B**). Pellegrini et al. observed a complete, asymmetric 2:1:2 complex ($P6_2$ unit cell) that implies that the assembly of a signalling complex takes place on a single heparin molecule. The green arrow in **Figure 5.15 C** indicates the observed orientation of the heparin chain that only interacts with one receptor. (Harmer et al.,

2004) analysed the interactions of the nonreducing heparin end to FGFR in detail and concluded that essential interactions for complex formation could theoretically not be undergone by the reducing heparin ends. Thus, it may be assume that on the cell surface this complex could not be formed with heparin in the inverted orientation (see red arrow in **Figure 5.15 C**). Furthermore, in biological context the reducing heparin end would have to be continued. However, this would be strongly impeded by the adjacent bulky N264-linked glycosylation, as discussed in **Section 5.6.2** (see **Figure 5.13 G**).

A structure based search for the minimal HS sequence required for FGFR and FGF binding was performed by screening of all available structures that involved a FGFR and/or FGF bound heparin (Pellegrini et al., 2001). The minimal binding sequence to FGF was convincingly determined as GlcN(6OSO3)–IdoA(2OSO3)–GlcN(NSO3). In the typical heparin conformation these three sulphates appear aligned on one side of the saccharide chain. The minimal binding sequence to FGFR involves both sulphates of a single glucosamine (GlcN(6OSO3) and GlcN(NSO3)). Making use of the PISA server (Krissinel & Henrick, 2007), here all three signalling complex structures are subjected to a similar but more detailed HS-protein interaction analysis in order to map which HS unit binds to which protein domain (**Figure 5.15 D-F**). The heparin chains for each complex are depicted as squares and numbered with ascending numbers starting from the reducing end. Heparan sulphates that form interactions with protein are represented by coloured circles on either cyan (FGFR interaction) or blue (FGF interaction) or. Strikingly, in the new model all 15 HS that are present on heparinDP10 form protein interaction (**Figure 5.15 D**). Both FGF1 molecules interact with the typical triad of O-6-, O-2- and N-sulphates. FGF1 B forms further interactions to two additional HS of the next triad and thereby reveals the most intense interaction to a heparin chain observed to date. Each receptor forms hydrogen bonds with 4 HS, involving one minimal FGFR binding sequence (GlcN(6OSO3) and GlcN(NSO3)). As the Schlessinger structure involves two heparinDP10 chains, theoretically there are 30 heparan sulphates available. However, protein interactions are only formed to saccharides

5-10 of each heparin decasaccharide. These 12 saccharides present a total of 18 HS units of which only 14 form protein interactions. FGFR1 A and B form remarkably ample interfaces to heparin but in very different patterns, revealing the high degree of asymmetry on the heparin level. Each receptor interacts with 6 HS by formation of at least one minimal receptor-heparin binding sequence. Both FGF2 molecules reveal identical interactions by binding to one typical HS triad. Of all analysed FGF-heparin interactions, FGF1 B in the Pellegrini structure is the only ligand that does not form the triad interactions to a 6-O-, 2-O-, N-sulphate motif. The strong 34° kink in the heparin chain (see **Section 5.7.1** and **5.7.3**) severely alters the ideal helical heparin conformation and thereby limits the interface to FGF1-B to two distant GlcN(NSO3) located at glucosamine 1 and glucosamine 5. In addition, out of all compared structures, the Pellegrini model reveals the weakest receptor-heparin interface. Whereas FGFR1 A only coordinates two HS (one minimal binding sequence), FGFR1 B does not bind to heparin at all. Only 8 out of 15 available heparan sulphates interact with protein, forming an arrangement that is going to be less stable and less dependent on specific sulphation patterns on cell surface HSPG than the new model.

Taking these observations together, it seems that the new presented structure represents the only signalling complex architecture that has no directional and positional restrictions for the heparin binding mode. All three structures comprise heparin decasaccharide, which has been identified in various studies to be a potent biological activator of FGFR (Franklin J Moy et al., 1997; Goodger et al., 2008; Ornitz et al., 1992; Wu et al., 2003). Strikingly, only in the new model does heparin adopt the ideal helical conformation and involve 100 % of all available heparan sulphates in protein interactions. This describes the most compact heparin-binding mode in a FGFR signalling complex and also suggests on high specificity for heparin binding arising from evolutionary development.

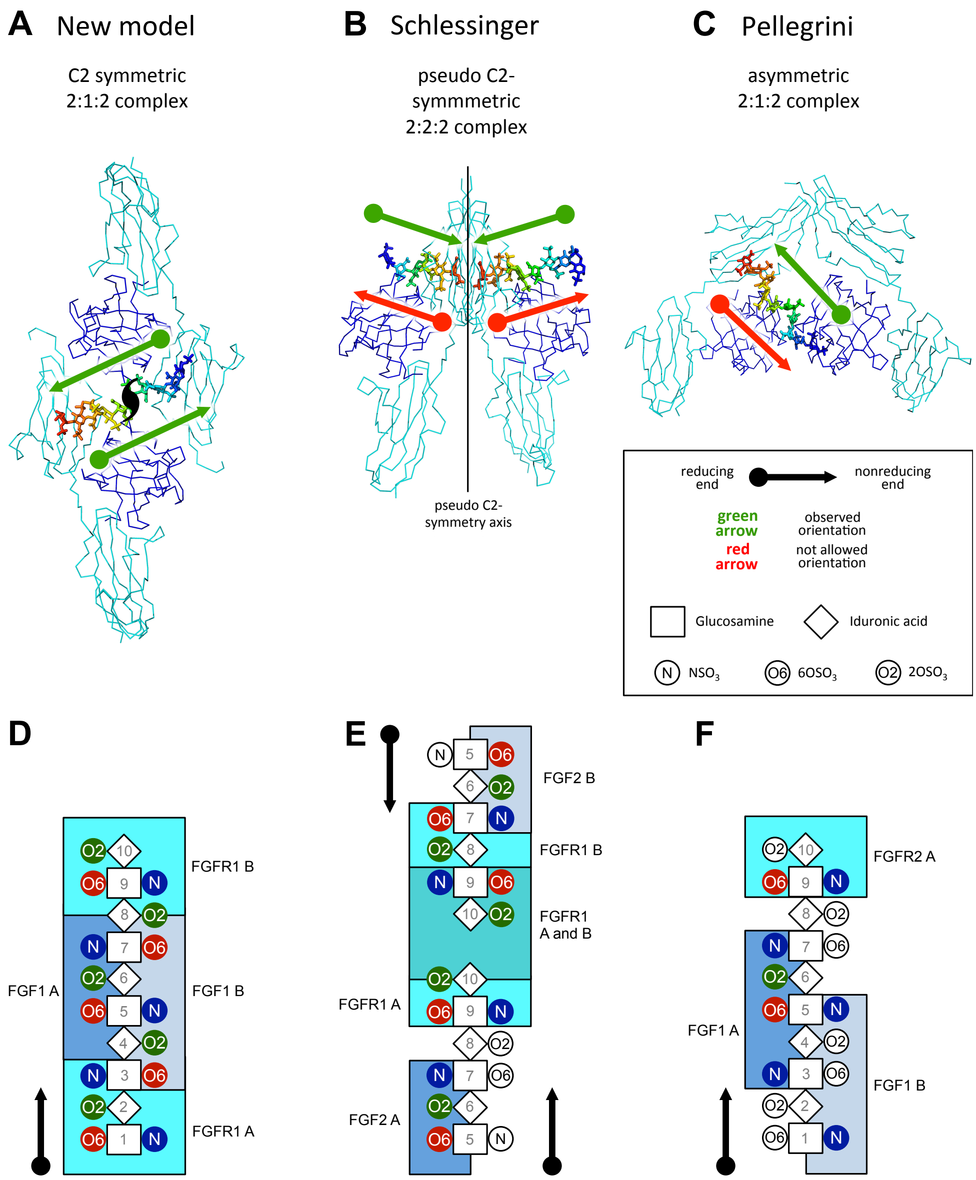


Figure 5.15 Heparin orientation and HS interaction map

Analysis of heparin orientation observed in (**A**) the new 2:1:2 model with C2 symmetry, (**B**) the pseudo C2-summetric 2:2:2 Schlessinger model and (**C**) the asymmetric 2:1:2 Pellegrini model. FGFRs and FGFs are shown as ribbons and coloured cyan or blue respectively. Heparin chains are presented as sticks models and are coloured in rainbow mode, showing the reducing end in blue and the nonreducing end in red. Green arrows indicate heparin orientations that were observed in the crystal structure and red arrows indicate heparin orientations that are not compatible with the respective crystal packing or with the biological environment. A detailed analysis of HS groups involved in FGF and FGFR binding is presented for the new model in (**D**), for the Schlessinger model in (**E**) and for the Pellegrini model in (**F**). Heparin molecules are represented as a chain of glucosamines (square) and iduronic acids (diamonds), numbered from the reducing to the nonreducing end and aligned vertically relative to each other in a way that reflects their structural alignment. In order to mimic the three-dimensional arrangement of the sulphate groups of adjacent disaccharide units relative to the heparin helix, they are drawn in opposite positions. N-sulphate groups at a protein-interface are shown in blue, 6-O-sulphate groups in red and 2-O-sulphate groups in green (representation adopted from Pellegrini et al., 2001). Cyan background labels HS that interact with FGFRs and blue background labels those HS that interact with FGFs. Black arrows indicate heparin orientation as in the upper panel.

5.7.3 Solely the new model is consistent with cooperative FGF dimerisation

The question of how two receptors and two ligands find each other to form a heparin mediated complex still needs to be addressed. Studies have suggested that heparin-linked FGF oligomers are the biologically active form of the ligand (DiGabriele et al., 1998; Franklin J Moy et al., 1997.; Spivak-Kroizman et al., 1994). The observation from SEC-MALLS experiments (presented in **Section 5.3.2**), suggesting that heparin mediated FGF1 dimerisation proceeds cooperatively, supports the hypothesis that FGF dimerisation precedes recruitment of FGFR to form a ternary signalling complex. This hypothesis is further supported by previous SEC studies (Harmer et al., 2006; Robinson et al., 2005) and an ITC study, reporting that the FGF2-octasaccharide interaction exhibits pronounced cooperativity (Brown et al., 2013). Several crystal structures have been described for FGF molecules in complexes with heparin-like sugars (Faham et al., 1996; Ornitz et al., 1995; X. Zhu, Hsu, & Rees, 1993), revealing the FGF-binding site for oligosaccharides. The first biologically active dimer of human FGF1 in complex with a fully sulphated, homogeneous heparin decasaccharide was solved by DiGabriele and coworkers (1998) at a resolution of 2.9 Å (**Figure 5.16 A**). Heparin decasaccharide cross links FGF1 in a trans configuration with two FGF1 polypeptides related by a quasi 2-fold symmetry axis that runs perpendicular to the helical axis of the heparin chain. There are no direct FGF-FGF contacts, which means that all contacts are mediated through the heparin molecule. Strikingly, aligning FGF1A of 1AXM to FGF A of all available FGFR signalling complexes (**Figure 5.16 B-D**) reveals that the new presented ternary complex is the only one that shares the same symmetry axis at identical position (compare 1AXM 2-fold symmetry axis to C2 symmetry axis depicted in **Figure 5.15 A**), including very similar heparin-FGF contacts. The heparin-ligand interface in DiGabriele's dimer, stretches over five saccharides (leaving five disordered saccharides), whereas heparin in the new structure is slightly reoriented by heparin-receptor interactions, enabling additional contacts to R122 and K128 of the ligand (residues flanking the heparin binding site on

FGF1), resulting in a total number of 6 saccharides that interact with FGF1. The asymmetric heparin-FGF1 unit in the Pellegrini structure (**Figure 5.16 C**) shows that the rotational position and the contacts of FGF1 B to heparin strongly deviate from the typical heparin-FGF interface (compare position of all β7-β8 loops in black circles). It is important to note that this asymmetric FGF dimer arrangement is not the result of two 'free ligands' binding to heparin, but rather imposed by the crystallographic arrangement of two FGFR-FGF units to one heparinDP10 molecule. In stark contrast, the Schlessinger structure is lacking a ligand dimer and is thereby not compatible with all studies suggesting that heparin-linked FGF oligomers are the biologically active form of the protein and the primary units for ternary signalling complex formation on the cell surface. However, as presented in **Figure 5.14**, the isolated heparin-ligand units are in perfect agreement with the interface observed by DiGabriele et al. Nevertheless, this model remains particularly difficult to reconcile with reported correlation between heparin length and their physiological activity (Bârzu et al., 1989; Ishihara, 1994; Pye et al., 1998).

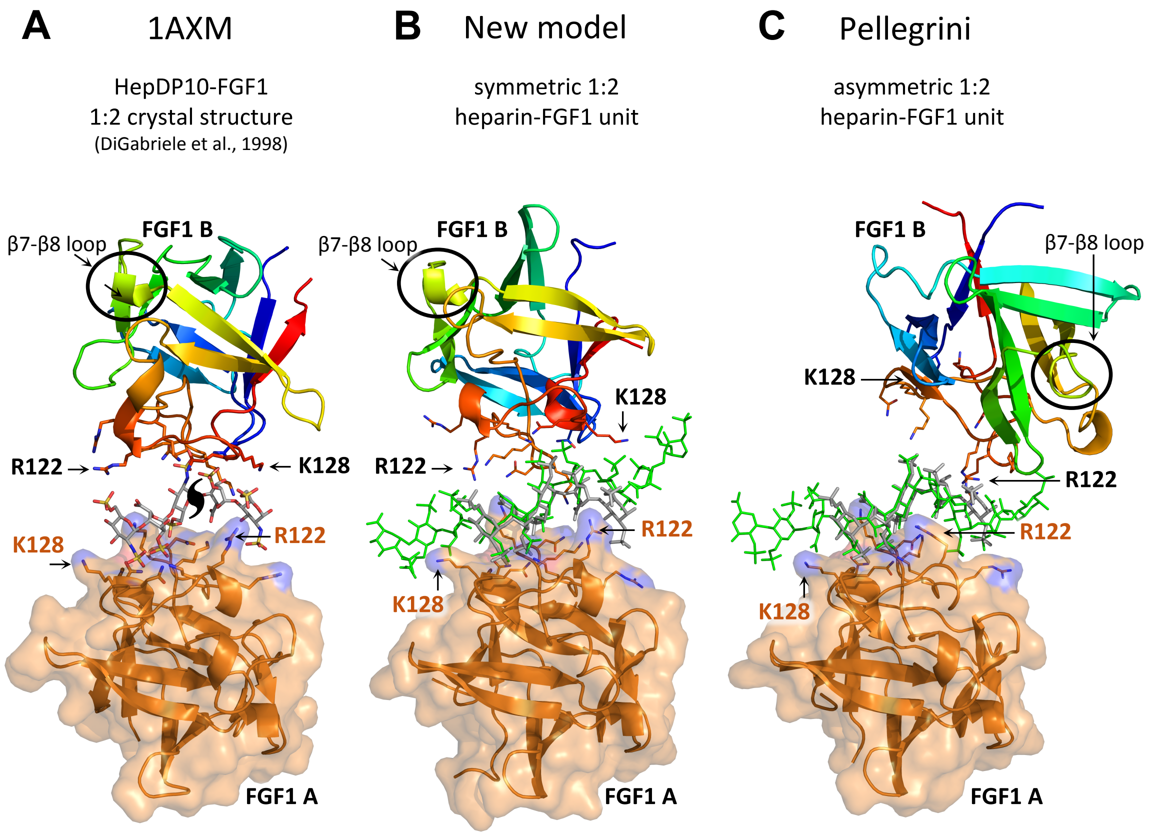


Figure 5.16 Comparison of heparin mediated FGF dimer structures

Comparison of the heparinDP10-FGF1 crystal structure (DiGabriele et al., 1998; pdb ID: 1AXM) in (**A**) to the heparin-FGF units of the new signalling complex presented in this work (**B**) and the asymmetric model by Pellegrini et al. in (**C**) (pdb ID: 1E0O). All structures were aligned to their FGF A molecules which are shown as cartoon in an orange transparent sphere. FGF B of all structures is shown as cartoon in rainbow representation. Selected arginine and lysine residues involved in heparin interaction are shown as sticks, coloured by element. R122 and K128 form the first and the last residue of the heparin binding site and serve as marker for positional comparison of all three structures. Equally the β7-β8 loop serves as a marker, which highlights the positional deviation of FGF B in the Pellegrini structure.

5.8 Conclusions and future Perspectives

An important objective pursued through comparison of FGFR signalling complexes is to find indications that help to explain the order of events during complex formation on the cell surface. The field has been debating whether receptor activation occurs upon a preformed FGF dimer (A. Brown et al., 2013; Robinson et al., 2005), by assembly of two preformed 1:1:1 receptor-heparin-ligand complexes (Mohammadi et al., 2005; Sterner et al., 2014) or whether both mechanisms are relevant (Goodger et al., 2008). **Table 10** summarises the crystallisation conditions of FGFR crystal structures representing putative biologically relevant higher order arrangements. Strikingly, with exception of the new structure, all crystallisation conditions contain high (≥ 1M) concentration of a sulphate salt. The last column of the table lists the molecules that are bound to the heparin-binding pocket created by assembly of two receptors and two ligands. The new model was crystallised from a mixture of all three complex components in defined concentration, allowing free assembly of any complex arrangement in presence of one GlcNAc residue at each N-linked glycosylation site. Crystals grew in conditions that out of all signalling complex conditions were closest to physiological conditions. It seems natural that the correct complex arrangement is mainly going to be driven by electrostatic interactions of the heavily negatively charged heparin to very positively charged lysine and arginine patches on the receptor and the ligand. As discussed in **Section 5.7.2**, the result is a symmetric 2:1:2 complex arrangement, in which all heparan sulphate units are hydrogen bonded to protein. FGFR1-FGF1 interaction in absence of heparin could not be observed in this work, supposedly due to electrostatic repulsion and obstructing N-linked glycosylations. This strongly suggests that the signalling complex was formed by stepwise addition of two receptors to a central heparin-mediated FGF-dimer, which is consistent with our previous observations of cooperative FGF dimerisation (**Section 5.3.2**). Importantly, the described order of events avoids at all times formation of protein-protein interactions without compensation of positive charges by heparin and should be compatible with physiological conditions on the cell-surface.

In stark contrast, the Schlessinger structure (see 1FQ9 in **Table 10**) was obtained by soaking of heparin to a previously published 2:2 structure formed by two preformed receptor-ligand dimers in presence of 1.6 M sulphate concentration (see 1EVT by **Plotnikov et al., 2000** in **Table 5.3**). In 1EVT the heparin binding site is occupied with 4 sulphates that compensate for otherwise strong electrostatic repulsion of the tightly packed positively charged heparin binding patches on receptors and ligands. It is questionable whether this complex assembly would occur in absence of sulphate under physiological conditions. As heparin is required for FGFR activation, it seems plausible that the Schlessinger architecture on the cell surface would be formed of two pre-formed 1:1:1 complexes. Thus, FGFR activation through a 2:2:2 complex would depend on the concentration of 1:1:1 complexes on the cell surface, and on receptor-receptor and secondary receptor-ligand interactions, not on heparin directly crosslinking two receptor-ligand complexes. The fact that nonreducing heparin ends are essential for the formation of this arrangement must be a major limiting factor for efficient FGFR signalling. In contrast the new presented signalling structure is independent from heparin orientation and location (see **Section 5.7.2**). Experiments with synthetic heparan sulphate blocks have shown that nonreducing heparin ends are stronger FGFR activators than highly sulphated domains anywhere else on the polysaccharide chain, which has been used as a supporting argument for the relevance of the Schlessinger model (Sterner et al., 2014). In fact, considering the very similar interaction pattern of heparin to the primary receptor-ligand dimer in the Schlessinger model and the new model (**Figure 5.14**), nonreducing ends might also be the stronger activators of signalling in the arrangement of the new structure, however not essential for activation. The analysis of further FGFR-FGF crystal structures without heparin, published by co-workers of Schlessinger, reveals that the Schlessinger arrangement with canyon-forming interaction between the FGFR ectodomains is not observed when crystallisation is achieved in a low salt buffer (compare pdb entry 1CVS with entries 1EVT and 1EV2). Another independent group has published the first heparin-free canyon-forming crystal structure (see 1DJS in Table 5.3; Stauber et al., 2000) that contains 7 sulphates in the heparin-binding site. Taking these facts together, it is likely that the crystallisation process, and not heparin, is responsible for the association of two FGFR molecules in a dimer with a canyon-like cleft.

Structure	**Crystallisation condition**		**Molecules bound to heparin binding site**
	Compound	**Concentration (pH)**	
New structure **FGFR1$_{glyc}$-HepDP10-FGF1** (2:1:2)	amino acids Imidazole/MES PEG monomethyl ether 550/PEG 20.000	0.1 M 0.1 M (6.5) 30 %	heparin
1EVT (Plotnikov et al., 2000) **FGFR1-FGF2** (2:2) Schlessinger arrangement	**$(NH_4)_2SO_4$** Tris-HCL Glycerol	**1.6 M** 0.1 M (8.5) 20 %	4 sulphates
1FQ9 (Schlessinger et al., 2000) **FGFR1-hepDP10-FGF2** (2:2:2)	**$(NH_4)_2SO_4$** Tris-HCL Glycerol	as above + heparin (soak)	Heparin
1DJS (Stauber et al., 2000) **FGFR2-FGF1** (2:2) Schlessinger arrangement	**$(NH_4)_2SO_4$** Hepes	**1.6 M** 10 mM (7.5)	7 sulphates
1E0O (Pellegrini et al., 2000) **FGFR2-hepDP10-FGF1** (2:1:2)	**Li_2SO_4** Tris-HCL NiSO4	**1 M** 0.1 M (8.5) 10 mM	4 sulphates + heparin

Table 10 – Crystallisation conditions of FGFR complex structures

The single heparin chain in the asymmetric signalling complex (Pellegrini et al., 2000) crosslinks two receptor-ligand units via a central FGF dimer. Importantly, in this arrangement one FGFR molecule is attached to the complex just by interactions to a ligand, which leaves a complete heparin-binding site unoccupied (**Figure 5.17 C**). Just as in the heparin-free, canyon-forming structure (pdb ID: 1EVT), the electrostatic repulsion of the positively charged patches of receptor and ligand are compensated by 4 sulphates that originate from 1 M lithium sulphate in the crystallisation condition. Therefore, this arrangement (including the cis-Pro235 residue (**Section 5.7.1**)) is again likely to represent the result of the crystallisation process rather than a complex architecture that could occur under physiological conditions. Furthermore, as presented in **Section 5.6.2** (see **Figure 5.13 E** and **G**), this arrangement seems unlikely to be feasible in presence of N-linked glycosylations.

