

Technische Universität München
Max-Planck-Institut für Biochemie

**Structural analysis of the human
Fibroblast Growth Factor Receptor 4 kinase**

**Expression and purification of the
human calcium channel ORAI1**

Elena Lesca

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität
München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Michael Groll

Prüfer der Dissertation: 1. apl. Prof. Dr. Dr. h. c. Robert Huber (i. R.)

2. Priv.-Doz. Dr. Sonja Alexandra Dames

Die Dissertation wurde am 29.09.2014 bei der Technischen Universität München
eingereicht und durch die Fakultät für Chemie am 25.11.2014 angenommen.

At the moment of submitting the present dissertation, a part of it was already published in a scientific journal:

Lesca, E., Lammens, A., Huber, R., Augustin, M., (2014). Structural analysis of the human Fibroblast Growth Factor Receptor 4 Kinase. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.09.004

Acknowledgements

My heartfelt thanks to Prof. Robert Huber, I learnt a lot in your group. I would like to thank you for having encouraged me when results were poor and for having shown me how cautious I have to be with words in a scientific publication. It was a great opportunity to study under your supervision.

Dear Martin, I owe you a lot. You shaped my skills, as a craftsman creates a pot from clay. Thank you for having let me free to follow most of my ideas and for having put the wrong ones on the right rail. I appreciated all the time you dedicated to teach me how to interpret my results, how much I have to be precise and careful, how many efforts and trials are necessary to really validate a result. Thanks for all.

A deep thank to my lab mates: Sohee, Elisabeth, Claudia, Breyan, Marcelino and Robert. I was happy to share my thoughts, worries, ideas, moods and free time with you. You light up my days here. Thanks for always encouraging me to overcome my shyness and be a little proud of myself.

Fanny, Massimo and Luigi, I am deeply happy you understand my desire to do research abroad. I am grateful for all your support, from important to simply things. You have always been there, through long Skype calls and delirious e-mails. I feel strong with you on my side.

Dear Giuliano, you have been part of my PhD period too, a lovely twist of fate. Thanks for patiently understanding my life-style changes and for taking care of me in the sweetest way, fighting with my changeable moods and Lufthansa prices to visit me.

Table of Content

Abbreviations

Part I. Structural analysis of the human Fibroblast Growth Factor Receptor 4 kinase

<i>Summary</i>	8
1. Introduction	9
1.1. Signal transduction and receptors tyrosine kinases	9
1.2. Protein kinases: function and structural elements	12
1.3. Conformational plasticity of kinases	13
1.4. Fibroblast growth factor receptor family	15
1.5. Therapeutic approaches	18
1.6. FGFR4 and its therapeutic value	19
1.7. Structure determination by X-ray crystallography	21
1.8. Molecular replacement	22
2. Materials and Methods	25
2.1. Molecular cloning.....	25
2.2. Insect cell culture and heterologous protein expression	26
2.3. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis analysis (SDS-PAGE).....	26
2.4. Western blot analysis.....	27
2.5. Determination of DNA and protein concentration	27
2.6. Purification protocol.....	27
2.7. Phosphorylation and dephosphorylation – Mass spectrometry analysis (MALDI).....	28
2.8. Addition of inhibitors	28
2.9. Proteins concentration steps	29
2.10. Phosphorylation western blot	29
2.11. Thermofluor analysis.....	30
2.12. Protein crystallization.....	30
2.13. X-ray data collection, data processing and structures solution	30
2.14. PDB accession numbers	31
3. Results	32
3.1. Cloning and expression of FGFR4-KD gene	32
3.2. Purification of FGFR4-KD-wt and FGFR4-KD-V550E mutant.....	33
3.3. Phosphorylation states and mass spectrometry analysis	36
3.4. Crystallization screenings.....	37
3.5. Apo-FGFR4-KD structures	38