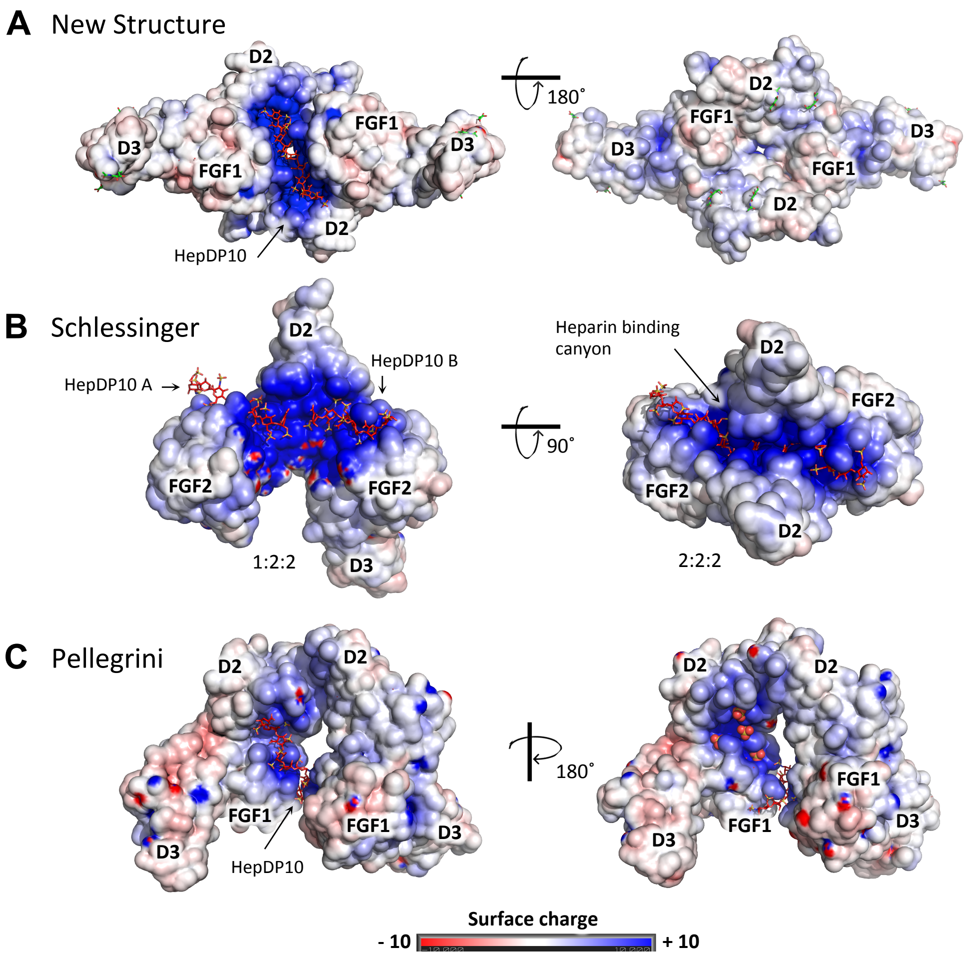


Figure 5.17 – Electrostatic potential of FGFR signalling complex structures

Electrostatic surface potential of all three signalling complex structures at a potential range from red (−10kbT/ec) to blue (+10kbT/ec). (**A**) The new structure depicted from top and bottom view. Heparin is shown in red. GlcNAc residues are shown as green sticks. (**B**) The left panel shows the Schlessinger structure after removal of one receptor, allowing a direct view into the inside of the heparin binding canyon occupied by heparinDP10 A and B. The right panel shows the top view onto the heparin canyon of the complete Schlessinger complex. (**C**) Front and rear side of the Pellegrini structure, revealing the phosphate occupancy of the heparin binding site of one receptor-ligand pair.

Considering protein production, complex formation protocol, crystallisation condition, crystal packing and symmetry, position of N-linked glycosylations on the receptor, compatibility with suggested receptor-ligand binding specificity filter, length, number, orientation, conformation and interaction pattern of heparin, ligand dimer arrangement and electrostatics of all three ternary signalling complexes, the Schlessinger and Pellegrini structures encounter multiple difficulties in explaining full functionality of activation of FGFR signalling pathways in physiological conditions. At the same time the new model presented in this chapter reveals full compatibility with physiological conditions and agreement with the reviewed studies. It is important to evaluate compatibility of all three models with mechanisms of kinase dimerisation for efficient downstream activation upon receptor dimerisation. In addition and importantly, to date neither the Schlessinger nor the Pellegrini architecture could be confirmed to be present in solution, questions which are the subject of the following chapter.

Chapter 6

Structural analysis of the FGFR signalling complex in solution

Chapter 6 – Structural analysis of the FGFR signalling complex in solution

6.1 Summary

All three FGFR signalling complex structures lack clear evidence for their formation in solution. Here, Small Angle X-ray Scattering (SAXS) is used in combination with atomistic molecular dynamics simulations in order to study the low-resolution solution structure of the FGFR1-heprin-FGF1 complex. The presented data strongly suggests that glycosylated receptor prefers signalling complex formation in the new elongated architecture. A MD simulation derived structural ensemble revealed breathing motions of the complex around the heparin axis and improved SAXS curve fitting to the new structure. Furthermore, heparinDP16 was observed to perform least complex dissociation in solution.

6.2 Introduction

With the addition of the new complex structure presented in **Chapter 5**, at present there are three crystal structures representing three different FGFR signalling complex architectures. Although these models have some important features in common, the bigger scale architectures and the possible mechanisms of activation that come along with them, differ strongly. Although comprehensive analysis by size exclusion chromatography (Goodger et al., 2008; Harmer et al., 2006; Robinson et al., 2005), analytical ultracentrifugation and mass spectrometry (Harmer et al., 2004; Ibrahimi et al., 2005) have provided valuable insights to the behaviour of the signalling complex in solution, in no case could the authors confirm unambiguously from which complex architecture the collected data were generated. There are various methods to be considered in order to obtain structural data of protein complexes without the bias of crystal packing in non-physiological conditions. My attempts to generate a 2D reconstruction by negative stain electron microscopy (based on methods described in Ohi, Li, Cheng, & Walz, 2004) failed due to small particle size and heterogeneity of the complex sample. I attempted solubilisation screenings of multiple full-length receptor constructs from expression in HEK293S cells and cell membrane extraction as applied to other membrane protein systems (Miller & Aricescu, 2014) but these efforts did not yield suitable amounts of soluble protein for complex formation and cryo-EM analysis. In attempts to express the protein natively anchored in cell-derived membrane vesicles (method described by Zeev-Ben-Mordehai et al., 2014) the receptor could not be detected in the collected vesicles by western blot analysis. SAXS is a well-established source for structural information of protein samples in solution and deemed as particularly helpful when conflicting high resolution crystal structures are available (reviewed by Schneidman-Duhovny et al., 2012). SAXS data of high quality could be obtained from signalling complex samples comprising different complex stoichiometries and varying heparin length. Based on preliminary fitting of measured SAXS curves of signalling complex samples to all three crystal structures (data not shown) and based on thorough comparison of these structures' conformity with published data (see **Sections 5.6.2 – 5.8**), further analysis will be limited to the 2:2:2 Schlessinger two-end model and the new elongated 2:1:2 structure.

6.3 Aims

The aims of the work presented in this chapter are to compare signalling complex crystal structures with structural data of complex samples collected by SAXS. This will provide important, currently lacking information on which complex stoichiometry and configuration is adopted in solution.

6.4 SAXS curves fitted to AllosMod structure ensembles

6.4.1 Modelling of flexible termini and $Man_5GlcNAc_2$ moieties

SAXS measurements are very sensitive to protein aggregation. As described in the previous chapter, FGFR1 precipitation could be observed after enzymatic deglycosylation. In order to preserve receptor stability and to keep the process of complex formation closest to biological conditions, the receptor has been used without EndoF1 treatment in all presented SAXS measurements. Thus, fitting of the experimental SAXS curves to the crystal structures required modelling of the 12 flexible, branched oligosaccharides and the 8 flexible protein termini that were disordered in the crystal structure but part of the protein constructs during SAXS measurements. All presented SAXS experiments were performed with the same FGFR1 and FGF1 constructs as used for crystallisation. The original Schlessinger structure contains the ligand homolog FGF2. Therefore, the glycosylated FGFR1-FGF1 unit of the new structure was used for creation of a homology model in 2:2:2 Schlessinger architecture (as described in **Section 5.6.2**; **Figure 5.13 A**). This homology model was required in order to ensure that SAXS curve fitting yields the most accurate solution in case of formation of the Schlessinger-like signalling complex architecture in solution.

Figure 6.1 describes the modelling process for addition of flexible termini and N-glycans to the signalling complexes using the AllosMod server (Guttman et al., 2013). To begin with, the heparin coordinates were removed from the refined crystal structure (AllosMod server does not cope with heteroatoms of that kind). An AllosMod integrated MODLLER (Eswar et al., 2007) function, created 50 models with differently oriented flexible termini. Subsequently, for each of these 50 models, 20 models were generated with $Man_5GlcNAc_2$ in variable orientations (on the basis of an input file indicating the correct type and connectivity of sugars). As this modelling process did not alter the crystal structure framework, the previously removed heparin coordinates could simply be appended to each final AllosMod output pdb. The same procedure was carried out for the FGFR1-FGF1 signalling complex structure in Schlessinger architecture. Finally, two separate ensembles were created, each comprising 1000 models with flexible termini and flexible N-linked glycans.

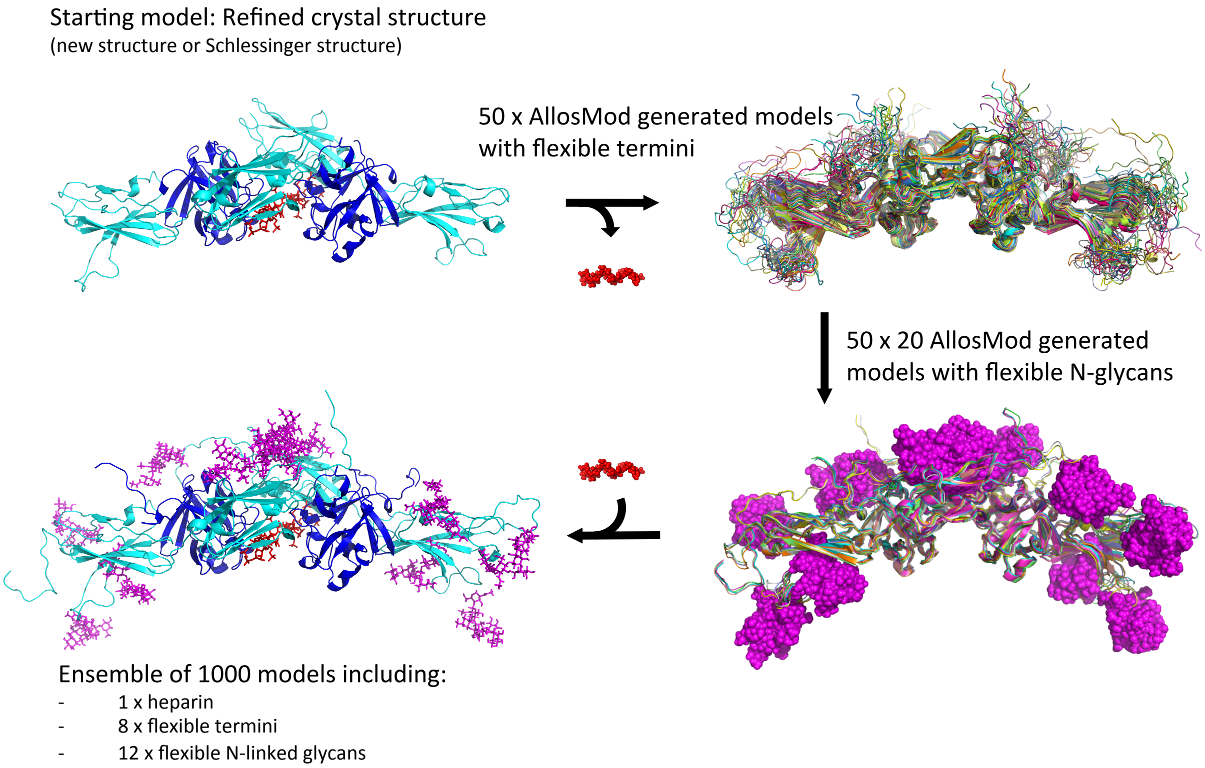


Figure 6.1 Generation of structure ensembles with AllosMod

Schematic workflow for structure ensemble generation, sampling flexible termini and N-linked glycosylations that were disordered in the respective crystal structure. The same procedure was applied to a homology model in Schlessinger architecture, built of the FGFR1-FGF1 unit from the new structure.

6.4.2 The Schlessinger model is not found in solution

SAXS data collection was performed for all samples as described in **Section 2.7** and in **Section 4.8**. Briefly, FGFR1-heparin-FGF1 complex samples were mixed in either 2:1:2 or 2:2:2 stoichiometry freshly before each measurement, diluted to the correct concentration (ranging from 1 to 9 mg/ml), spun down at 15,000 g to remove any aggregates and directly applied to the SAXS analysis pipeline. As in all other experiments, due to high probability of inaccuracies in heparin stock concentrations, all heparin amounts were increased by 10 %.

The experimental scattering data of the FGFR1-heparinDP10-FGF1 complex samples in 2:1:2 and 2:2:2 ratio were fitted to both generated structure ensembles by the ensemble optimisation method. First, CRYSOL (Svergun et al., 1995) was used to calculate theoretical SAXS patterns of each single ensemble structure. Subsequently, GAJOE (Bernadó et al., 2007) was applied to search for an ensemble of 20 structures that generates the optimum fit to the experimental scattering curve. The top left panel in **Figure 6.2** clearly shows, that the experimental scattering curve of the 2:2:2 complex sample does not yield a good fit to the Schlessinger structure ensemble (χ^2 = 11.37). This confirms SEC-MALLS experiments in **Section 5.4.3** providing evidence that excess of heparin leads to heavy signalling complex dissociation. However according to SEC-MALLS, it is likely that the 2:2:2 complex sample contains a small fraction of signalling complex. As the SAXS curve integrates intensities detected from all present protein species, strong dissociation impedes the identification of complex architecture. The same experimental curve yielded an improved fit to the structure ensemble of the new model with χ^2 = 7.71 (top right panel), indicating that the small signalling complex fraction adopts a rather elongated architecture. However, the bad fit does not allow any clear conclusions.

The 2:1:2 complex sample formed a significantly larger signalling complex fraction in SEC-MALLS screening (**Section 5.4.3, Figure 5.6 C**) which is also expressed in improved χ^2 values when fitting the 2:1:2 experimental curve to both crystal structures. Strikingly, the fit to the ensemble of the new structure yielded a χ^2 value of 2.52, compared to a χ^2 value of 8.70 for the Schlessinger ensemble. This is a first indication that the FGFR1-heparinDP10-FGF1 complex in solution might actually adopt the architecture of the new structure presented in this work. Another indicator for quality of fit can be the difference between the experimentally determined radius of gyration (Rg) and the average Rg of all optimised ensemble models. The experimental Rg value integrates scattering signal from all protein species that are present in the sample. Thus, a lower Rg generally indicates a higher degree of dissociation or formation of a signalling complex with smaller Rg. Interestingly, fitting the Schlessinger ensemble to the 2:2:2 sample (top left panel in **Figure 6.2**) yields a ΔRg of only 0.05 Å, which could indicate good structural agreement, if the respective χ^2 value was lower. However, in this case this low ΔRg value is strong evidence, that the 2:2:2 stoichiometry does lead to complex dissociation rather than to formation of the Schlessinger complex with lower Rg. In comparison, ΔRg for the ensemble of the new structure against the 2:1:2 complex sample is 2.70 Å. This is a relatively big difference but as the χ^2 value is relatively low and Rg (calc) relatively high, this ΔRg value is evidence for the presence of a signalling complex species with bigger Rg value than the Schlessinger complex Rg.

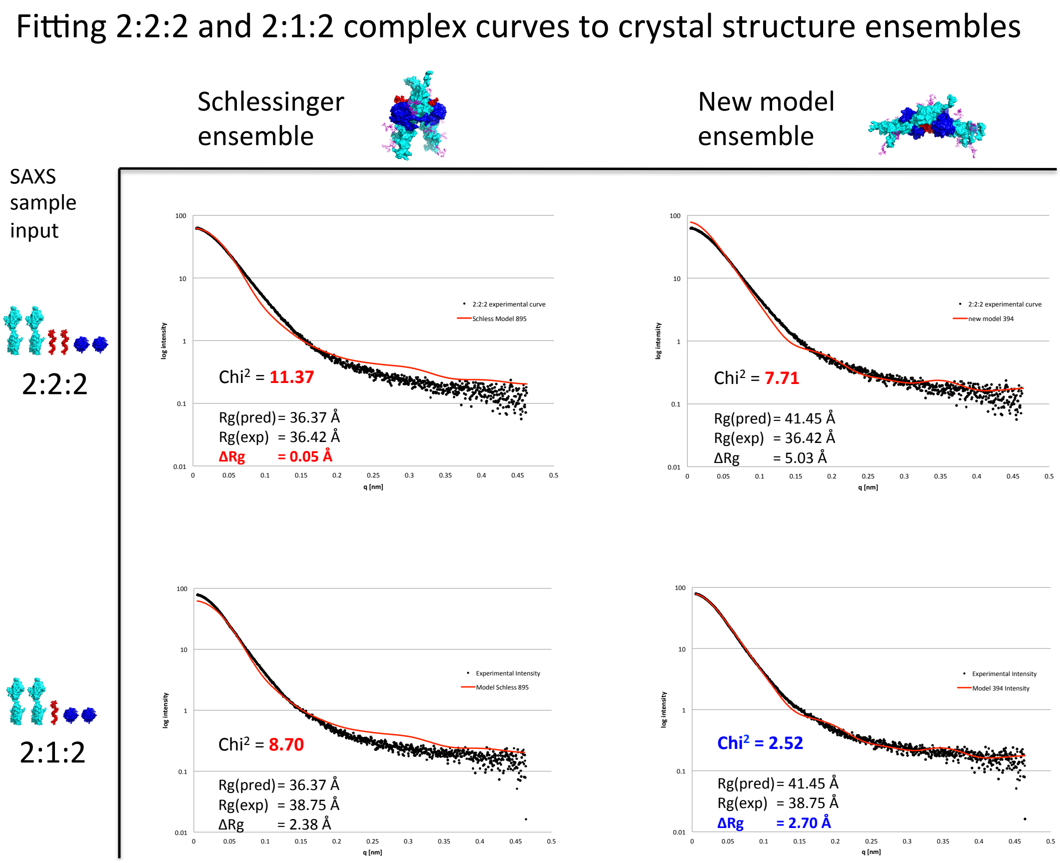


Figure 6.2 SAXS analysis of the FGFR1-heparinDP10-FGF1 sample prepared in different stoichiometries

SAXS analysis of FGFR1-heparinDP10-FGF1 signalling complex samples prepared in Schlessinger stoichiometry (2:2:2; top two panels) and in the stoichiometry observed in the Pellegrini structure and in the new structure (2:1:2; lower two panels). The experimental curves (black curves) of these two samples were fitted to optimised structure ensembles (red curve) comprising flexible protein termini and flexible N-linked glycosylations. Experimental scattering data is shown to a maximal momentum transfer of q = 0.45 nm. Insets describe the goodness of fit value (χ^2), radius of gyration of the calculated curve (Rg(calc)), experimental radius of gyration (Rg(exp)) and the difference between the Rg(calc) and Rg(exp).

6.4.3 Ab initio modelled SAXS envelopes confirm elongated complex architecture

The SAXS curve measured from a 2:1:2 signalling complex sample was used to generate ab initio shape reconstructions using the bead modelling program DAMMIF (Franke & Svergun, 2009). The models generated by ten DAMMIF cycles imposing 2-fold symmetry were subsequently averaged with application of P2 symmetry, using the DAMAVER program suite (Volkov & Svergun, 2003). A second averaged model was generated without any imposed symmetry.

Figure 6.3 A presents top, side and front view of the P2 symmetry ab initio model. Intriguingly, the envelope matches perfectly with the high-resolution crystal structure of the new elongated structure. The resolution of the envelope is good enough to identify single Ig domains and FGF1 molecules. Manual overlay of the crystal structure (with modelled termini and N-glycans) from top, side and front view of the ab initio envelope, reveals that only flexible, disordered components stick out of the envelope. The maximum dimension of the envelope measures 145 Å, which is only 3 Å above the maximum dimension of the crystal structure. These data strongly suggest that the predominant complex species in solution adopts the elongated architecture. Ab initio modelling with released P2 symmetry yielded an asymmetric envelope of similar shape but with less definition of single domains. This indicates that the signalling complex might not be in perfect symmetry in solution as observed in the crystal structure. A lack of dissociated complex species and of asymmetry in the ensemble structures used for SAXS curve fitting in **Section 6.4.3** might be the reasons for poor fit with a $\chi^2 = 2.52$.

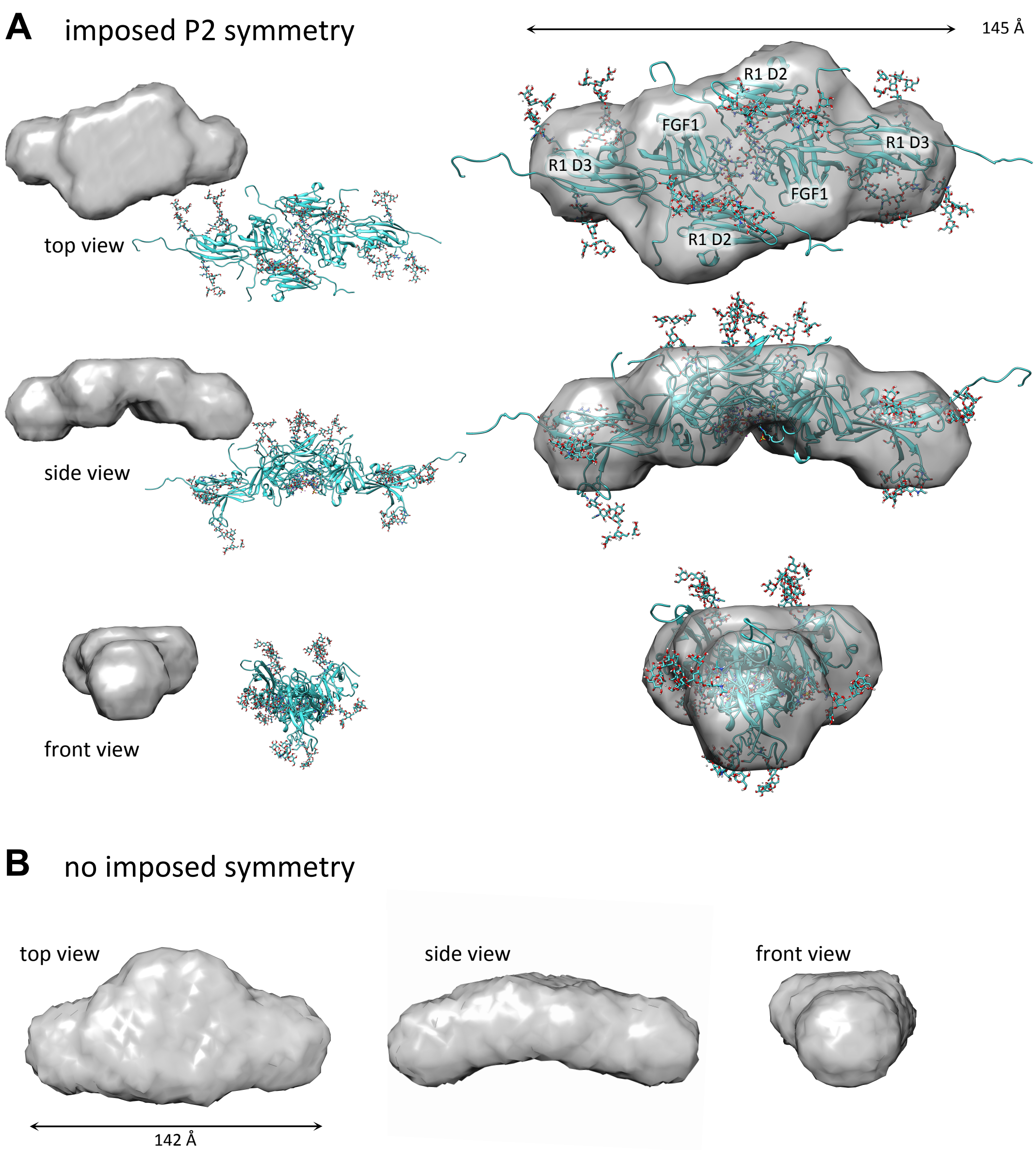


Figure 6.3 Ab initio modelling of the FGFR1-heparinFGF1 complex

(**A**) Ab initio model generated of SAXS diffraction data collected from a 2:1:2 FGFR1-heparin-FGF1 signalling complex sample. Ten DAMMIF cycles with imposed 2-fold symmetry were averaged using DAMAVER. The left panel shows top, side and front views of the envelope shown as grey spheres. The middle panel shows the same views of the new crystal structure after modelling of flexible termini and N-glycans. The right panel shows manual fitting of the crystal structure into the ab initio envelope. (**B**) Top, side and front views of an ab initio model generated from the same sample as in (A) but without imposed symmetry.

6.5 SAXS interpretation on the basis of MD simulations

6.5.1 Generation of a MD simulation derived structural ensemble

The analysis in the preceding chapter has indicated, that the complex architecture of the new structure might be predominant in the FGFR1-heparin-FGF1 2:1:2 signalling complex sample in solution. However, the best fit that could be obtained using the structure ensembles that sample flexible protein termini and N-linked glycosylations, was not sufficient for a clear identification of the complex architecture in solution. In a next step, atomistic molecular dynamics simulations were applied in order to create improved structure ensembles that also sample large scale protein motions within the crystal complex.

Atomistic MD simulations of the Schlessinger structure and of the new structure, including heparin and all flexible termini (excluding N-glycans) were set up as described in **Section 2.1.2** using the GROMACS 4.5.4 (Van Der Spoel et al., 2005) software package in combination with the Charmm27 force field (MacKerell et al., 2001). For each system, three simulation repeats of 50 ns were run and concatenated. Interestingly, the side views of a selected production run of the new structure at t = 0 and t = 50 ns reveal breathing motions of both receptor-ligand units around the longitudinal heparin axis (see blue arrows in **Figure 6.4 A**, top right panel). At the same time, the top view shows that on one side of the heparin chain, secondary receptor-ligand interactions are intensified, whereas the same interactions are released on the other side of the heparin chain. This causes bending of the heparin chain and produces an asymmetric complex that matches very well to the top view of the asymmetric ab initio model (**Figure 6.3 B**, top view). These motions represent important complex configurations that were not sampled in the previous structural ensemble. In contrast, similar snapshots taken from a Schlessinger structure simulation show only minimal deviations from the starting structure. Both

Ig D2 domains seem to align in the longitudinal heparin axis (see blue arrows in **Figure 6.4 B**, bottom right panel) and the heparin ends bend down to form further interactions with the ligand.

A thousand frames were extracted in even time steps from simulations of both complex architectures. N-linked glycosylations were added to each model as described in **Section 6.3.1**. Finally these 2000 models were merged into one big ensemble, together with all possible dissociation species of both complex architectures. The complete ensemble contained a total of 2020 different models.

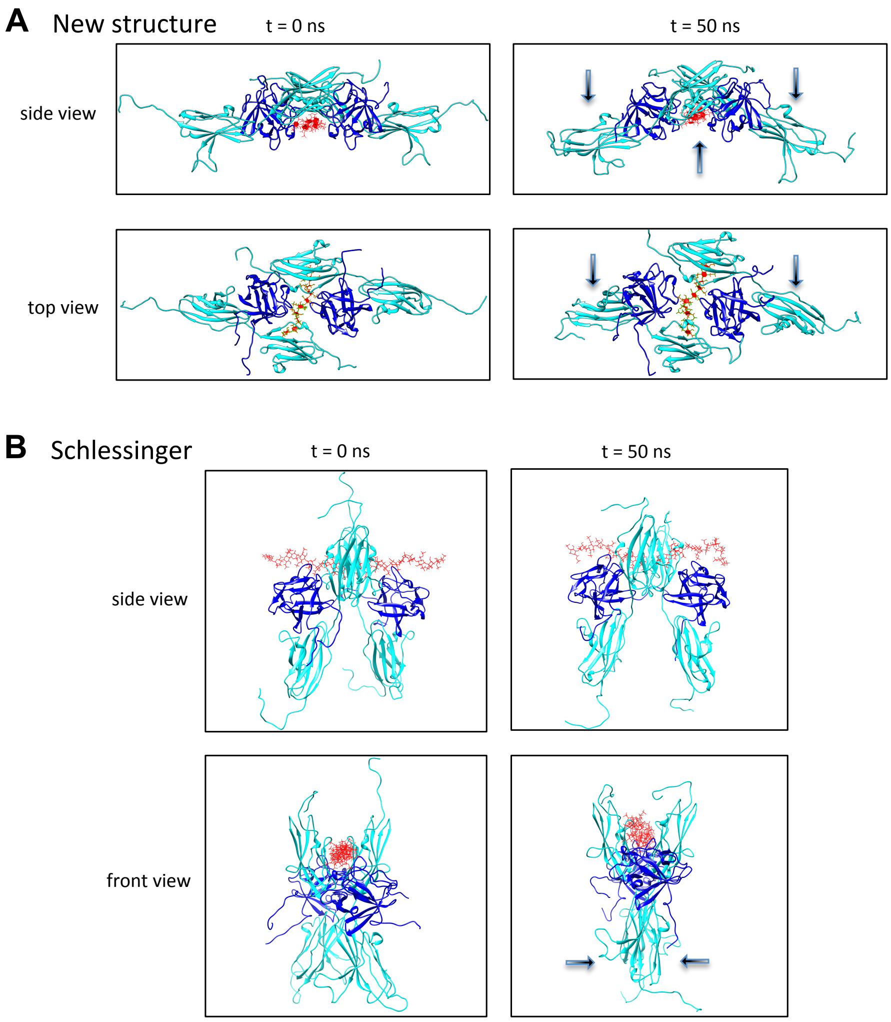


Figure 6.4 Atomistic MD simulations of FGFR1 signalling complexes

Snapshots of one selected simulation at t = 0 (left panel) and at t = 50 ns (right panel) for the new structure (**A**) and the Schlessinger structure (**B**). All structures are coloured identically with the receptor in cyan, the ligand in blue and the heparin chain in red. Blue arrows indicate motions of complex sections through the course of the 50 ns simulation.

6.5.2 The new signalling complex model represents the complex architecture in solution

The analysis in **Section 6.3.2** suggested that most likely FGFR1-heparinDP10-FGF1 in 2:2:2 or 2:1:2 stoichiometry do not form signalling complex in the Schlessinger architecture but rather adopt an elongated architecture as observed in the new structure. The new molecular dynamics structure ensemble generated in **Section 6.5.1** was used to analyse the solution complex architecture in more detail on the basis of a SAXS scattering curves series. Four signalling complex samples using heparinDP8, DP10, DP12 and DP16 in 2:1:2 stoichiometry were used for SAXS data collection in order to analyse the dependency of complex stability on the heparin length. Again, iterative ensemble optimisation was applied using CRYSOL and GAJOE in order to calculate a small ensemble of 20 structures that generates the best fit to the experimental curve. In contrast to the previous optimal ensemble search with a separate ensemble for each complex architecture, here, GAJOE search is performed on one big ensemble (containing Schlessinger model, the new model and all dissociated complex species).

Intriguingly, the new MD simulation-derived ensemble yielded a significant improvement of χ^2 values (see **Figure 6.5 A**; coloured curve: calculated curve from ensemble optimisation method; black curve: experimental scattering), with the best fit found for the heparinDP12 sample (χ^2 = 1.21, green curve) and the least good fit for the heparinDP16 sample (χ^2 = 1.63, red curve). The fit for the heparinDP10 sample has improved from χ^2 = 2.52 to χ^2 = 1.25. However, as the structural ensemble of 2020 models contained a broad mixture of different complexes and even single proteins, it is important to analyse the individual models that GAJOE has combined to generate these improved fits. **Figure 6.5. B** plots the Rg distribution within each optimised ensemble, which reveals the percentage of full signalling complex models. Interestingly, the optimised ensemble to fit the heparinDP8 sample (blue curve) contained the smallest fraction of signalling complex structures (40 %) and thereby the highest fraction of dissociated species (60

%). This suggests that of the tested heparin chains, heparinDP8 has the lowest potential of keeping the signalling complex together in solution. Importantly, not a single model in Schlessinger architecture has been included to the optimised ensemble and all selected signalling complex models in the new architecture adopted a slightly bent configuration, as observed in the MD simulations. The same is valid for the heparinDP10 and DP12 complex samples; only models of the new structure plus dissociated species were included to the respective optimised ensembles. These results strongly suggest that

a) The elongated architecture of the new structure represents the signalling complex architecture in solution
b) The receptor-ligand units bend towards each other around the longitudinal heparin axis in solution
c) Only about 50 % of the samples prepared with heparinDP8, 10 or 12 form signalling complex in solution

Against expectations, the complex sample prepared with heparin DP16 was fitted with an optimised ensemble containing almost exclusively models based on the new structure (red curve in **Figure 6.5 B**). The ensemble contained only one single structure in Schlessinger architecture, but no dissociated species. The reason for the slightly weaker fit with $\chi^2 = 1.63$ could be that all simulations were run with heparinDP10, which leaves 6 unmodelled saccharide units. A representation of the new structures electrostatic potential at a lower potential range from red (−3kbT/ec) to blue (+3kbT/ec) (**Figure 6.5 C**) reveals that the heparin binding canyon has unoccupied, positively charged flanking regions on either sides. HeparinDP16 could theoretically be modelled perfectly into the extended binding canyon with formation of additional hydrogen bonds. These additional bonds might be the reason why according to the SAXS analysis, heparinDP16 forms a more stable complex. However, new simulations of the signalling complex containing heparinDP16 have to be run in order to amend the structural ensemble and new rounds of optimised ensemble search will show whether the fit improves.

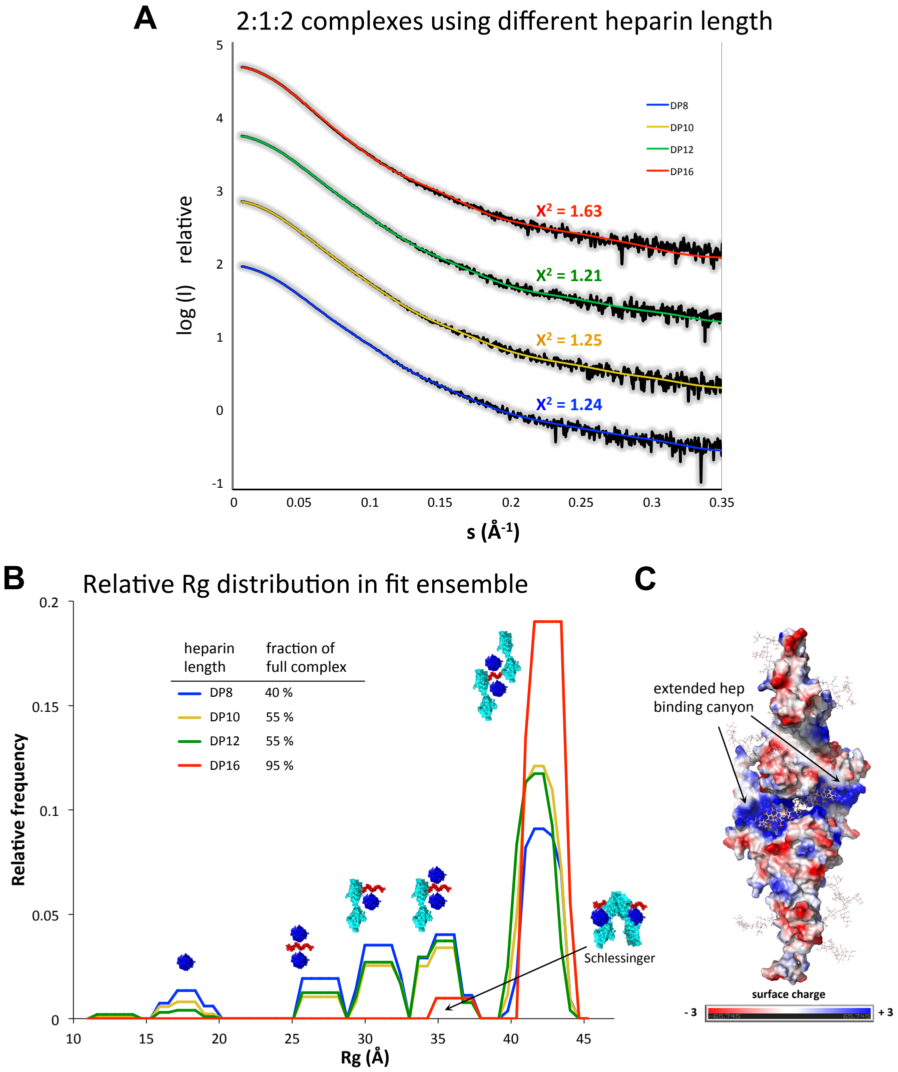


Figure 6.5 Analysis of signalling complex formation in heparin length dependency

(**A**) Fitting of experimental SAXS curves (black curves) by optimised ensemble method of 2:1:2 signalling complex samples formed with heparinDP8 (blue), 10 (yellow), 12 (green) and 16 (red). Complex concentration in all samples was 3 mg/ml. (**B**) Plot of the relative Rg distribution in each optimised ensemble. The inset lists the percentage of elongated signalling complex structures within the 20 selected models. Each single model represents 5 %. Small cartoon representations highlight the complex species that corresponds to the respective Rg values. (**C**) Electrostatic potential representation of the new structure at potential range from red (–3kbT/ec) to blue (+3kbT/ec).

6.6 Conclusion and future perspectives

This chapter describes the first solution structural data of an FGFR signalling complex. The good fit of experimental SAXS data to structural ensembles derived from MD simulations, suggest that the signalling complex adopts the elongated architecture as observed in the crystal structure presented in **Chapter 5**. However, it is important to note that the crystal structure, the SAXS data and the MD simulations were obtained from systems, which neither included the Ig domain 1 nor the cellular membrane. Cofactors on the cell surface might play important roles in FGFR complex formation. Whereas Schlessinger and Pellegrini obtained their crystal structures in absence of N-linked glycosylations, the SAXS data were collected from $Man_5GlcNAc_2$ glycosylated receptor sample. It could be argued that the complex formation in the absence of the freedom of motion-limiting cell membrane context but in the presence of bulky N-linked glycans is artificially pushed towards an elongated complex architecture. However, the discussion in **Section 5.6.2** (**Figure 5.13**) has shown that the more compact formation of the Schlessinger or Pellegrini architecture encounters steric hindrances in the presence of N-linked glycosylations. Furthermore, in the crystallisation trials leading to the new crystal structure, only one single GlcNAc residue was present at each glycosylation site of the receptor, which would not have disturbed the potential formation of a complex in Schlessinger or Pellegrini mode.

The good agreement of the complex solution data with the crystal structure suggests that even if the new model is not a biologically active complex form, it is still likely that many previous studies, providing evidence for the Schlessinger or the Pellegrini structure, were actually presenting data that were collected from a complex arranged as observed in the new crystal structure. It could well be, that the new arrangement can not be formed on the cell surface, but it seems that it is the most favourable arrangement in solution and in low sulphate crystallisation conditions. Two new structural based mutant receptor constructs have been produced that by introduction of a new N-glycosylation site either disrupt the Schlessinger receptor-receptor interactions or the secondary receptor-ligand interface in the new structure. Preliminary expression tests have shown that these constructs can be expressed efficiently as FGFR1 D2D3 constructs (for in vitro experiments) and as fluorescently tagged, full length constructs (for interaction studies in COS7 cells). SEC-MALLS complex formation and FRET-FLIM experiments probing for receptor-receptor interaction, combined with TIRF-dSTORM experiments probing for receptor clustering will help us to get a more clear picture of the biological importance of the various presented signalling complex architectures.

Chapter 7

Further discussion and future directions

Chapter 7 – Further discussion and future directions

7.1 Summary

This thesis represents a structural and computational study of fibroblast growth factor receptor signalling complex formation with a focus on extracellular and TMD dimerisation events. As discussed in **Chapter 3**, coarse-grained and atomistic simulations of wild-type and mutant FGFR3 TMDs have revealed new putative mechanistic details on the mode of action of the Achondroplasia mutation. Coarse-grained simulations of FGFR4 have yielded the first structural models for wild-type and mutant FGFR4 TMD dimers. **Chapter 4** describes attempts to crystallise ligand bound FGFR4. Along the way, novel methods for expression and purification of glycosylated FGFR, and for heparin mediated signalling complex formation were developed. A detailed study of FGF-heparin complex formation is described in **Chapter 5**, followed by the presentation of a new FGFR1-heparin-FGF1 signalling complex crystal structure. This structure represents a new arrangement, which is in agreement with SAXS structural solution data as presented in **Chapter 6**.

7.2 Further discussion

Discussions at the end of each results chapter have underscored that there are many important questions that are yet to be addressed in order to gain a better understanding of the mechanism of RTK signalling in health and disease. Three crystal structures of the extracellular signalling complex suggest fundamentally different molecular arrangements, which suggest differing molecular signalling mechanisms. The only similarity across all complex architectures is that they feature heparin as cross linker for at least one receptor-ligand dimer.