3.6. Phosphorylation western blot and Thermofluor analysis	43
3.7. Dovitinib-FGFR4-KD structure	45
3.8. Ponatinib-FGFR4-KD structure	47
4. Discussion	49
4.1. Autoinhibition in FGFR4-KD	49
4.2. FGFR4-KD in cancer	50
4.3. FGFR4-KD and drug design	51
<i>Part II. Expression and purification of the human calcium channel ORAI1</i>	
<i>Summary</i>	53
1. Introduction	54
1.1. Membrane proteins functions	54
1.2. Calcium signaling.....	55
1.3. Calcium Release-Activated Channels (CRAC).....	56
1.4. Orai1, the pore-forming subunit	57
1.5. STIMs: stromal interaction molecules	60
2. Materials and Methods	62
2.1. Molecular cloning.....	62
2.2. Insect cell culture and heterologous protein expression.....	63
2.3. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis analysis (SDS-PAGE).....	63
2.4. Western Blot Analysis.....	64
2.5. Determination of of DNA and protein concentration.....	64
2.6. Membrane protein isolation and preparation.....	64
2.7. Purification protocol.....	64
2.8. Proteins concentration steps	65
2.9. Detergents screenings.....	65
2.10. Protein crystallization.....	66
3. Results and Discussion	67
3.1. Cloning and expression of Orai1 variants	67
3.2. Detergents screening	69
3.3. Oligomeric states of Orai1	69
3.4. Optimization of Orai1 purification.....	70
3.5. Purification of the CRAC complex: Orai1 with STIM1-CAD domain.....	70
3.6. Crystallisation of Orai1	74
3.7. Future outlook	74
References	76

Abbreviations

Å, angstrom, 1×10^{-10} m	Ni-NTA, nickel-nitriloacetat acid
ABL, Abelson murine leukemia viral oncogene homolog 1	NMR, nuclear magnetic resonance
ABS, absorbance	NUS, antitermination protein
AMP-PNP, adenylyl-imidodiphosphate	OD _(x) , optical density at a wavelength of x nm
ATP, adenosine-5'-triphosphate	PCR, polymerase chain reaction
CHK1, checkpoint kinase 1	PDB, Protein Data Bank
CHS, cholesteryl hemisuccinate	<i>PDC</i> , protein- detergent complex
c-MET, hepatocyte growth factor receptor kinase	PEG, polyethylene glycol
CRAC, Calcium release-activated calcium channel protein 1	PMSF, phenylmethylsulfonyl fluoride
DDM, n-Dodecyl β-D-maltoside	PVDF, polyvinylidene fluorid
DDR1, discoidin domain receptor family member 1	RIPK2, receptor-interacting serine/threonine-protein kinase 2
DNA, deoxyribonucleic acid	RMS, rhabdomyosarcoma
DTT, 1,4-dithiothreitol	RMSD, root-mean-square-deviation
E. coli, Escherichia coli	Ror2, neurotrophic tyrosine kinase 2
EG, ethylene glycol	RTK, receptor tyrosine kinase
EM, electron microscopy	SAD, single wavelength anomalous dispersion
ER, endoplasmic reticulum	<i>SCID</i> , severe combined immunodeficiency
FGF, fibroblast growth factors	SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
FGFR, fibroblast growth factor receptor	SEC, size-exclusion chromatography
GPCR, G coupled-protein receptor	<i>Sf</i> , <i>Sodoptera frugiperda</i>
GST, Glutathione S-transferase	SIR, single isomorphous replacement
HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	SLS, Swiss Light Source
IRK, insulin receptor kinase	SOAR, Store-Operated Calcium Channels
KD, kinase domain	TCEP, tris(2-carboxyethyl)phosphine
kDa, kilodalton	TEV, Tobacco Etch Virus protease
LCP, lipid cubic phase	TLS, translation-liberation-screw-rotation
MAD, multiwavelength anomalous dispersion	TM, transmembrane helix
MES, 2-(N-morpholino)ethanesulfonic acid	TRIS, tris(hydroxymethyl)aminomethane
MIR, multiple isomorphous replacement	TrkA, high affinity nerve growth factor receptor
MR, molecular replacement	v/v, volume/volume
MW, molecular weight	w/v, weight/volume
	WT, wild-type

Part I**Structural analysis of the human
Fibroblast Growth Factor Receptor 4 kinase****Summary**

The family of Fibroblast Growth Factor Receptors (FGFRs) plays an important and well characterized role in a variety of pathological disorders. FGFR4 is involved in myogenesis and muscle regeneration, and mutations affecting the kinase domain of FGFR4 may cause cancer, e.g. breast cancer or rhabdomyosarcoma. Whereas FGFRs 1-3 have been structurally characterized, the structure of the FGFR4 kinase domain has not yet been reported.

In this study we present four structures of the kinase domain of FGFR4, in its apo-form and in complex with different types of small-molecule inhibitors. The two apo-FGFR4 kinase domain structures show an activation segment similar in conformation to an autoinhibitory segment observed in the hepatocyte growth factor receptor kinase but different from the known structures of other FGFR kinases.