7.2.1 The Schlessinger Model

In theory the Schlessinger two-end model requires independent formation of two 1:1:1 receptor-heparin-ligand ternary complexes that stabilise a symmetric 2:2:2 signalling complex by direct receptor–receptor interactions and by interactions of heparin and ligand to the secondary complex elements. However, in a biological context this depends on the rather less frequent event of two long heparin chains cross linking the independent 1:1 binary complexes not only selectively with their non-reducing ends but also in correct directionality (with the non-reducing ends facing towards the pseudo-C2 symmetry axis of the signalling complex; see **Figure 5.15 B and E**; **Section 5.7.2**). Notably, this complex assembly does not rely on a single molecule triggering receptor dimerisation. Remarkably, the crystals of that structure were only producible in the absence of heparin, suggesting heparin independent receptor dimerisation. In that case heparin would only enhance FGFR signalling, but not control dimerisation and signalling complex formation as described very early (Ornitz et al., 1992; Spivak-Kroizman et al., 1994). However, this complex architecture remains appealing due to structure based mutation studies presenting “direct evidence for the symmetric two-end model for FGFR dimerisation” (Ibrahimi et al., 2005), by identification of a pathogenic mutation located in the receptor-receptor dimer interface. In addition, analysing the putative orientation of the Schlessinger structure relative to the cellular membrane, it is the only structure that would allow *cis*-dimerisation of the transmembrane domains, as shown in **Figure 7.1** Nevertheless, this structure does not conform to studies providing evidence that heparin drives FGF dimerisation in a cooperative fashion (see **Section 5.3.2** and **Section 5.7.3**).

7.2.2 The Pellegrini model

In the Pellegrini structure the distance between last ordered C-terminal receptor residues is 74 Å, which does not allow TMD dimerisation in *cis*. A TMD dimer in cross conformation, flexible motions of the receptor-ligand dimer units moving the C-termini closer or TMD dimerisation in *trans* could solve this problem. However, there are further serious doubts about the potential biological relevance of the Pellegrini model: **a**) Only one receptor-ligand unit is cross linked by heparin, leaving a unsaturated, unstable and highly positively charged heparin binding pocket (see **Figure 5.17**). **b**) The heparin chain adopts a prominent atypical 34° kink between the 6th and the 7th saccharide for FGF dimerisation (see **Section 5.7.1**). **c**) One ligand in the asymmetric 2:1 FGF-heparin complex is distorted (not binding with the canonical heparin-binding site; see **Section 5.7.3**). **d**) Proline235 that is located in the D2–D3 linker adopts a *cis*-conformation, which results in rotation of Ig domain 3. In biology *cis*-prolines occur very rarely, but more importantly, the reoriented D3 is not compatible with the well-established mechanism for ligand binding selectivity (see **Figure 5.14 C**). This partial list of arguments strongly suggests that the Pellegrini structure is the result of the crystallisation process (biased by high salt concentration and tight packing) rather than the complex that would be formed under physiological conditions.

7.2.3 The new model and structural analysis in solution

The new model is generally consistent with all experimental data that has been published to support the Schlessinger or the Pellegrini structure. It is the only model that can be formed independent of heparin position and heparin directionality (see **Figure 5.15 A**), that involves all 15 heparinDP10 HS moieties in hydrogen bonding and that features a single heparin chain that cross links both receptors and both ligands (see **Figure 5.15 D**). The new model is also unique in that the central heparin mediated FGF dimer adopts the same symmetric *trans*-configuration and the same interaction pattern as the previously published active 2:1 FGF-heparin complex (DiGabriele et al., 1998).

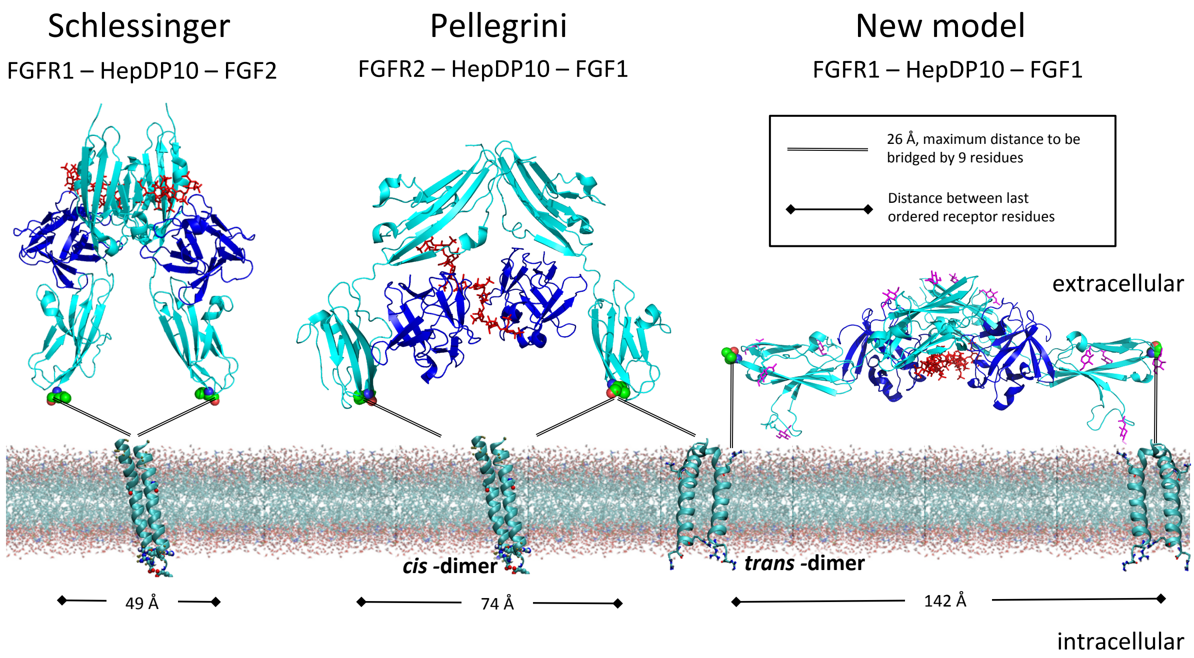


Figure 7.1 Signalling complexes in the context of the cellular membrane

Presentation of the FGFR signalling complexes according to Schlessinger, Pellegrini and the new model. All structures are shown in identical representations: receptor in cyan cartoon, ligand in blue cartoon, heparin in red sticks, the last ordered residue at the receptor C-terminus as green spheres plus colouring by element, GlcNAc residues in purple sticks. The double line represents the maximum lengths of 26 Å that can be bridged by 9 residues that are missing between the last ordered C-terminal residue of the crystal structures and the first residue of the predicted FGFR transmembrane domain. The TMD dimer cartoons are snapshots of an atomistic FGFR3 dimer simulation produced in this work. They are embedded into a model of a POPC lipid bilayer. The black lines below each model display the distance between two C-terminal receptor residues.

The lack of both crystal structures of full-length FGFR monomers or dimers and of any structural data of the extracellular FGFR signalling complex in solution represents a huge problem. Besides structure based mutation studies, most studies that aimed to discriminate between the Schlessinger and the Pellegrini model applied size exclusion chromatography, analytical ultracentrifugation and mass spectrometry – methods that collect data from soluble protein (or protein in gaseous phase) but do not generate conclusive structural information. In essence, it is generally uncertain which complex architecture was present during data collection for these many published FGFR studies. This thesis presents the first structural data for a FGFR signalling complex in solution by application of small angle X-ray scattering. The presented analysis suggests that receptor-heparin-ligand complex sample prepared in 2:2:2 (Schlessinger)-stoichiometry results in strong complex dissociation. The collected spectra could not be fitted to any of the three complex architectures (see **Figure 6.2** in **Section 6.3.2**). Analysis of a 2:1:2 SAXS sample in combination with structural ensembles generated by atomistic MD simulations yielded a convincing fit with the new model (**Figure 6.5 A** in **Section 6.5.2**). Interestingly, complex samples with heparinDP8 revealed impeded efficiency of complex formation (**Figure 6.5 B**) – an observation that only makes sense in the context of the new complex model, as longer saccharide chains are capable of cross linking receptors on both sides of the central FGF dimer with formation of an additional set of contacts on an extended heparin binding site on Ig domain 2 (**Figure 6.5 C**). This conclusion is supported by the SAXS data obtained from the complex sample formed with heparinDP16, which leads to maximisation of heparin-protein interactions in the ternary complex. The good fit of the heparinDP10 complex data to the new crystal structure suggests that, even if the new model does not represent a biologically active complex architecture, it is still likely that previous studies actually presented data that were collected from a complex arranged as observed in the new crystal structure. It could well be, that the new arrangement cannot be formed on the cell surface, but it seems that the new arrangement is the most favourable in solution. SAXS experiments performed with FGFR4 (the only FGFR homolog of which no structural ectodomain data is available) strongly suggest that FGFR4 adopts the same signalling complex architecture as FGFR1 (see **Section 4.8**). Considering that FGFR4 has the lowest sequence identity to any other FGFR homolog, this shows that the new architecture is not only specific to the FGFR1-FGF1 receptor-ligand combination but that it can also be adopted by other receptor homologs

7.2.4 FGFR signalling through the membrane

Multiple pathogenic TMD mutations leave no doubt that the TM domains play an important role in the process of signal transduction from the extracellular space to the cytoplasm. However detailed mechanisms are yet to be determined. The simulations of TMD systems presented in **Chapter 3** have shown that most of the studied FGFR3 and FGFR4 mutations (including the Achondroplasia mutant) do not form part of the identified dimerisation interfaces (see **Section 3.5.2.1** for FGFR3 **Section 3.6.1** for FGFR4). This suggests that mutant TMDs act through more sophisticated mechanisms than simple aberration of the dimerisation kinetics. In the new signalling complex model, the C-terminal ordered receptor residues are separated by 142 Å. Even though atomistic MD simulations have shown "breathing" motions around the heparin axis that drive the receptor's C-termini closer, it is clear that the TMDs can not get close enough to form dimers. Thus, if TMD dimers are required for signalling, they can only be formed to the TMD of an adjacent receptor complex (as the modelled TMD dimer in **Figure 7.1** suggests), which would consequently result in local receptor clustering. It is evident that in the cell, RTK receptor systems can feature formation of clusters upon activation (Franco et al., 2003; Sacchettini, 2001; Seiradake et al., 2013). Direct evidence for FGFR clustering has not yet been shown. However, it is known that FGFRs bind to multiple other proteins, both extracellularly (e.g. N-CAM (Cavallaro et al., 2001), N-cadherin (Utton, et al., 2001)) and intracellularly (FRS2α, PLCγ1 and further adaptor proteins) that could drive FGFR clustering in response to FGF. Furthermore, concluding from the discussion in **Section 5.7.2**, the new model seems to represent the only arrangement that is compatible with formation of multiple FGFR–heparin–FGF complex units upon a single HS chain. Studies of RTK juxtamembrane domains have described detailed mechanisms for kinase activation by conformational coupling across the plasma membrane (Bocharov et al., 2013; Endres et al., 2013; Hubbard, 2009; Jura et al., 2009). These mechanisms can come into action independently from TMD dimerisation between two single receptors or by formation of a larger TMD-linked receptor network. In fact, whereas the long disordered juxtamembrane domains are theoretically capable of bridging substantial distances, there is no direct evidence yet that TMD dimerisation is required for FGFR signalling.

7.2.5 Future work

From the preceding discussion it is clear that extensive experimental efforts are needed to gain a better understanding of FGFRs signalling mechanisms across the membrane. Consideration of extracellular and intracellular binding partners will be key in both functional and structural studies. Significant advances in electron microscopic methods could provide particularly valuable structural information on the binding mode of additional factors and give insights into the kinase domain arrangement. However, as briefly reported in **Section 6.2**, the major bottleneck for structural studies of the full-length FGFR is the isolation of adequate amounts of monodisperse protein sample. At this stage, the most promising method for production of full-length receptor is expression in cell-derived membrane vesicles (Zeev-Ben-Mordehai et al., 2014). Two major challenges here currently are **a**) cell surface HSPG that might be preventing FGFR from budding out in vesicles and **b**) diminished receptor presence at the cell surface due to FGF that is contained in the expression medium, leading to activation and subsequent down regulation of the receptor. In connection with this thesis, structure based mutations were cloned that introduce new N-glycosylation sites to either disrupt the Schlessinger receptor-receptor interactions or the secondary receptor-ligand interface in the new structure. Förster resonance energy transfer (FRET) applications in various flavours are capable of detecting changes of protein distances in live cells and can therefore help to discriminate between different structural arrangements. Applied intelligently, FRET-FLIM (fluorescence-lifetime imaging) can be used for detection of dimerisation intensities (see (Padilla-Parra, , 2015) for FRET-FLIM). The mutant constructs will potentially show different FRET-FLIM signals, indicating which complex arrangement is formed in live cells. In addition, TIRF-dSTORM experiments applied in previous studies (Healey et al., 2015) could show whether any of the mutant receptor constructs induce receptor clustering. In the future, structural studies will also have to address the role of HS specificity in complexes assembled of different receptor and ligand types. Naturally occurring HS sequences present specific patterns of sulphate groups that elicit both activating or inhibiting physiological responses. The underlying rules to these processes need to be understood in more detail in order to develop heparan sulphate based treatments. Lastly, future studies have to contribute to better understanding of how canonical FGF signalling pathways generate specific cellular responses and how enhanced expression or disruption of these pathways can lead to developmental defects and cancer. Answering these questions could drive the development of a new generation of drugs that block or attenuate FGFR activity.

Appendix

Appendix A – Expression Vectors

A.1 Mammalian Expression: pHLsec

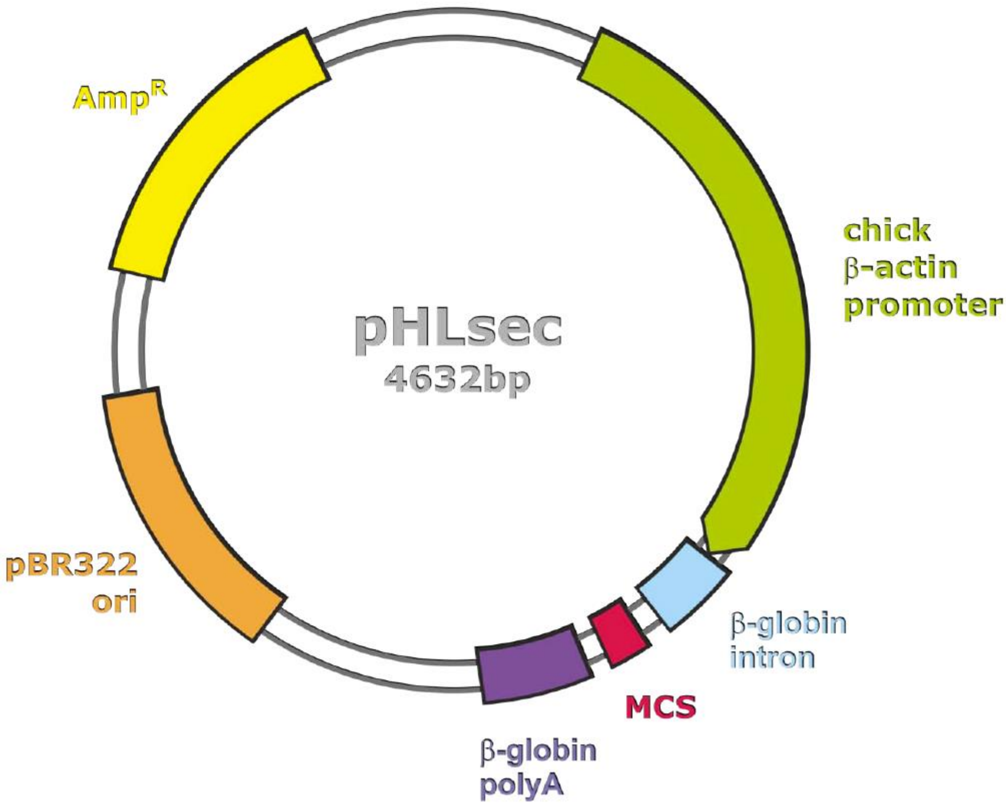


Figure A.1 Key features of the pHLsec vector backbone

Figure provided by Dr. Radu Aricescu. Multiple Cloning sites for pHLSec and subsequent vector variations used are provided below, and are based on Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. (Aricescu et al., 2006).

Section of the pHLsec vector Multiple Cloning Site

```
GAATTCAAGCTTGCCACCATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAG
CCTTCTCTCCGTGCTGCTGATGGGTTGCGTAGCTGAAACCGGT...INSERT...GGTACCAAGCACC
ACCATCACCATCACTAATGATCACTCGAG
```

EcoR1 - HinD3 - Kozak with start codon - Signal Particle – Linker – Age1 - INSERT - Kpn1 – Linker - His-6 tag – Stop - Xho1

A.2 Bacterial Expression: pET22b(+)

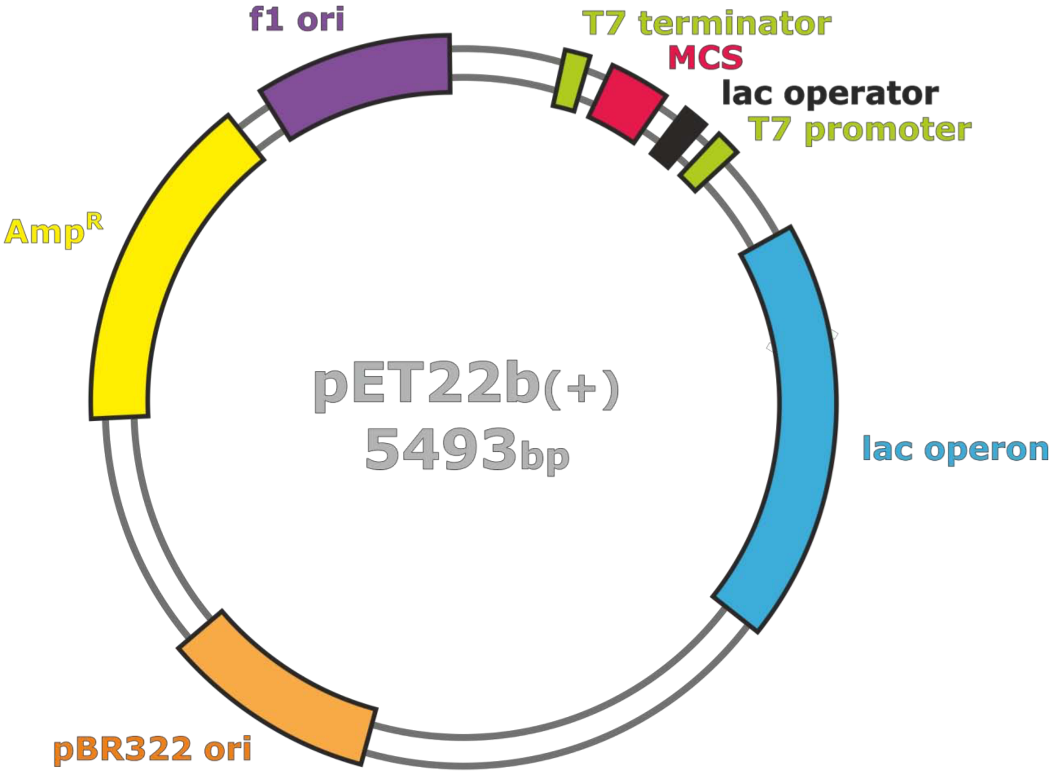


Figure A.2 Key features of the pET22b(+) vector (Invitrogen)

Key features of the expression vector pET22b(+) for protein expression in *E.coli*. A section of the Multiple Cloning Site is depicted below. The two restriction sites Age1 and Kpn1 have been added to the vector by Dr. Radu Aricescu (University of Oxford, department of Structural Biology). A second variation of this vector has been used which has a removed 6His-tag.

Section of the pET22b(+) vector Multiple Cloning Site

AAGGAGATATA**CATATGACCGGT**...INSERT...**GGTACCAAGCAC**CACCATCACCAT**CACTAA**TGA **TCA**CTCGAG

RBS - TATA-box - Nde1+Start - Age1 - INSERT - Kpn1 - Linker - His-6 tag - Stop - Xho1

Appendix B – hFGFR sequence alignment

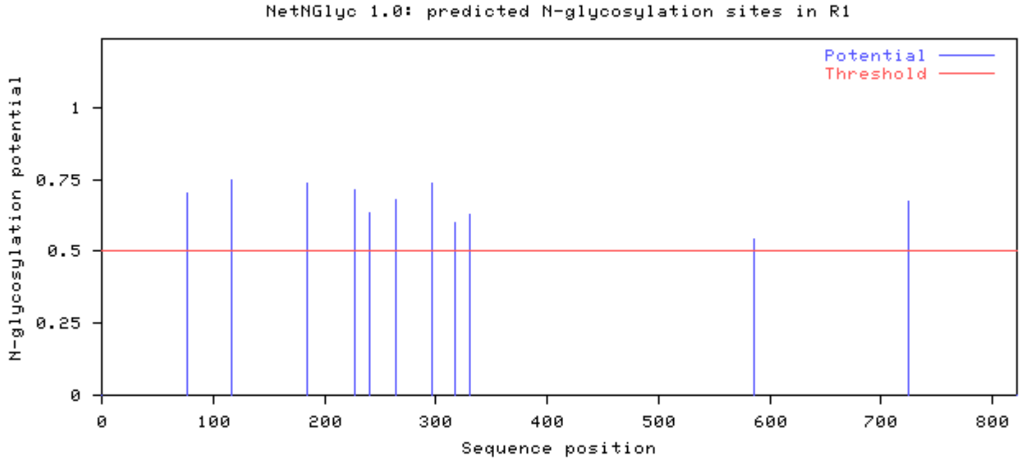
```
             1           11          21          31          41
Consensus       --sw-aclll  avlavatlap  arpspslpeq  aqvepe----  --apqqqeql
Conservation
R1          1   MWSWKCLLFW  AVLVTATLCT  ARPSPTLPEQ  AQPWG-----  --APVEVESF  43
R2          1   MVSWGRFICL  VVVTMATLSL  ARPSFSLVED  TTLEPEEPPT  KYQISQPEVY  50
R3          1   -MGAPACALA  LCVAVAIVAG  ASSESLGTEQ  RVVGRAAEVP  GPEPGQQEQL  49
R4          1   -----MRLLL  ALLGVLLSVP  GPPVLSLEAS  EEVELEPCLA  PSLEQQEQEL  45

             51          61          71          81          91
Consensus       vvapGdalel  rCpla--d--  qii-WykdGv  qLapsnRtri  igerlqiqsa
Conservation
R1          44  LVHPGDLLQL  RCRLR--DDV  QSINWLRDGV  QLAESNRTRI  TGEEVEVQDS  91
R2          51  VAAPGESLEV  RCLLK--DAA  VIS-WTKDGV  HLGPNNRTVL  IGEYLQIKGA  97
R3          50  VFGSGDAVEL  SCPPPGGGPM  GPTVWVKDGT  GLVPSERVLV  GPQRLQVLNA  99
R4          46  TVALGQPVRL  CCGRAE----  RGGHWYKEGS  RLAPAGRVRG  WRGRLEIASF  91

             101         111         121         131         141
Consensus       speDsGlYaC  rasstqiset  cyfsvnvtda  isSgdDeDdk  dhaedsnket
Conservation
R1          92  VPADSGLYAC  VTSSPSGSDT  TYFSVNVSDA  LPSSEDDDDD  DDSSSEEKET  141
R2          98  TPRDSGLYAC  TASRTVDSET  WYFMVNVTDA  ISSGDDEDDT  DGAEDFVSEN  147
R3          100 SHEDSGAYSC  RQRLTQR-VL  CHFSVRVTDA  PSSGDDEDGE  DEAEDTGVDT  148
R4          92  LPEDAGRYLC  LARGSMIVLQ  NLTLITGDSL  TSSNDDEDPK  SHRDPSNRHS  141

             151         161         171         181         191
Consensus       yn--------  APYWThpekM  eKkLhAVPAa  nTVkFrCPaa  GnPtPtirWL
Conservation
R1          142 DNTKPNRMPV  APYWTSPEKM  EKKLHAVPAA  KTVKFKCPSS  GTPNPTLRWL  191
R2          148 SNNKR-----  APYWTNTEKM  EKRLHAVPAA  NTVKFRCPAG  GNPMPTMRWL  192
R3          149 G---------  APYWTRPERM  DKKLLAVPAA  NTVRFRCPAA  GNPTPSISWL  189
R4          142 YP------QQ  APYWTHPQRM  EKKLHAVPAG  NTVKFRCPAA  GNPTPTIRWL  185

             201         211         221         231         241
Consensus       KnGkeFkgeh  RIGGyklRhq  hWSliMeSVV  PSDkGnYTCv  VENeyGSIrh
Conservation
R1          192 KNGKEFKPDH  RIGGYKVRYA  TWSIIMDSVV  PSDKGNYTCI  VENEYGSINH  241
R2          193 KNGKEFKQEH  RIGGYKVRNQ  HWSLIMESVV  PSDKGNYTCV  VENEYGSINH  242
R3          190 KNGREFRGEH  RIGGIKLRHQ  QWSLVMESVV  PSDRGNYTCV  VENKFGSIRQ  239
R4          186 KDGQAFHGEN  RIGGIRLRHQ  HWSLVMESVV  PSDRGTYTCL  VENAVGSIRY  235

             251         261         271         281         291
Consensus       tYqLDVIERS  PHRPILQAGL  PANatavlGs  dVEfhCKVYS  DaQPHIQWIK
Conservation
R1          242 TYQLDVVERS  PHRPILQAGL  PANKTVALGS  NVEFMCKVYS  DPQPHIQWLK  291
R2          243 TYHLDVVERS  PHRPILQAGL  PANASTVVGG  DVEFVCKVYS  DAQPHIQWIK  292
R3          240 TYTLDVLERS  PHRPILQAGL  PANQTAVLGS  DVEFHCKVYS  DAQPHIQWLK  289
R4          236 NYLLDVLERS  PHRPILQAGL  PANTTAVVGS  DVELLCKVYS  DAQPHIQWLK  285

             301         311         321         331         341
Consensus       HievNGSkiG  pDglPYvqvL  KtAgvNttdk  EiEVLylrNV  sfEDAGEYTC
Conservation
R1          292 HIEVNGSKIG  PDNLPYVQIL  KTAGVNTTDK  EMEVLHLRNV  SFEDAGEYTC  341
R2          293 HVEKNGSKYG  PDGLPYLKVL  KAAGVNTTDK  EIEVLYIRNV  TFEDAGEYTC  342
R3          290 HVEVNGSKVG  PDGTPYVTVL  KTAGANTTDK  ELEVLSLHNV  TFEDAGEYTC  339
R4          286 HIVINGSSFG  ADGFPYVQVL  KTADINSS--  EVEVLYLRNV  SAEDAGEYTC  333

             351         361         371         381         391
Consensus       LAGNSIGISh  hSAWLtVLpa  eeerpeaaas  -prYleiiiY  ciGafliacm
Conservation
R1          342 LAGNSIGLSH  HSAWLTVLEA  LEERPAVMTS  -PLYLEIIIY  CTGAFLISCM  390
R2          343 LAGNSIGISF  HSAWLTVLPA  PGREKEITAS  -PDYLEIAIY  CIGVFLIACM  391
R3          340 LAGNSIGFSH  HSAWLVVLPA  EEELVEADEA  GSVYAGILSY  GVGFFLFILV  389
R4          334 LAGNSIGLSY  QSAWLTVLPE  EDPTWTAAAP  EARYTDIILY  ASGSLALAVL  383

             401         411         421         431         441
Consensus       vaavilyrmk  sptkkpd---  -pavhklaKi  iplprqrqvS  aeSsaSmnSn
Conservation
R1          391 VGSVIVYKMK  SGTKKSDFHS  QMAVHKLAKS  IPLRRQVTVS  ADSSASMNSG  440
R2          392 VVTVILCRMK  NTTKKPDFSS  QPAVHKLTKR  IPLRRQVTVS  AESSSSMNSN  441
R3          390 VAAVTLCRLR  SPPKKG----  -LGSPTVHKI  SRFPLKRQVS  LESNASMSSN  434
R4          384 LLLAGLYRGQ  ALHGRHP---  -RPPATVQKL  SRFPLARQFS  LESGSSGKSS  429

             451         461         471         481         491
Consensus       tpLVRiar--  lSssgtpmLA  gvseyeLPeD  PkWElpRdrL  tLGKPLGEGC
Conservation
R1          441 VLLVRPSR--  LSSSGTPMLA  GVSEYELPED  PRWELPRDRL  VLGKPLGEGC  488
R2          442 TPLVRITTRL  SSTADTPMLA  GVSEYELPED  PKWEFPRDKL  TLGKPLGEGC  491
R3          435 TPLVRIAR--  LSSGEGPTLA  NVSELELPAD  PKWELSRARL  TLGKPLGEGC  482
R4          430 SSLVRGVR--  LSSSGPALLA  GLVSLDLPLD  PLWEFPRDRL  VLGKPLGEGC  477

             501         511         521         531         541
Consensus       FGQVVmAEAi  GiDkdkpaqa  vtVAVKMLKd  dAteKDLsDL  vSEMEmMKmI
Conservation
R1          489 FGQVVLAEAI  GLDKDKPNRV  TKVAVKMLKS  DATEKDLSDL  ISEMEMMKMI  538
R2          492 FGQVVMAEAV  GIDKDKPKEA  VTVAVKMLKD  DATEKDLSDL  VSEMEMMKMI  541
R3          483 FGQVVMAEAI  GIDKDRAAKP  VTVAVKMLKD  DATDKDLSDL  VSEMEMMKMI  532
R4          478 FGQVVRAEAF  GMDPARPDQA  STVAVKMLKD  NASDKDLADL  VSEMEVMKLI  527
```

```
                   351        361        371        381        391
Consensus          LAGNSIGISh hSAWLtVLpa eeerpeaaas -prYleiiiY ciGafliacm
Conservation
R1            342  LAGNSIGLSH HSAWLTVLEA LEERPAVMTS -PLYLEIIIY CTGAFLISCM  390
R2            343  LAGNSIGISF HSAWLTVLPA PGREKEITAS -PDYLEIAIY CIGVFLIACM  391
R3            340  LAGNSIGFSH HSAWLVVLPA EEELVEADEA GSVYAGILSY GVGFFLFILV  389
R4            334  LAGNSIGLSY QSAWLTVLPE EDPTWTAAAP EARYTDIILY ASGSLALAVL  383

                   401        411        421        431        441
Consensus          vaavilyrmk sptkkpd--- -pavhklaKi iplprqrqvS aeSsaSmnSn
Conservation
R1            391  VGSVIVYKMK SGTKKSDFHS QMAVHKLAKS IPLRRQVTVS ADSSASMNSG  440
R2            392  VVTVILCRMK NTTKKPDFSS QPAVHKLTKR IPLRRQVTVS AESSSSMNSN  441
R3            390  VAAVTLCRLR SPPKKG---- -LGSPTVHKI SRFPLKRQVS LESNASMSSN  434
R4            384  LLLAGLYRGQ ALHGRHP--- -RPPATVQKL SRFPLARQFS LESGSSGKSS  429

                   451        461        471        481        491
Consensus          tpLVRiar-- lSssgtpmLA gvseyeLPeD PkWElpRdrL tLGKPLGEGC
Conservation
R1            441  VLLVRPSR-- LSSSGTPMLA GVSEYELPED PRWELPRDRL VLGKPLGEGC  488
R2            442  TPLVRITTRL SSTADTPMLA GVSEYELPED PKWEFPRDKL TLGKPLGEGC  491
R3            435  TPLVRIAR-- LSSGEGPTLA NVSELELPAD PKWELSRARL TLGKPLGEGC  482
R4            430  SSLVRGVR-- LSSSGPALLA GLVSLDLPLD PLWEFPRDRL VLGKPLGEGC  477

                   501        511        521        531        541
Consensus          FGQVVmAEAi GiDkdkpaqa vtVAVKMLKd dAteKDLsDL vSEMEmMKmI
Conservation
R1            489  FGQVVLAEAI GLDKDKPNRV TKVAVKMLKS DATEKDLSDL ISEMEMMKMI  538
R2            492  FGQVVMAEAV GIDKDKPKEA VTVAVKMLKD DATEKDLSDL VSEMEMMKMI  541
R3            483  FGQVVMAEAI GIDKDRAAKP VTVAVKMLKD DATDKDLSDL VSEMEMMKMI  532
R4            478  FGQVVRAEAF GMDPARPDQA STVAVKMLKD NASDKDLADL VSEMEVMKLI  527

                   551        561        571        581        591
Consensus          GkHKNIINLL GaCTQdGPLY ViVEyAaKGN LREyLrARRP PGleysydip
Conservation
R1            539  GKHKNIINLL GACTQDGPLY VIVEYASKGN LREYLQARRP PGLEYCYNPS  588
R2            542  GKHKNIINLL GACTQDGPLY VIVEYASKGN LREYLRARRP PGMEYSYDIN  591
R3            533  GKHKNIINLL GACTQGGPLY VLVEYAAKGN LREFLRARRP PGLDYSFDTC  582
R4            528  GRHKNIINLL GVCTQEGPLY VIVECAAKGN LREFLRARRP PGPDLSPDGP  577

                   601        611        621        631        641
Consensus          rppEeqlsfk dLVSCaYQvA RGMeYLaSqK CIHRDLAARN VLVTEdNVMK
Conservation
R1            589  HNPEEQLSSK DLVSCAYQVA RGMEYLASKK CIHRDLAARN VLVTEDNVMK  638
R2            592  RVPEEQMTFK DLVSCTYQLA RGMEYLASQK CIHRDLAARN VLVTENNVMK  641
R3            583  KPPEEQLTFK DLVSCAYQVA RGMEYLASQK CIHRDLAARN VLVTEDNVMK  632
R4            578  RSSEGPLSFP VLVSCAYQVA RGMQYLESRK CIHRDLAARN VLVTEDNVMK  627

                   651        661        671        681        691
Consensus          IADFGLARdi hhiDYYKKTt NGRLPVKWMA PEALFDRvYT HQSDVWSFGv
Conservation
R1            639  IADFGLARDI HHIDYYKKTT NGRLPVKWMA PEALFDRIYT HQSDVWSFGV  688
R2            642  IADFGLARDI NNIDYYKKTT NGRLPVKWMA PEALFDRVYT HQSDVWSFGV  691
R3            633  IADFGLARDV HNLDYYKKTT NGRLPVKWMA PEALFDRVYT HQSDVWSFGV  682
R4            628  IADFGLARGV HHIDYYKKTS NGRLPVKWMA PEALFDRVYT HQSDVWSFGI  677

                   701        711        721        731        741
Consensus          LIWEIFTLGG SPYPGiPVEE LFkLLkEGHR MDkPanCtne LYmmMReCWH
Conservation
R1            689  LLWEIFTLGG SPYPGVPVEE LFKLLKEGHR MDKPSNCTNE LYMMMRDCWH  738
R2            692  LMWEIFTLGG SPYPGIPVEE LFKLLKEGHR MDKPANCTNE LYMMMRDCWH  741
R3            683  LLWEIFTLGG SPYPGIPVEE LFKLLKEGHR MDKPANCTHD LYMIMRECWH  732
R4            678  LLWEIFTLGG SPYPGIPVEE LFSLLREGHR MDRPPHCPPE LYGLMRECWH  727

                   751        761        771        781        791
Consensus          AaPSQRPTFK QLVEdLDril tltsneEYLD LsqpleqYSP sgpDtr-ssc
Conservation
R1            739  AVPSQRPTFK QLVEDLDRIV ALTSNQEYLD LSMPLDQYSP SFPDTRSSTC  788
R2            742  AVPSQRPTFK QLVEDLDRIL TLTTNEEYLD LSQPLEQYSP SYPDTR-SSC  790
R3            733  AAPSQRPTFK QLVEDLDRVL TVTSTDEYLD LSAPFEQYSP GGQDTP-SSS  781
R4            728  AAPSQRPTFK QLVEALDKVL LAVS-EEYLD LRLTFGPYSP SGGDAS--ST  774

                   801        811        821        831
Consensus          sSgdDSVFsh dplPyepcsp qypaq----- ----
Conservation
R1            789  SSGEDSVFSH EPLPEEPCLP RHPAQLANGG LKRR  822
R2            791  SSGDDSVFSP DPMPYEPCLP QYPHINGSVK T---  821
R3            782  SSGDDSVFAH DLLPPAPPSS GGSRT----- ----  806
R4            775  CSSSDSVFSH DPLPLGSSSF PFGSGVQT-- ----  802
```

Appendix D – NetNGlyc 1.0 Server prediction

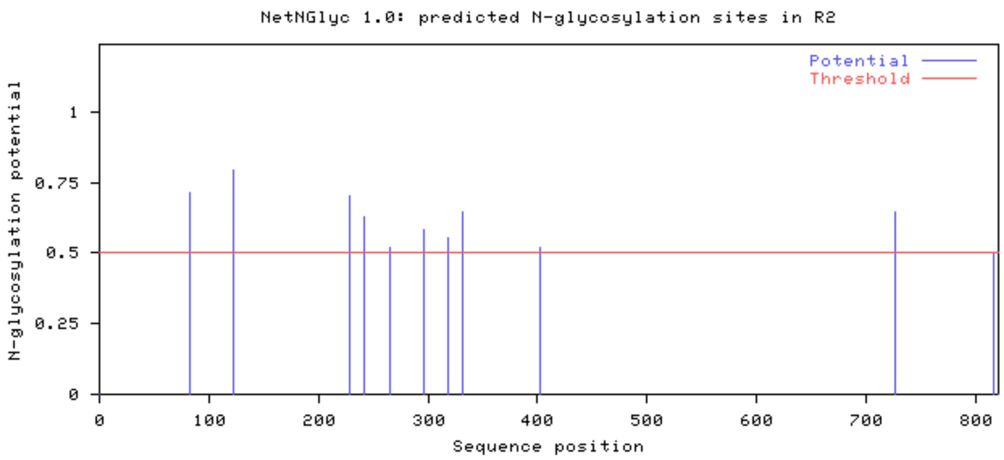
D.1 FGFR1

Protein	Position		Potential	Jury	N-Glyc	
R1	77	NRTR	0.7043	(9/9)	++	
R1	117	NVSD	0.7477	(9/9)	++	
R1	185	NPTL	0.7384	(9/9)	++	WARNING: PRO-X1.
R1	227	NYTC	0.7137	(9/9)	++	
R1	240	NHTY	0.6360	(7/9)	+	
R1	264	NKTV	0.6773	(8/9)	+	
R1	296	NGSK	0.7346	(9/9)	++	
R1	317	NTTD	0.6013	(6/9)	+	
R1	330	NVSF	0.6298	(7/9)	+	



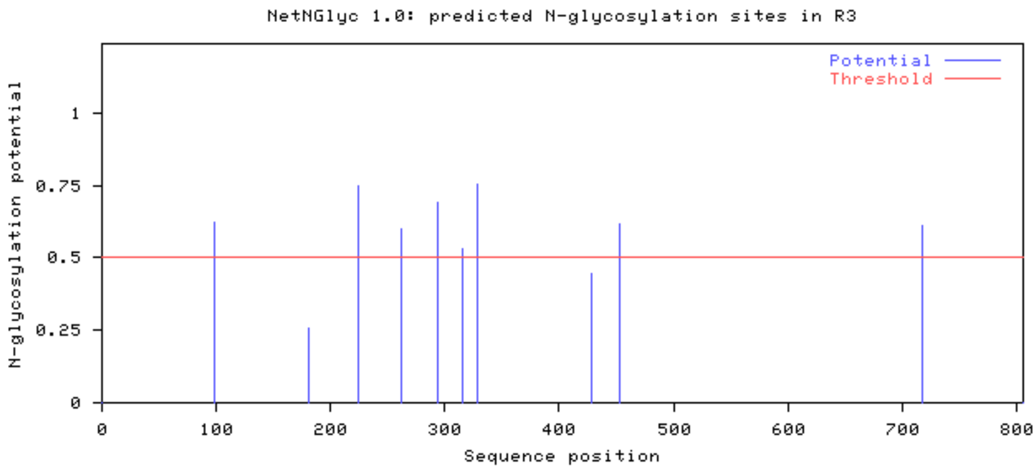
D.2 FGFR2

Protein	Position		Potential	Jury	N-Glyc
R2	83	NRTV	0.7158	(9/9)	++
R2	123	NVTD	0.7922	(9/9)	+++
R2	228	NYTC	0.7024	(9/9)	++
R2	241	NHTY	0.6288	(8/9)	+
R2	265	NAST	0.5183	(6/9)	+
R2	297	NGSK	0.5840	(6/9)	+
R2	318	NTTD	0.5520	(6/9)	+
R2	331	NVTF	0.6438	(8/9)	+
R2	402	NTTK	0.5212	(6/9)	+



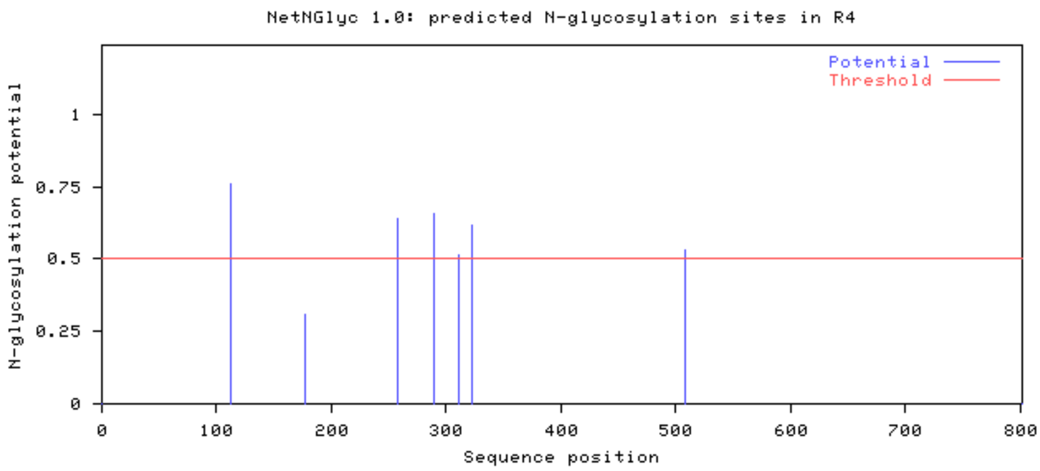
D.3 FGFR3

Protein	Position	Potential	Jury	N-Glyc
R3	98 NASH	0.6239	(7/9)	+
R3	181 NPTP	0.2596	(9/9)	---
R3	225 NYTC	0.7502	(9/9)	+++
R3	262 NQTA	0.5986	(7/9)	+
R3	294 NGSK	0.6898	(9/9)	++
R3	315 NTTD	0.5316	(7/9)	+
R3	328 NVTF	0.7571	(9/9)	+++
R3	428 NASM	0.4462	(6/9)	-
R3	453 NVSE	0.6166	(7/9)	+
R3	718 NCTH	0.6120	(6/9)	+



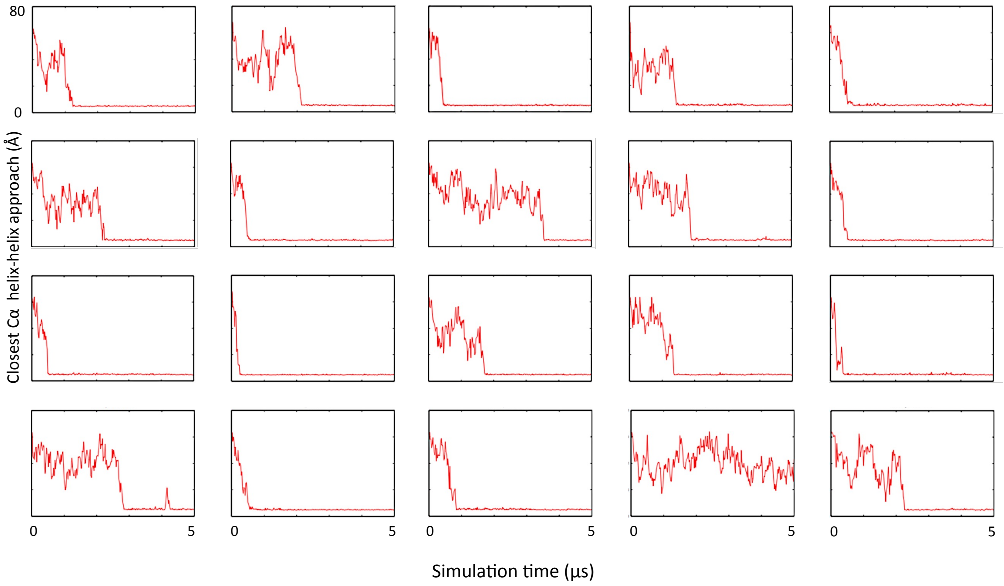
D.4 FGFR4

Protein	Position	Potential	Jury	N-Glyc
R4	112 NLTL	0.7628	(9/9)	+++
R4	177 NPTP	0.3104	(8/9)	--
R4	258 NTTA	0.6381	(8/9)	+
R4	290 NGSS	0.6580	(9/9)	++
R4	311 NSSE	0.5146	(6/9)	+
R4	322 NVSA	0.6156	(6/9)	+

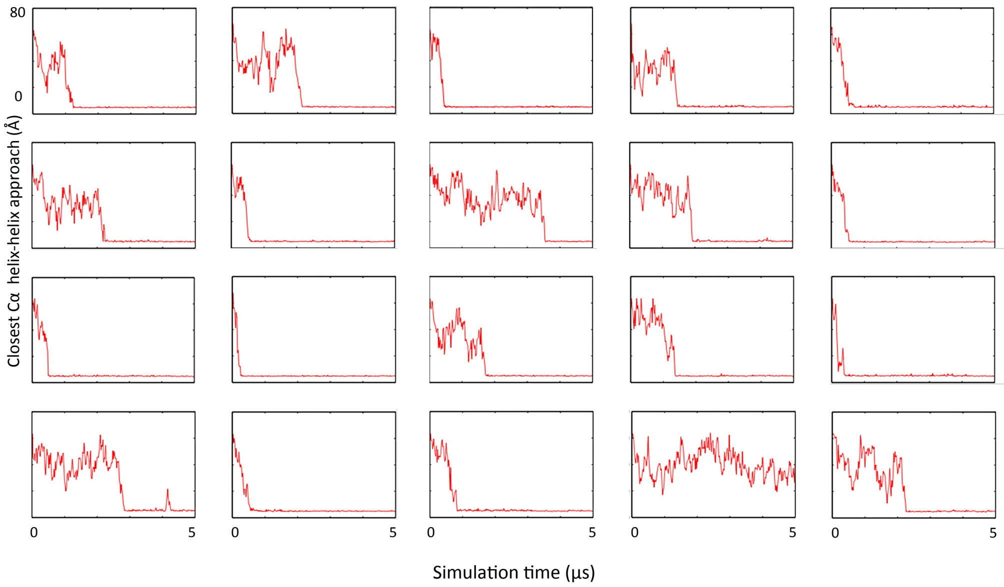


Appendix E – Cα closest interhelix distance plots

E.1 FGFR3 wt

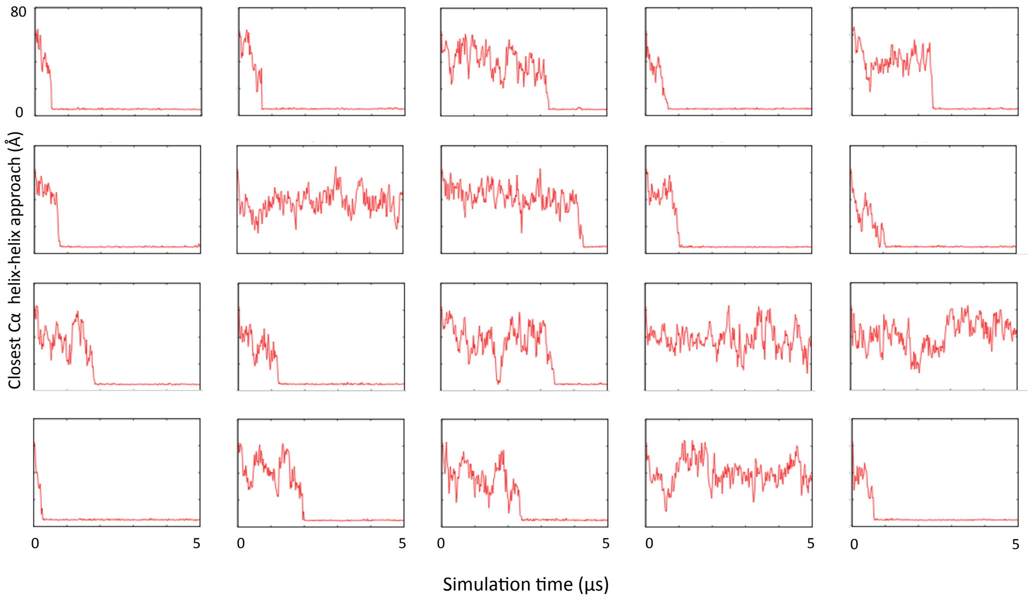


E.2 FGFR3 G380R

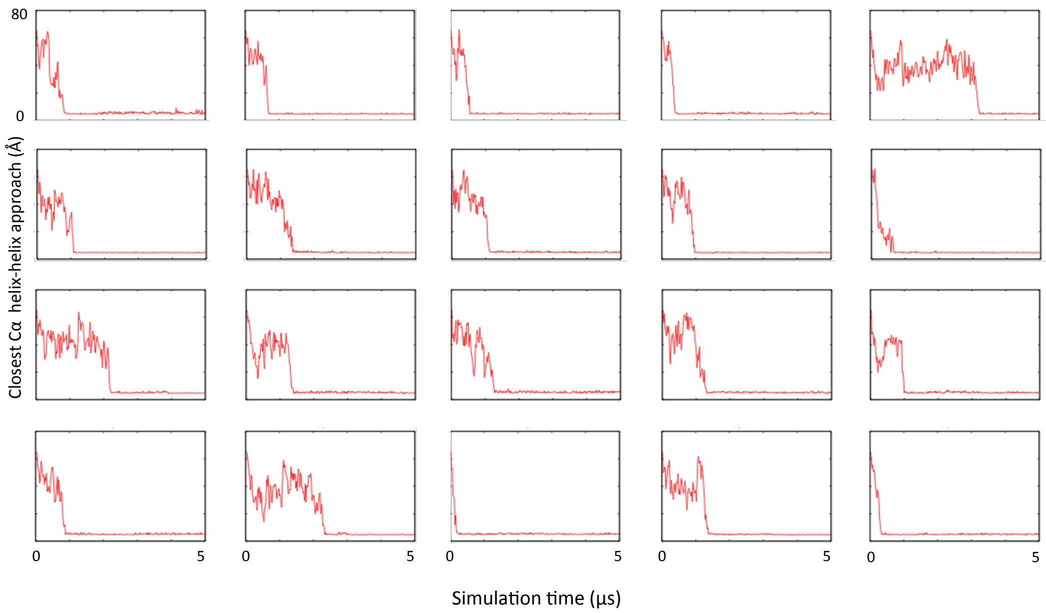


.

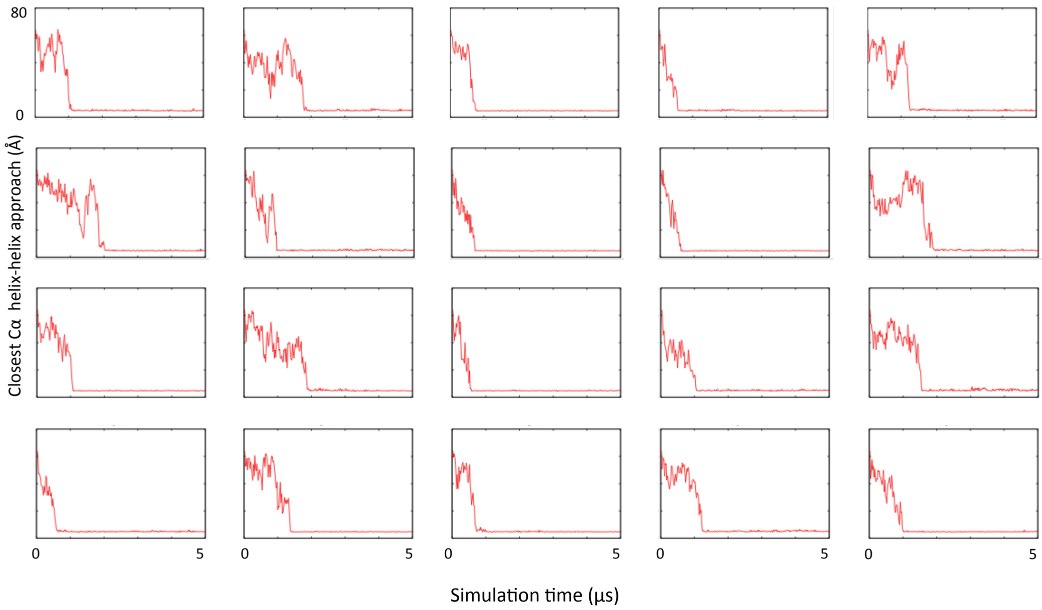
E.3 FGFR3 Y373C



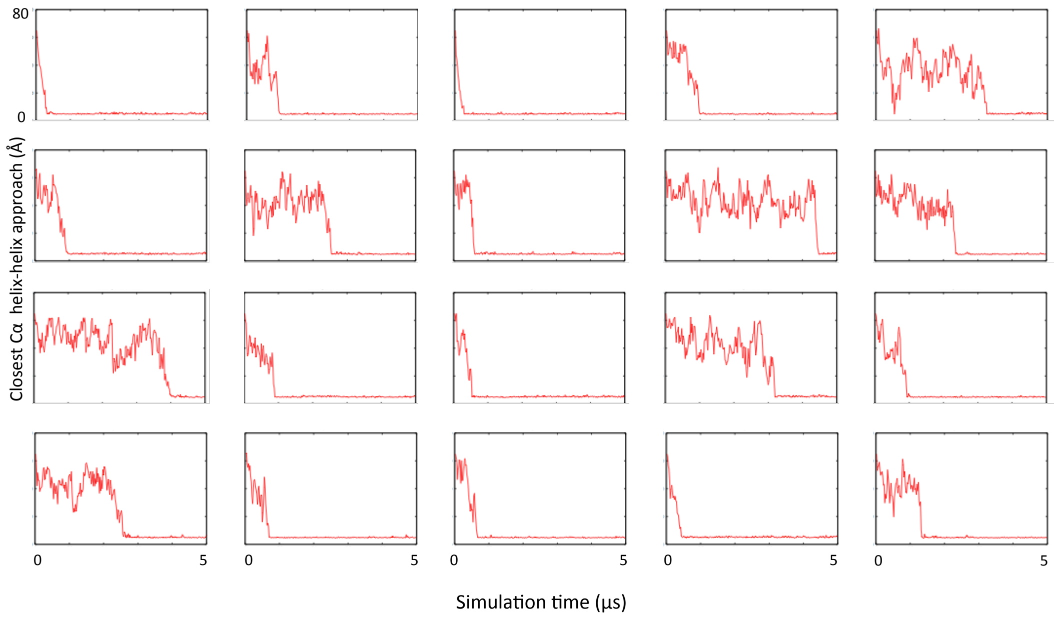
E4 FGFR4 wt



E.5 FGFR4 Y367C



E.6 FGFR4 G388R



Appendix F – List of Formulae

Formula 1

$$\text{Completeness} = \frac{\text{No. unique reflections observed}}{\text{No. unique reflections expected}} \text{ x } 100$$

Formula 2

$$\text{Redundancy} = \frac{\text{No. reflections observed}}{\text{No. unique reflections observed}}$$

Formula 3

$$\chi^2 = \sum_{hkl} \sum_{i} \left(\frac{I(hkl;i) - < I(hkl) >}{\sigma_{I(hkl)}} \right)^2$$

where $I(hkl;i)$ is the intensity of a single measurement i of the reflection hkl, $<I(hkl)>$ is the mean intensity from multiple measurements of the reflection hkl and $\sigma I_{I(hkl)}$ is the standard deviation for the measurement $I(hkl)$

Formula 4

$$R_{merge} = \frac{\sum_{hkl} \sum_{i} I(hkl;i) - < I(hkl) >}{\sum_{hkl} \sum_{i} I(hkl;i)}$$

where $I(hkl;i)$ is the intensity of a single measurement i of the reflection hkl and $<I(hkl)>$ is the mean intensity from multiple measurements of the reflection hkl

Formula 5

$$R_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i I(hkl;i) - < I(hkl) >}{\sum_{hkl}\sum_i I(hkl;i)}$$

As R_{merge} with an additional $\sqrt{\frac{1}{n-1}}$ term, where n is the redundancy of the data

Formula 6

$$I_{hkl} = |F_{hkl}|^2$$

where $F_{hkl} = \sum_j f_j \exp\left(2\pi i\left(hx_j + ky_j + lz_j\right)\right) \exp\left(-\frac{B_j sin^2\theta}{\lambda^2}\right)$ and f_j is the atomic scattering factor

Formula 7

$$R(\kappa, \phi, \psi) = \int_{F_{min}}^{F_{max}} P_{obs}(u) P_{model}(\kappa, \phi, \psi, u)\, du$$

where P_{obs} (u) and $P_{model}(\kappa, \phi, \psi, u)$ are the observed and model Patterson functions:

$P(u) = \sum_{hkl}|F_{hkl}|^2 \exp\left(-2\pi i(hx + ky + lz)\right)$

u represents a point (x, y, z) in 3D space for the Patterson function κ,ϕ,ψ are the rotation angles of the model solution

r_{max} is the mazimum intramolecular vector in the Patterson function

r_{min} is the minimum limit set to exclude the large origin peak in the Patterson function

Formula 8

$$C(t) = \frac{\sum\left(|F_{obs}|^2 - \overline{|F_{obs}|^2}\right)\left(|F_{calc}(t)|^2 - \overline{|F_{calc}(t)|^2}\right)}{\sqrt{\sum\left(|F_{obs}|^2 - \overline{|F_{obs}|^2}\right)\sum\left(|F_{calc}(t)|^2 - \overline{|F_{calc}(t)|^2}\right)}}$$

The correlation coefficient C(t) can be used to evaluate translation search solutions. This is a measure of agreement between observer and calculated squared structure factor amplitudes and is calculated for each trial solution *t* in the translation search.

Formula 9

$$Z - \text{score} = \frac{LLG_{solution} - < LLG_{random} >}{rmsd(LLG_{solution}, LLG_{random})}$$

Formula 10

$$R_{work} = \frac{\sum_{hkl}\left||F_{obs}| - k|F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

where $|F_{obs}|$ are the experimentally observed structure factor amplitudes and $|F_{calc}|$ are the structure factor amplitudes calculated from the model

Formula 11

$$R_{free} = \frac{\sum_{test\ set\ hkl}\left||F_{obs}| - k|F_{calc}|\right|}{\sum_{test\ set\ hkl}|F_{obs}|}$$

Calculated as for R_{work} with approximately the 5 % of data that has been excluded from the refinement process.

Appendix G – Abbreviations

AB	acidic box
AT	atomistic
BAD	BCL-2 antagonist of cell death
BMP	bone morphogenic protein
CG	coarse-grained
COM	centre of mass
Da	Dalton
DAG	diacylglycerol
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
DP	degree of polymerisation
E. coli	Escherichia coli
EGFR	epidermal growth factor receptor
EndoF1	endoglycosidase F1
Eph	ephrin receptor
FBS	fetal bovine serum
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
fl	flexible loop
FOXO	forkhead box class O
FRS2α	FGFR substrate 2α
GAB1	GRB2-associated binding protein 1
GlcNAc	N-acetyl-glucosamine
GnTI	N-acetylglucosaminyltransferase I
GPI	glycosyl phosphatidyl inositol
GRB2	growth factor receptor-bound 2
GROMACS	GROningen MAchine for Chemical Simulations
GTP	guanosine triphosphate
HEK	human embryonic kidney
hep	heparin
HPLC	high performance liquid chromatography
HS	heparan sulphate
HSPG	heparan sulphate proteoglycan
Ig	immunoglobulin
IMAC	immobilised metal affinity chromatography
InR	insulin receptor
IP_3	inositol-1,4,5,-trisphosphate
IPTG	isopropyl β-thiogalactopyranoside
ITC	isothermal titration calorimetry
JMD	juxtamembrane
LB	lysogeny broth
LLG	Log-likelihood gain
M	molar
Man	mannose
MAPK	mitogen-activated protein kinase
MARCKS	myristoylated alanine-rich C kinase substrate

MD	molecular dynamics
MPD	2-Methyl-2,4-Pentanediol
MW	molecular weight
N-CAM	neural cell-adhesion molecule
NFAT	nuclear factor of activated T
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PLCγ1	phospholipase Cγ1
POPC	palmitoyl oleoyl phosphatidyl choline
POPS	palmitoyl oleoyl phosphatidyl serine
RAS	rat sarcoma
RT	room temperature
RTK	receptor tyrosine kinase
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SNP	single-nucleotide polymorphism
SOS	son of sevenless
SPR	surface plasmon resonance
TDI	Thanatophoric Dysplasia Type I
TMD	transmembrane domain
wt	wild-type

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(Pt 2), 213–221.

Alegria-Schaffer, A., Lodge, A., & Vattem, K. (2009). Performing and optimizing Western blots with an emphasis on chemiluminescent detection. *Methods in Enzymology*, *463*, 573–599.

Aricescu, A. R., & Jones, E. Y. (2007). Immunoglobulin superfamily cell adhesion molecules: zippers and signals. *Current Opinion in Cell Biology*, *19*(5), 543–550.

Aricescu, A. R., Lu, W., & Jones, E. Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D: Biological Crystallography*, *62*(10), 1243–1250.

B Mulloy, M. J. Forster, C. Jones, D. B. Davies (1993). N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochemical Journal*, *293*(Pt 3), 849.

Bange, J., Prechtl, D., Cheburkin, Y., Specht, K., Harbeck, N., Schmitt, M., et al. (2002). Cancer progression and tumor cell motility are associated with the FGFR4 Arg(388) allele. *Cancer Research*, *62*(3), 840–847.

Bârzu, T., Lormeau, J. C., Petitou, M., Michelson, S., & Choay, J. (1989). Heparin-derived oligosaccharides: affinity for acidic fibroblast growth factor and effect on its growth-promoting activity for human endothelial cells - Google Search. *Journal of Cellular Physiology*, *140*(3), 538–548.

Beenken, A., & Mohammadi, M. (2009). The FGF family: biology, pathophysiology and therapy. *Nature Reviews Drug Discovery*, *8*(3), 235–253.

Beenken, A., Eliseenkova, A. V., Ibrahimi, O. A., Olsen, S. K., & Mohammadi, M. (2012). Plasticity in interactions of fibroblast growth factor 1 (FGF1) N terminus with FGF receptors underlies promiscuity of FGF1. *The Journal of Biological Chemistry*, *287*(5), 3067–3078.

Bell, C. A., Tynan, J. A., Hart, K. C., Meyer, A. N., Robertson, S. C., & Donoghue, D. J. (2000). Rotational coupling of the transmembrane and kinase domains of the Neu receptor tyrosine kinase. *Molecular Biology of the Cell*, *11*(10), 3589–3599.

Belov, A. A., & Mohammadi, M. (2013). Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology. *Cold Spring Harbor Perspectives in Biology*, *5*(6).

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690.

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, *129*(17), 5656–5664.

Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., & Zako, M. (1999). Functions of cell surface heparan sulfate proteoglycans. *Annual Review of Biochemistry*, *68*, 729–777.

Blencke, S., Zech, B., Engkvist, O., Greff, Z., Orfi, L., Horváth, Z., et al. (2004). Characterization of a conserved structural determinant controlling protein kinase sensitivity to selective inhibitors. *Chemistry & Biology*, *11*(5), 691–701.

Blume-Jensen, P., & Hunter, T. (2001). Oncogenic kinase signalling. *Nature*, *411*(6835), 355–365.

Bocharov, E. V., Lesovoy, D. M., Goncharuk, S. A., Goncharuk, M. V., Hristova, K., & Arseniev, A. S. (2013). Structure of FGFR3 Transmembrane Domain Dimer: Implications for Signaling and Human Pathologies. *Structure (London, England : 1993)*.

Bond, P. J., & Sansom, M. S. P. (2006). Insertion and assembly of membrane proteins via simulation. *Journal of the American Chemical Society*, *128*(8), 2697–2704.

Bond, P. J., Wee, C. L., & Sansom, M. S. P. (2008). Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry*, *47*(43), 11321–11331.

Botelho, A. V., Huber, T., Sakmar, T. P., & Brown, M. F. (2006). Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophysical Journal*, *91*(12), 4464–4477.

Böttcher, R. T., & Niehrs, C. (2005). Fibroblast growth factor signaling during early vertebrate development. *Endocrine Reviews*, *26*(1), 63–77.

Braun, P., & Heijne, von, G. (1999). The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. *Biochemistry*, *38*(30), 9778–9782.

Brown, A., Robinson, C. J., Gallagher, J. T., & Blundell, T. L. (2013). Cooperative Heparin-Mediated Oligomerization of Fibroblast Growth Factor-1 (FGF1) Precedes Recruitment of FGFR2 to Ternary Complexes. *Biophysical Journal*, *104*(8), 1720–1730.

Brunet, A., Bonni, A., Zigmond, M. J., Lin, M. Z., Juo, P., Hu, L. S., et al. (1999). Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell*, *96*(6), 857–868.

Byron, S. A., Chen, H., Wortmann, A., Loch, D., Gartside, M. G., Dehkhoda, F., et al. (2013). The N550K/H mutations in FGFR2 confer differential resistance to PD173074, dovitinib, and ponatinib ATP-competitive inhibitors. *Neoplasia (New York, NY)*, *15*(8), 975–988.

Carpenter, G., & Ji, Q.-S. (1999). Phospholipase C-γ as a Signal-Transducing Element. *Experimental Cell Research*, *253*(1), 15–24.

Carter, E. P., Fearon, A. E., & Grose, R. P. (2015). Careless talk costs lives: fibroblast growth factor receptor signalling and the consequences of pathway malfunction. *Trends in Cell Biology*, *25*(4), 221–233.

Cavallaro, U., Niedermeyer, J., Fuxa, M., & Christofori, G. (2001). N-CAM modulates tumour-cell adhesion to matrix by inducing FGF-receptor signalling. *Nature Cell Biology*, *3*(7), 650–657.

Cerutti, D. S., Duke, R. E., Darden, T. A., & Lybrand, T. P. (2009). Staggered Mesh Ewald: An extension of the Smooth Particle-Mesh Ewald method adding great versatility. *Journal of Chemical Theory and Computation*, *5*(9), 2322.

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., et al. (2007). Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure (London, England : 1993)*, *15*(3), 267–273.

Chen, H., Xu, C.-F., Ma, J., Eliseenkova, A. V., Li, W., Pollock, P. M., et al. (2008). A crystallographic snapshot of tyrosine trans-phosphorylation in action. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(50), 19660–19665.

Chen, L., Novicky, L., Merzlyakov, M., Hristov, T., & Hristova, K. (2010). Measuring the energetics of membrane protein dimerization in mammalian membranes. *Journal of the American Chemical Society*, *132*(10), 3628–3635.

Chen, P., Zhang, L., Weng, T., Zhang, S., Sun, S., Chang, M., et al. (2014). A Ser252Trp mutation in fibroblast growth factor receptor 2 (FGFR2) mimicking human Apert syndrome reveals an essential role for FGF signaling in the regulation of endochondral bone formation. *PloS One*, *9*(1), e87311.

Cho, J. Y., Guo, C., Torello, M., Lunstrum, G. P., Iwata, T., Deng, C., & Horton, W. A. (2004). Defective lysosomal targeting of activated fibroblast growth factor receptor 3 in achondroplasia. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(2), 609–614.

Crispin, M., Yu, X., & Bowden, T. A. (2013). Crystal structure of sialylated IgG Fc: Implications for the mechanism of intravenous immunoglobulin therapy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), E3544–E3546.

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Res*. 2004. 14: 1188-1190

d'Avis, P. Y., Robertson, S. C., Meyer, A. N., Bardwell, W. M., Webster, M. K., & Donoghue, D. J. (1998). Constitutive activation of fibroblast growth factor receptor 3 by mutations responsible for the lethal skeletal dysplasia thanatophoric dysplasia type I. *Cell Growth & Differentiation : the Molecular Biology Journal of the American Association for Cancer Research*, *9*(1), 71–78.

Datta, S. R., Dudek, H., Tao, X., Masters, S., Fu, H., Gotoh, Y., & Greenberg, M. E. (1997). Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell*, *91*(2), 231–241.

Davis, S. J., & van der Merwe, P. A. (2006). The kinetic-segregation model: TCR triggering and beyond. *Nature Immunology*, *7*(8), 803–809.

Deng, C. X., Wynshaw-Boris, A., Shen, M. M., Daugherty, C., Ornitz, D. M., & Leder, P. (1994). Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes & Development*, *8*(24), 3045–3057.

Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A., & Leder, P. (1996). Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell*, *84*(6), 911–921.

DiGabriele, A. D., Hendrickson, W. A., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., & Schlessinger, J. (1998). Structure of a heparin-linked biologically active dimer of fibroblast growth factor : Article : Nature. *Nature*, *393*(6687), 812–817.

Dorairaj, S., & Allen, T. W. (2007). On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 4943–4948.

Duchesne, L., Tissot, B., Rudd, T. R., Dell, A., & Fernig, D. G. (2006). N-glycosylation of fibroblast growth factor receptor 1 regulates ligand and heparan sulfate co-receptor binding. *The Journal of Biological Chemistry*, *281*(37), 27178–27189.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*(Pt 12 Pt 1), 2126–2132.

Endres, N. F., Das, R., Eastwood, M. P., Wemmer, D. E., Shaw, D. E., & Kuriyan, J. (2013). Architecture and Membrane Interactions of the EGF Receptor. *Cell*, *152*(3), 557–569.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., et al. (2007). Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [Et Al.]*, *Chapter 2*, Unit 2.9.

Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J., & Rees, D. C. (1996). Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science (New York, NY)*, *271*(5252), 1116–1120.

Felix, J., Elegheert, J., Gutsche, I., Shkumatov, A. V., Wen, Y., Bracke, N., et al. (2013). Human IL-34 and CSF-1 establish structurally similar extracellular assemblies with their common hematopoietic receptor. *Structure (London, England : 1993)*, *21*(4), 528–539.

Feyerabend, T. B., Li, J.-P., Lindahl, U., & Rodewald, H.-R. (2006). Heparan sulfate C5-epimerase is essential for heparin biosynthesis in mast cells. *Nature Chemical Biology*, *2*(4), 195–196.

Finger, C., Escher, C., & Schneider, D. (2009). The single transmembrane domains of human receptor tyrosine kinases encode self-interactions. *Science Signaling*, *2*(89), ra56.

Fleishman, S. J., Schlessinger, J., & Ben-Tal, N. (2002). A putative molecular-activation switch in the transmembrane domain of erbB2. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), 15937–15940.

Franco, R., Canals, M., Marcellino, D., Ferré, S., Agnati, L., Mallol, J., et al. (2003). Regulation of heptaspanning-membrane-receptor function by dimerization and clustering. *Trends in Biochemical Sciences*, *28*(5), 238–243.

Franke, D., & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initioshape determination in small-angle scattering. *Journal of Applied Crystallography*, *42*(2), 342–346.

Franklin J Moy, Michal Safran, Andrew P Seddon, Doug Kitchen, Peter Böhlen, David Aviezer, et al. (1997). Properly Oriented Heparin−Decasaccharide-Induced Dimers Are the Biologically Active Form of Basic Fibroblast Growth Factor. *Pubs.Acs.org*.

French, S., & Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallographica Section A*, *34*(4), 517–525.

Gallagher, J. T. (2001). Heparan sulfate: growth control with a restricted sequence menu. *The Journal of Clinical Investigation*, *108*(3), 357–361.

Gallo, L. H., Nelson, K. N., Meyer, A. N., & Donoghue, D. J. (2015). Functions of Fibroblast Growth Factor Receptors in cancer defined by novel translocations and mutations. *Cytokine & Growth Factor Reviews*.

Galvin, B. D., Hart, K. C., Meyer, A. N., Webster, M. K., & Donoghue, D. J. (1996). Constitutive receptor activation by Crouzon syndrome mutations in fibroblast growth factor receptor (FGFR)2 and FGFR2/Neu chimeras. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(15), 7894–7899.

Gamper, N., & Shapiro, M. S. (2007). Regulation of ion transport proteins by membrane phosphoinositides. *Nature Reviews Neuroscience*, *8*(12), 921–934.

Garman, E. F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D: Biological Crystallography*, *66*(Pt 4), 339–351.

George Hedger, M. S. P. Sansom, H. Koldsø (2015). The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Scientific Reports*, *5*. 9198, doi:10.1038/srep09198

Goetz, R., & Mohammadi, M. (2013). Exploring mechanisms of FGF signalling through the lens of structural biology. *Nature Reviews Molecular Cell Biology*. 14(3):166-80. doi: 10.1038/nrm3528

Goetz, R., Beenken, A., Ibrahimi, O. A., Kalinina, J., Olsen, S. K., Eliseenkova, A. V., et al. (2007). Molecular insights into the klotho-dependent, endocrine mode of action of fibroblast growth factor 19 subfamily members. *Molecular and Cellular Biology*, *27*(9), 3417–3428.

Goetz, R., Nakada, Y., Hu, M.-C., Kurosu, H., Wang, L., Nakatani, T., et al. (2010). Isolated C-terminal tail of FGF23 alleviates hypophosphatemia by inhibiting FGF23-FGFR-Klotho complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 407–412.

Goodger, S. J., Robinson, C. J., Murphy, K. J., Gasiunas, N., Harmer, N. J., Blundell, T. L., et al. (2008). Evidence that heparin saccharides promote FGF2 mitogenesis through two distinct mechanisms. *The Journal of Biological Chemistry*, *283*(19), 13001–13008.

Gotoh, N. (2008). Regulation of growth factor signaling by FRS2 family docking/scaffold adaptor proteins. *Cancer Science*, *99*(7), 1319–1325.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, *446*(7132), 153–158.

Grigoryan, G., Moore, D. T., & DeGrado, W. F. (2011). Transmembrane Communication: General Principles and Lessons from the Structure and Function of the M2 Proton Channel, K+ Channels, and Integrin Receptors. *Dx.Doi.org*, *80*(1), 211–237.

Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B., & Chène, C. (1996). Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Protein Science : Publication of the Protein Society*, *5*(12), 2617–2622.

Gupta, R., & Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 310–322.

Guttman, M., Weinkam, P., Sali, A., & Lee, K. K. (2013). All-Atom Ensemble Modeling to Analyze Small-Angle X-Ray Scattering of Glycosylated Proteins. *Structure (London, England : 1993)*, *21*(3), 321–331.

Hall, B. A., Halim, K. B. A., Buyan, A., Emmanouil, B., & Sansom, M. S. P. (2014). Sidekick for Membrane Simulations: Automated Ensemble Molecular Dynamics Simulations of Transmembrane Helices. *Journal of Chemical Theory and Computation*, *10*(5), 2165–2175.

Han, X., Mihailescu, M., & Hristova, K. (2006). Neutron diffraction studies of fluid bilayers with transmembrane proteins: structural consequences of the achondroplasia mutation. *Biophysical Journal*, *91*(10), 3736–3747.

Harada, M., Murakami, H., Okawa, A., Okimoto, N., Hiraoka, S., Nakahara, T., et al. (2009). FGF9 monomer-dimer equilibrium regulates extracellular matrix affinity and tissue diffusion. *Nature Genetics*, *41*(3), 289–298.

Harmer, N. J., Ilag, L. L., Mulloy, B., Pellegrini, L., Robinson, C. V., & Blundell, T. L. (2004). Towards a resolution of the stoichiometry of the fibroblast growth factor (FGF)-FGF receptor-heparin complex. *Journal of Molecular Biology*, *339*(4), 821–834.

Harmer, N. J., Robinson, C. J., Adam, L. E., Ilag, L. L., Robinson, C. V., Gallagher, J. T., & Blundell, T. L. (2006). Multimers of the fibroblast growth factor (FGF)-FGF receptor-saccharide complex are formed on long oligomers of heparin. *The Biochemical Journal*, *393*(Pt 3), 741–748.

Hartwig, J. H., Thelen, M., Rosen, A., Janmey, P. A., Nairn, A. C., & Aderem, A. (1992). MARCKS is an actin filament crosslinking protein regulated by protein kinase C and calcium-calmodulin.

He, L., Horton, W., & Hristova, K. (2010). Physical basis behind achondroplasia, the most common form of human dwarfism. *The Journal of Biological Chemistry*, *285*(39), 30103–30114.

He, L., Shobnam, N., Wimley, W. C., & Hristova, K. (2011). FGFR3 heterodimerization in achondroplasia, the most common form of human dwarfism. *The Journal of Biological Chemistry*, *286*(15), 13272–13281.

Healey, E. G., Bishop, B., Elegheert, J., Bell, C. H., Padilla-Parra, S., & Siebold, C. (2015). Repulsive guidance molecule is a structural bridge between neogenin and bone morphogenetic protein. *Nature Structural & Molecular Biology*, *22*(6), 458–465.

Herr, A. B., Ornitz, D. M., Sasisekharan, R., Venkataraman, G., & Waksman, G. (1997). Heparin-induced Self-association of Fibroblast Growth Factor-2. Retrieved from

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472.

Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447.

Holzmann, K., Grunt, T., Heinzle, C., Sampl, S., Steinhoff, H., Reichmann, N., et al. (2012). Alternative Splicing of Fibroblast Growth Factor Receptor IgIII Loops in Cancer. *Journal of Nucleic Acids*, *2012*(15), 1–12.

Hubbard, S. R. (2004). Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nature Reviews Molecular Cell Biology*, *5*(6), 464–471.

Hubbard, S. R. (2009). The juxtamembrane region of EGFR takes center stage. *Cell*, *137*(7), 1181–1183.

Hubert, P., Sawma, P., Duneau, J.-P., Khao, J., Hénin, J., Bagnard, D., & Sturgis, J. (2010). Single-spanning transmembrane domains in cell growth and cell-cell interactions: More than meets the eye? *Cell Adhesion & Migration*, *4*(2), 313–324.

Ibrahimi, O. A., Eliseenkova, A. V., Plotnikov, A. N., Yu, K., Ornitz, D. M., & Mohammadi, M. (2001). Structural basis for fibroblast growth factor receptor 2 activation in Apert syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(13), 7182–7187.

Ibrahimi, O. A., Yeh, B. K., Eliseenkova, A. V., Zhang, F., Olsen, S. K., Igarashi, M., et al. (2005). Analysis of mutations in fibroblast growth factor (FGF) and a pathogenic mutation in FGF receptor (FGFR) provides direct evidence for the symmetric two-end model for FGFR dimerization. *Molecular and Cellular Biology*, *25*(2), 671–684.

Ibrahimi, O. A., Zhang, F., Eliseenkova, A. V., Itoh, N., Linhardt, R. J., & Mohammadi, M. (2004). Biochemical analysis of pathogenic ligand-dependent FGFR2 mutations suggests distinct pathophysiological mechanisms for craniofacial and limb abnormalities. *Human Molecular Genetics*, *13*(19), 2313–2324.

Ishihara, M. (1994). Structural requirements in heparin for binding and activation of FGF-1 and FGF-4 are different from that for FGF-2. *Glycobiology*, *4*(6), 817–824.

Janosi, L., Prakash, A., & Doxastakis, M. (2010). Lipid-modulated sequence-specific association of glycophorin A in membranes. *Biophysical Journal*, *99*(1), 284–292.

Jefferis, R. (2012). Isotype and glycoform selection for antibody therapeutics. *Archives of Biochemistry and Biophysics*, *526*(2), 159–166.

Jonker, J. W., Suh, J. M., Atkins, A. R., Ahmadian, M., Li, P., Whyte, J., et al. (2012). A PPARγ-FGF1 axis is required for adaptive adipose remodelling and metabolic homeostasis. *Nature*, *485*(7398), 391–394.

Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., et al. (2009). Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell*, *137*(7), 1293–1307.

Juretić, D., Zoranić, L., & Zucić, D. (2002). Basic charge clusters and predictions of membrane protein topology. *Journal of Chemical Information and Computer Sciences*, *42*(3), 620–632.

Kalinina, J., Byron, S. A., Makarenkova, H. P., Olsen, S. K., Eliseenkova, A. V., Larochelle, W. J., et al. (2009). Homodimerization controls the fibroblast growth factor 9 subfamily's receptor binding and heparan sulfate-dependent diffusion in the extracellular matrix. *Molecular and Cellular Biology*, *29*(17), 4663–4678.

Kalinina, J., Dutta, K., Ilghari, D., Beenken, A., Goetz, R., Eliseenkova, A. V., et al. (2012). The Alternatively Spliced Acid Box Region Plays a Key Role in FGF Receptor Autoinhibition. *Structure (London, England : 1993)*, *20*(1), 77–88.

Kalli, A. C., Hall, B. A., Campbell, I. D., & Sansom, M. S. P. (2011). A helix heterodimer in a lipid bilayer: prediction of the structure of an integrin transmembrane domain via multiscale simulations. *Structure (London, England : 1993)*, *19*(10), 1477–1484.

Kalli, A. C., Wegener, K. L., Goult, B. T., Anthis, N. J., Campbell, I. D., & Sansom, M. S. P. (2010). The structure of the talin/integrin complex at a lipid bilayer: an NMR and MD simulation study. *Structure (London, England : 1993)*, *18*(10), 1280–1288.

Kaszuba, K., Grzybek, M., Orłowski, A., Danne, R., Róg, T., Simons, K., et al. (2015). N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4334–4339.

Killian, J. A., & Heijne, von, G. (2000). How proteins adapt to a membrane-water interface. *Trends in Biochemical Sciences*, *25*(9), 429–434.

Kim, C., Schmidt, T., Cho, E.-G., Ye, F., Ulmer, T. S., & Ginsberg, M. H. (2012). Basic amino-acid side chains regulate transmembrane integrin signalling. *Nature*, *481*(7380), 209–213.

Kouhara, H., Hadari, Y. R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I., & Schlessinger, J. (1997). A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell*, *89*(5), 693–702.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*(3), 774–797.

Lamothe, B., Yamada, M., Schaeper, U., Birchmeier, W., Lax, I., & Schlessinger, J. (2004). The docking protein Gab1 is an essential component of an indirect mechanism for fibroblast growth factor stimulation of the phosphatidylinositol 3-kinase/Akt antiapoptotic pathway. *Molecular and Cellular Biology*, *24*(13), 5657–5666.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948.

Lemmon, M. A., & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell*, *141*(7), 1117–1134.

Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brünger, A. T., & Engelman, D. M. (1994). A dimerization motif for transmembrane α–helices. *Nature Structural Biology*, *1*(3), 157–163.

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica Section D: Biological Crystallography*, *62*(Pt 1), 48–57.

Li, D., Lee, J., & Caffrey, M. (2011). Crystallizing Membrane Proteins in Lipidic Mesophases. A Host Lipid Screen. *Crystal Growth & Design*, *11*(2), 530–537.

Li, E., You, M., & Hristova, K. (2005). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis and forster resonance energy transfer suggest weak interactions between fibroblast growth factor receptor 3 (FGFR3) transmembrane domains in the absence of extracellular domains and ligands. *Biochemistry*, *44*(1), 352–360.

Li, E., You, M., & Hristova, K. (2006). FGFR3 dimer stabilization due to a single amino acid pathogenic mutation. *Journal of Molecular Biology*, *356*(3), 600–612.

Liao, R. G., Jung, J., Tchaicha, J., Wilkerson, M. D., Sivachenko, A., Beauchamp, E. M., et al. (2013). Inhibitor-sensitive FGFR2 and FGFR3 mutations in lung squamous cell carcinoma. *Cancer Research*, *73*(16), 5195–5205.

Livingstone, C. D., & Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer Applications in the Biosciences : CABIOS*, *9*(6), 745–756.

MacKerell, A. D., Banavali, N., & Foloppe, N. (2001). Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, *56*(4), 257–265.

Maître, J.-L., & Heisenberg, C.-P. (2013). Three Functions of Cadherins in Cell Adhesion. *Current Biology*, *23*(14), R626–R633.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & de Vries, A. H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, *111*(27), 7812–7824.

Martínez-Frías, M. L., de Frutos, C. A., Bermejo, E., Nieto, M. A., ECEMC Working Group. (2010). Review of the recently defined molecular mechanisms underlying thanatophoric dysplasia and their potential therapeutic implications for achondroplasia. *American Journal of Medical Genetics. Part A*, *152A*(1), 245–255.

Matsushita, M., Kitoh, H., Ohkawara, B., Mishima, K., Kaneko, H., Ito, M., et al. (2013). Meclozine Facilitates Proliferation and Differentiation of Chondrocytes by Attenuating Abnormally Activated FGFR3 Signaling in Achondroplasia. *PloS One*, *8*(12), e81569.

Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, *33*(2), 491–497.

Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–

674.
Meyers, G. A., Orlow, S. J., Munro, I. R., Przylepa, K. A., & Jabs, E. W. (1995). Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation in Crouzon syndrome with acanthosis nigricans. *Nature Genetics*, *11*(4), 462–464.
Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*.
Miller, P. S., & Aricescu, A. R. (2014). Crystal structure of a human GABAA receptor. *Nature*.
Mohammadi, M., Dikic, I., Sorokin, A., Burgess, W. H., Jaye, M., & Schlessinger, J. (1996). Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Molecular and Cellular Biology*, *16*(3), 977–989.
Mohammadi, M., Olsen, S. K., & Goetz, R. (2005). A protein canyon in the FGF-FGF receptor dimer selects from an à la carte menu of heparan sulfate motifs. *Current Opinion in Structural Biology*, *15*(5), 506–516.
Monsonego-Ornan, E., Adar, R., Feferman, T., Segev, O., & Yayon, A. (2000). The transmembrane mutation G380R in fibroblast growth factor receptor 3 uncouples ligand-mediated receptor activation from down-regulation. *Molecular and Cellular Biology*, *20*(2), 516–522.
Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., & Marrink, S.-J. (2008). The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, *4*(5), 819–834.
Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D: Biological Crystallography*, *53*(3), 240–255.
Nagata, M., Nuckolls, G. H., Wang, X., Shum, L., Seki, Y., Kawase, T., et al. (2011). The primary site of the acrocephalic feature in Apert Syndrome is a dwarf cranial base with accelerated chondrocytic differentiation due to aberrant activation of the FGFR2 signaling. *Bone*, *48*(4), 847–856.
Naski, M. C., Wang, Q., Xu, J., & Ornitz, D. M. (1996). Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nature Genetics*, *13*(2), 233–237.
Oberklaid, F., Danks, D. M., Jensen, F., Stace, L., & Rosshandler, S. (1979). Achondroplasia and hypochondroplasia. Comments on frequency, mutation rate, and radiological features in skull and spine. *Journal of Medical Genetics*, *16*(2), 140–146.
Ohi, M., Li, Y., Cheng, Y., & Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biological Procedures Online*, *6*, 23–34.
Olsen, S. K., Ibrahimi, O. A., Raucci, A., Zhang, F., Eliseenkova, A. V., Yayon, A., et al. (2004). Insights into the molecular basis for fibroblast growth factor receptor autoinhibition and ligand-binding promiscuity. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(4), 935–940.
Olsen, S. K., Li, J. Y. H., Bromleigh, C., Eliseenkova, A. V., Ibrahimi, O. A., Lao, Z., et al. (2006). Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain. *Genes & Development*, *20*(2), 185–198.
Ornitz, D. M., & Itoh, N. (2001). Fibroblast growth factors. *Genome Biology*, *2*(3), reviews3005.1.
Ornitz, D. M., Herr, A. B., Nilsson, M., Westman, J., Svahn, C. M., & Waksman, G. (1995). FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science (New York, NY)*, *268*(5209), 432–436.

Ornitz, D. M., Xu, J., Colvin, J. S., McEwen, D. G., MacArthur, C. A., Coulier, F., et al. (1996). Receptor specificity of the fibroblast growth factor family. *The Journal of Biological Chemistry*, *271*(25), 15292–15297.

Ornitz, D. M., Yayon, A., Flanagan, J. G., Svahn, C. M., Levi, E., & Leder, P. (1992). Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells. *Molecular and Cellular Biology*, *12*(1), 240–247.

Orr-Urtreger, A., Bedford, M. T., Burakova, T., Arman, E., Zimmer, Y., Yayon, A., et al. (1993). Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2). *Developmental Biology*, *158*(2), 475–486.

Padilla-Parra, S., Audugé, N., Tramier, M., & Coppey-Moisan, M. (2015). Time-Domain Fluorescence Lifetime Imaging Microscopy: A Quantitative Method to Follow Transient Protein–Protein Interactions in Living Cells. *Cshprotocols.Cshlp.org*.

Parrinello, M. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182–7190.

Pellegrini, L. (2001). Role of heparan sulfate in fibroblast growth factor signalling: a structural view. *Current Opinion in Structural Biology*, *11*(5), 629–634.

Pellegrini, L., Burke, D. F., Delft, von, F., Mulloy, B., & Blundell, T. L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature*, *407*(6807), 1029–1034.

Peng, W. C., Lin, X., & Torres, J. (2009). The strong dimerization of the transmembrane domain of the fibroblast growth factor receptor (FGFR) is modulated by C-terminal juxtamembrane residues. *Protein Science : a Publication of the Protein Society*, *18*(2), 450–459.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of Synchrotron Radiation*, *20*(Pt 4), 660–664.

Perrimon, N., & Bernfield, M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. *Nature*, *404*(6779), 725–728.

Peters, K., Werner, S., Liao, X., Wert, S., Whitsett, J., & Williams, L. (1994). Targeted expression of a dominant negative FGF receptor blocks branching morphogenesis and epithelial differentiation of the mouse lung. *The EMBO Journal*, *13*(14), 3296–3301.

Petersen, T. N., Brunak, S., Heijne, von, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786.

Petoukhov, M. V., & Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Current Opinion in Structural Biology*, *17*(5), 562–571.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Placone, J., & Hristova, K. (2012). Direct Assessment of the Effect of the Gly380Arg Achondroplasia Mutation on FGFR3 Dimerization Using Quantitative Imaging FRET. *PloS One*, *7*(10), e46678.

Platonova, N., Miquel, G., Chiu, L.-Y., Taouji, S., Moroni, E., Colombo, G., et al. (2014). Dimerization Capacities of FGF2 Purified with or without Heparin-Affinity Chromatography. *PloS One*, *9*(10), e110055.

Plotnikov, A. N., Hubbard, S. R., Schlessinger, J., & Mohammadi, M. (2000). Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell*, *101*(4), 413–424.

Plotnikov, A. N., Schlessinger, J., Hubbard, S. R., & Mohammadi, M. (1999). Structural basis for FGF receptor dimerization and activation. *Cell*, *98*(5), 641–650.

Polanska, U. M., Duchesne, L., Harries, J. C., Fernig, D. G., & Kinnunen, T. K. (2009). N-Glycosylation regulates fibroblast growth factor receptor/EGL-15 activity in Caenorhabditis elegans in vivo. *The Journal of Biological Chemistry*, *284*(48), 33030–33039.

Powell, A. K., Fernig, D. G., & Turnbull, J. E. (2002). Fibroblast growth factor receptors 1 and 2 interact differently with heparin/heparan sulfate. Implications for dynamic assembly of a ternary signaling complex. *The Journal of Biological Chemistry*, *277*(32), 28554–28563.

Prakash, A., Janosi, L., & Doxastakis, M. (2011). GxxxG motifs, phenylalanine, and cholesterol guide the self-association of transmembrane domains of ErbB2 receptors. *Biophysical Journal*, *101*(8), 1949–1958.

Psachoulia, E., Fowler, P. W., Bond, P. J., & Sansom, M. S. P. (2008). Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization. *Biochemistry*, *47*(40), 10503–10512.

Pye, D. A., Vives, R. R., Turnbull, J. E., Hyde, P., & Gallagher, J. T. (1998). Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *The Journal of Biological Chemistry*, *273*(36), 22936–22942.

Rahmoune, H., Chen, H. L., Gallagher, J. T., Rudland, P. S., & Fernig, D. G. (1998). Interaction of heparan sulfate from mammary cells with acidic fibroblast growth factor (FGF) and basic FGF. Regulation of the activity of basic FGF by high and low affinity binding sites in heparan sulfate. *The Journal of Biological Chemistry*, *273*(13), 7303–7310.

Rassouli, H., Tabe Bordbar, M. S., Rezaei Larijani, M., Pakzad, M., Baharvand, H., & Salekdeh, G. H. (2013). Cloning, expression and functional characterization of in-house prepared human basic fibroblast growth factor. *Cell Journal*, *14*(4), 282–291.

Reardon, W., Winter, R. M., Rutland, P., Pulleyn, L. J., Jones, B. M., & Malcolm, S. (1994). Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. *Nature Genetics*, *8*(1), 98–103.

Red Brewer, M., Choi, S. H., Alvarado, D., Moravcevic, K., Pozzi, A., Lemmon, M. A., & Carpenter, G. (2009). The Juxtamembrane Region of the EGF Receptor Functions as an Activation Domain. *Molecular Cell*, *34*(6), 641–651.

Reddy, T. (2012). Structure, Flexibility, And Overall Motion Of Transmembrane Peptides Studied By NMR Spectroscopy And Molecular Dynamics Simulations, 1–515.

Reddy, T., Manrique, S., Buyan, A., Hall, B. A., Chetwynd, A., & Sansom, M. S. P. (2014). Primary and Secondary Dimer Interfaces of the Fibroblast Growth Factor Receptor 3 Transmembrane Domain: Characterization via Multiscale Molecular Dynamics Simulations. *Biochemistry*, *53*(2), 323–332.

Renhowe, P. A., Pecchi, S., Shafer, C. M., Machajewski, T. D., Jazan, E. M., Taylor, C., et al. (2009). Design, structure-activity relationships and in vivo characterization of 4-amino-3-benzimidazol-2-ylhydroquinolin-2-ones: a novel class of receptor tyrosine kinase inhibitors. *Journal of Medicinal Chemistry*, *52*(2), 278–292.

Robertson, S. C., Tynan, J. A., & Donoghue, D. J. (2000). RTK mutations and human syndromeswhen good receptors turn bad. *Trends in Genetics : TIG*, *16*(6), 265–271.

Robinson, C. J., Harmer, N. J., Goodger, S. J., Blundell, T. L., & Gallagher, J. T. (2005). Cooperative dimerization of fibroblast growth factor 1 (FGF1) upon a single heparin saccharide may drive the formation of 2:2:1 FGF1.FGFR2c.heparin ternary complexes. *The Journal of Biological Chemistry*, *280*(51), 42274–42282.

Roidl, A., Foo, P., Wong, W., Mann, C., Bechtold, S., Berger, H. J., et al. (2010). The FGFR4 Y367C mutant is a dominant oncogene in MDA-MB453 breast cancer cells. *Oncogene*, *29*(10), 1543–1552.

Rosenbaum, D. M., Rasmussen, S. G. F., & Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature*, *459*(7245), 356–363.

Rossmann, M. G. (2001). Molecular replacement – historical background. *Acta Crystallographica Section D: Biological Crystallography*, *57*(10), 1360–1366.
Rossmann, M. G., & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, *15*(1), 24–31.
Rousseau, F., Ghouzzi, el, V., Delezoide, A. L., Legeai-Mallet, L., Le Merrer, M., Munnich, A., & Bonaventure, J. (1996). Missense FGFR3 mutations create cysteine residues in thanatophoric dwarfism type I (TD1). *Human Molecular Genetics*, *5*(4), 509–512.
Russ, W. P., & Engelman, D. M. (1999). TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(3), 863–868.
Russ, W. P., & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *Journal of Molecular Biology*, *296*(3), 911–919.
Sacchettini, J. C., Baum, L. G., & Brewer, C. F. (2001). Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry*, *40*(10), 3009–3015.
Sambrook, J., & Russell, D. W. (2000). SDS-Polyacrylamide Gel Electrophoresis of Proteins. *Cshprotocols.Cshlp.org*.
Sankararamakrishnan, R., & Weinstein, H. (2000). Molecular dynamics simulations predict a tilted orientation for the helical region of dynorphin A(1-17) in dimyristoylphosphatidylcholine bilayers. *Biophysical Journal*, *79*(5), 2331–2344.
Sarabipour, S., & Hristova, K. (2013). FGFR3 Transmembrane Domain Interactions Persist in the Presence of Its Extracellular Domain. *Biophysical Journal*, *105*(1), 165–171.
Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., et al. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Molecular Cell*, *6*(3), 743–750.
Schneider, D., & Engelman, D. M. (2004). Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *Journal of Molecular Biology*, *343*(4), 799–804.
Schneidman-Duhovny, D., Kim, S. J., & Sali, A. (2012). Integrative structural modeling with small angle X-ray scattering profiles. *BMC Structural Biology*, *12*, 17.
Segrest, J. P., De Loof, H., Dohlman, J. G., Brouillette, C. G., & Anantharamaiah, G. M. (1990). Amphipathic helix motif: classes and properties. *Proteins*, *8*(2), 103–117.
Seiradake, E., Schaupp, A., Del Toro Ruiz, D., Kaufmann, R., Mitakidis, N., Harlos, K., et al. (2013). Structurally encoded intraclass differences in EphA clusters drive distinct cell responses. *Nature Structural & Molecular Biology*, *20*(8), 958–964.
Semenyuk, A. V., & Svergun, D. I. (1991). GNOM – a program package for small-angle scattering data processing. *Journal of Applied Crystallography*, *24*(5), 537–540.
Sengupta, D., & Marrink, S. J. (2010). Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. *Physical Chemistry Chemical Physics : PCCP*, *12*(40), 12987–12996.
Shiang, R., Thompson, L. M., Zhu, Y. Z., Church, D. M., Fielder, T. J., Bocian, M., et al. (1994). Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell*, *78*(2), 335–342.
Sierra, J. R., Cepero, V., & Giordano, S. (2010). Molecular mechanisms of acquired resistance to tyrosine kinase targeted therapy. *Molecular Cancer*, *9*, 75.
Sleeman, M., Fraser, J., McDonald, M., Yuan, S., White, D., Grandison, P., et al. (2001). Identification of a new fibroblast growth factor receptor, FGFR5. *Gene*, *271*(2), 171–182.
Snider, C., Jayasinghe, S., Hristova, K., & White, S. H. (2009). MPEx: a tool for exploring membrane proteins. *Protein Science : a Publication of the Protein Society*, *18*(12), 2624–

2628.
Sommer, A., & Rifkin, D. B. (1989). Interaction of heparin with human basic fibroblast growth factor: Protection of the angiogenic protein from proteolytic degradation by a glycosaminoglycan. *Journal of Cellular Physiology*, *138*(1), 215–220.
Spivak-Kroizman, T., Lemmon, M. A., Dikic, I., Ladbury, J. E., Pinchasi, D., Huang, J., et al. (1994). Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell*, *79*(6), 1015–1024.
Stansfeld, P. J., & Sansom, M. S. P. (2011). From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations.
Stauber, D. J., DiGabriele, A. D., & Hendrickson, W. A. (2000). Structural interactions of fibroblast growth factor receptor with its ligands. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(1), 49–54.
Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., et al. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, *32*(10), 1478–1488.
Steinberger, D., Mulliken, J. B., & Müller, U. (1996). Crouzon syndrome: previously unrecognized deletion, duplication, and point mutation within FGFR2 gene. *Human Mutation*, *8*(4), 386–390.
Sterner, E., Masuko, S., Li, G., Li, L., Green, D. E., Otto, N. J., et al. (2014). Fibroblast Growth Factor-Based Signaling through Synthetic Heparan Sulfate Block Copolymers Studied using High-cell Density 3D Cell Printing. *The Journal of Biological Chemistry*.
Stevens, T. J., & Arkin, I. T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins*, *39*(4), 417–420.
Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, *76*(6), 2879–2886.
Svergun, D., Barberato, C., & Koch, M. H. J. (1995). CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, *28*(6), 768–773.
The UniProt Consortium. (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, *43*(D1), D204–D212.
Thisse, B., & Thisse, C. (2005). Functions and regulations of fibroblast growth factor signaling during embryonic development. *Developmental Biology*, *287*(2), 390–402.
Touat, M., Ileana, E., Postel-Vinay, S., Andre, F., & Soria, J. C. (2015). Targeting FGFR Signaling in Cancer. *Clinical Cancer Research*, *21*(12), 2684–2694.
Treutlein, H. R., Lemmon, M. A., Engelman, D. M., & Brünger, A. T. (1992). The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry*, *31*(51), 12726–12732.
Trombetta, E. S., & Parodi, A. J. (2003). Quality control and protein folding in the secretory pathway. *Annual Review of Cell and Developmental Biology*, *19*, 649–676.
Trudel, S., Li, Z. H., Wei, E., Wiesmann, M., Chang, H., Chen, C., et al. (2005). CHIR-258, a novel, multitargeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma. *Blood*, *105*(7), 2941–2948.
Turner, N., & Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. *Nature Reviews Cancer*, *10*(2), 116–129.
Utton, M. A., Eickholt, B., Howell, F. V., Wallis, J., & Doherty, P. (2001). Soluble N-cadherin stimulates fibroblast growth factor receptor dependent neurite outgrowth and N-cadherin and the fibroblast growth factor receptor co-cluster in cells. *Journal of Neurochemistry*, *76*(5), 1421–1430.
Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16),

1701–1718.
Volkov, V. V., & Svergun, D. I. (2003). Uniqueness of ab initioshape determination in small-angle scattering. *Journal of Applied Crystallography*, *36*(3), 860–864.
Volynsky, P. E., Polyansky, A. A., Fakhrutdinova, G. N., Bocharov, E. V., & Efremov, R. G. (2013). Role of dimerization efficiency of transmembrane domains in activation of fibroblast growth factor receptor 3. *Journal of the American Chemical Society*.
Wallin, E., Tsukihara, T., Yoshikawa, S., Heijne, von, G., & Elofsson, A. (1997). Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. *Protein Science : a Publication of the Protein Society*, *6*(4), 808–815.
Walter R P Scott, Philippe H Hünenberger, Ilario G Tironi, Alan E Mark, Salomon R Billeter, Jens Fennen, et al. (1999). The GROMOS Biomolecular Simulation Program Package. *Pubs.Acs.org*.
Walter, T. S., Diprose, J. M., Mayo, C. J., Siebold, C., Pickford, M. G., Carter, L., et al. (2005). A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallographica Section D: Biological Crystallography*, *61*(Pt 6), 651–657.
Wang, J., Yu, W., Cai, Y., Ren, C., & Ittmann, M. M. (2008). Altered fibroblast growth factor receptor 4 stability promotes prostate cancer progression. *Neoplasia (New York, NY)*, *10*(8), 847–856.
Wassenaar, T. A., Pluhackova, K., Moussatova, A., Sengupta, D., Marrink, S. J., Tieleman, D. P., & Böckmann, R. A. (2015). High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *Journal of Chemical Theory and Computation*, *11*(5), 2278–2291.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.
Webster, M. K., & Donoghue, D. J. (1996). Constitutive activation of fibroblast growth factor receptor 3 by the transmembrane domain point mutation found in achondroplasia. *The EMBO Journal*, *15*(3), 520–527.
Weinstein, M., Xu, X., Ohyama, K., & Deng, C. X. (1998). FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development (Cambridge, England)*, *125*(18), 3615–3623.
Weiss, M. S. (2001). Global indicators of X-ray data quality. *Journal of Applied Crystallography*, *34*(2), 130–135.
Wesche, J., Haglund, K., & Haugsten, E. M. (2011). Fibroblast growth factors and their receptors in cancer. *The Biochemical Journal*, *437*(2), 199–213.
White, K. E., Cabral, J. M., Davis, S. I., Fishburn, T., Evans, W. E., Ichikawa, S., et al. (2005). Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *American Journal of Human Genetics*, *76*(2), 361–367.
Wiesmann, C., & de Vos, A. M. (1999). Putting two and two together: crystal structure of the FGF-receptor complex. *Structure (London, England : 1993)*, *7*(11), R251–5.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*(Pt 4), 235–242.
Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, *43*(1), 186–190.
Winterpacht, A., Hilbert, K., Stelzer, C., Schweikardt, T., Decker, H., Segerer, H., et al. (2000). A novel mutation in FGFR-3 disrupts a putative N-glycosylation site and results in hypochondroplasia. *Physiological Genomics*, *2*(1), 9–12.

Wu, Z. L., Zhang, L., Yabe, T., Kuberan, B., Beeler, D. L., Love, A., & Rosenberg, R. D. (2003). The involvement of heparan sulfate (HS) in FGF1/HS/FGFR1 signaling complex. *The Journal of Biological Chemistry*, *278*(19), 17121–17129.

Wuechner, C., Nordqvist, A. C., Winterpacht, A., Zabel, B., & Schalling, M. (1996). Developmental expression of splicing variants of fibroblast growth factor receptor 3 (FGFR3) in mouse. *The International Journal of Developmental Biology*, *40*(6), 1185–1188.

Xu, R., Rudd, T. R., Hughes, A. J., Siligardi, G., Fernig, D. G., & Yates, E. A. (2013). Analysis of the fibroblast growth factor receptor (FGFR) signalling network with heparin as coreceptor: evidence for the expansion of the core FGFR signalling network. *FEBS Journal*, *280*(10), 2260–2270.

Xu, X., Weinstein, M., Li, C., Naski, M., Cohen, R. I., Ornitz, D. M., et al. (1998). Fibroblast growth factor receptor 2 (FGFR2)-mediated reciprocal regulation loop between FGF8 and FGF10 is essential for limb induction. *Development (Cambridge, England)*, *125*(4), 753–765.

Yan, D., & Lin, X. (2009). Shaping morphogen gradients by proteoglycans. *Cold Spring Harbor Perspectives in Biology*, *1*(3), a002493.

You, M., Li, E., & Hristova, K. (2006). The achondroplasia mutation does not alter the dimerization energetics of the fibroblast growth factor receptor 3 transmembrane domain. *Biochemistry*, *45*(17), 5551–5556.

You, M., Spangler, J., Li, E., Han, X., Ghosh, P., & Hristova, K. (2007). Effect of pathogenic cysteine mutations on FGFR3 transmembrane domain dimerization in detergents and lipid bilayers. *Biochemistry*, *46*(39), 11039–11046.

You, M., Spangler, J., Niphadkar, N., Shobnam, N., & Hristova, K. (2012). Effect of Pathogenic Cysteine Mutations on FGFR3 Transmembrane Domain Dimerization in Detergents and Lipid Bilayers. *PloS One*, *7*(4), e34808.

Zeev-Ben-Mordehai, T., Vasishtan, D., Siebert, C. A., & Grünewald, K. (2014). The full-length cell-cell fusogen EFF-1 is monomeric and upright on the membrane. *Nature Communications*, *5*, 3912.

Zhang, X., Stappenbeck, T. S., White, A. C., Lavine, K. J., Gordon, J. I., & Ornitz, D. M. (2006). Reciprocal epithelial-mesenchymal FGF signaling is required for cecal development. *Development (Cambridge, England)*, *133*(1), 173–180.

Zhao, Y., Ren, J., Padilla-Parra, S., Fry, E. E., & Stuart, D. I. (2014). Lysosome sorting of β-glucocerebrosidase by LIMP-2 is targeted by the mannose 6-phosphate receptor. *Nature Communications*, *5*. doi:10.1038/ncomms5321

Zhu, H., Anchin, J., Ramnarayan, K., Zheng, J., Kawai, T., Mong, S., & Wolff, M. E. (1997). Analysis of high-affinity binding determinants in the receptor binding epitope of basic fibroblast growth factor. *Protein Engineering*, *10*(4), 417–421.

Zhu, X., Hsu, B. T., & Rees, D. C. (1993). Structural studies of the binding of the anti-ulcer drug sucrose octasulfate to acidic fibroblast growth factor. *Structure (London, England : 1993)*, *1*(1), 27–34.

Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*, *32*(11), 2359–2368.