The structures of FGFR4 in complex with the type I-inhibitor Dovitinib and the type II-inhibitor Ponatinib reveal the molecular interactions with different types of kinase inhibitors and may assist in the design and development of FGFR4-inhibitors.

1. Introduction

1.1. Signal transduction and receptors tyrosine kinases

Eukaryotic cells use many ways to communicate with each other and signal transduction indicates the fascinating steps through which signals cross the plasma membrane and reach the nucleus or other organelles where the signal is finally influencing gene expression in its various processes. DNA transcription is a delicate process and only specific proteins can pass the double nucleus membrane. Moreover, the activity of the proteins involved must be tightly regulated and many checkpoints are necessary to avoid errors and protect cell life.

The first checkpoint involves proteins on the plasma membrane (Fig. 1) that are responsible of distributing the signal to cytosolic proteins: G protein-coupled receptors (GPCRs), receptors with enzyme-linked activities (such as receptor tyrosine kinases), cytokine receptors and/or others.

The main part of my thesis is on *Receptor Tyrosine Kinases* (RTK). RTKs form a large family of proteins responsible for fundamental cellular processes: cell cycle, cell migration, cell metabolism and survival, cell proliferation and differentiation. RTKs are folded in an extracellular domain, a membrane spanning domain and one or more intracellular tyrosine kinase domain(s) (KDs). The extracellular domain contains various sub-domain types (Fig. 2) able to interact with and bind specific ligands. The binding causes conformational changes which pass through the membrane domain and reach the cytoplasmic domain. Its regulatory sequences are subject to autophosphorylation in trans and phosphorylation by partner kinases¹⁻³. RTK monomers can also oligomerize without ligands, however their interaction is not stable, so they remain “inactive”^{3,4}. In addition, the single membrane-spanning α -helix may contribute to dimerisation in some cases, although its precise role is not defined yet. One receptor in the dimer/oligomer phosphorylates one or more tyrosines in a neighboring RTK, and then the phosphorylated receptor serves as a site for assembly and activation of intracellular signaling proteins^{3,5-7}. Phosphorylation events and their subsequent conformational changes are the way through which the external signal is transmitted to the nucleus.

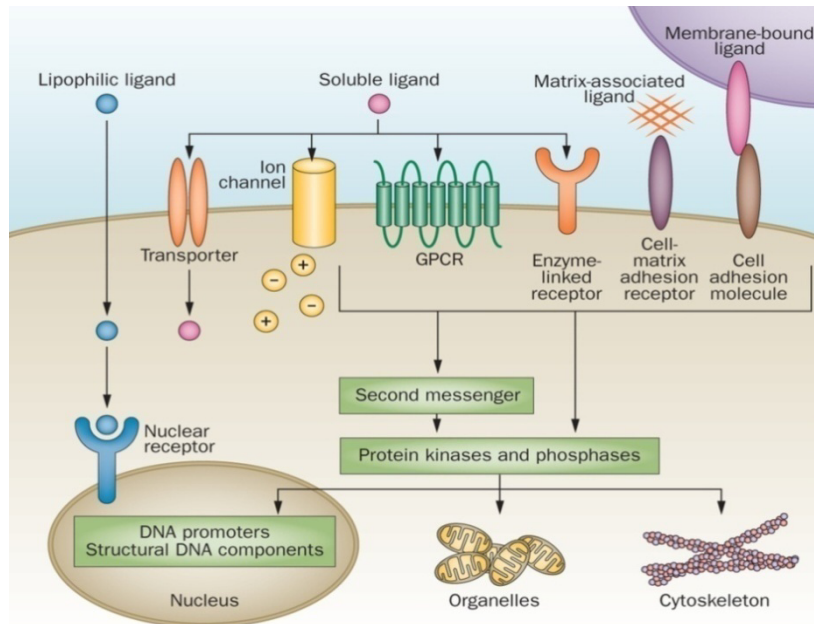


Fig. 1. Cell signaling receptors differ in their mechanisms of activation and signal transmission. They are broadly classified into three groups: GPCRs, receptors with enzyme-linked activities (such as receptor tyrosine kinases) and cytokine receptors. Some receptors mediate cell-to-cell or cell-to-matrix contact by binding transmembrane proteins or extracellular proteins, respectively, and link cell adhesion to outside-in and inside-out signaling. Lipophilic ligands that can cross the cell membrane by simple diffusion target intracellular receptors that act directly as transcriptional regulators (Reiser et al., 2014⁸).

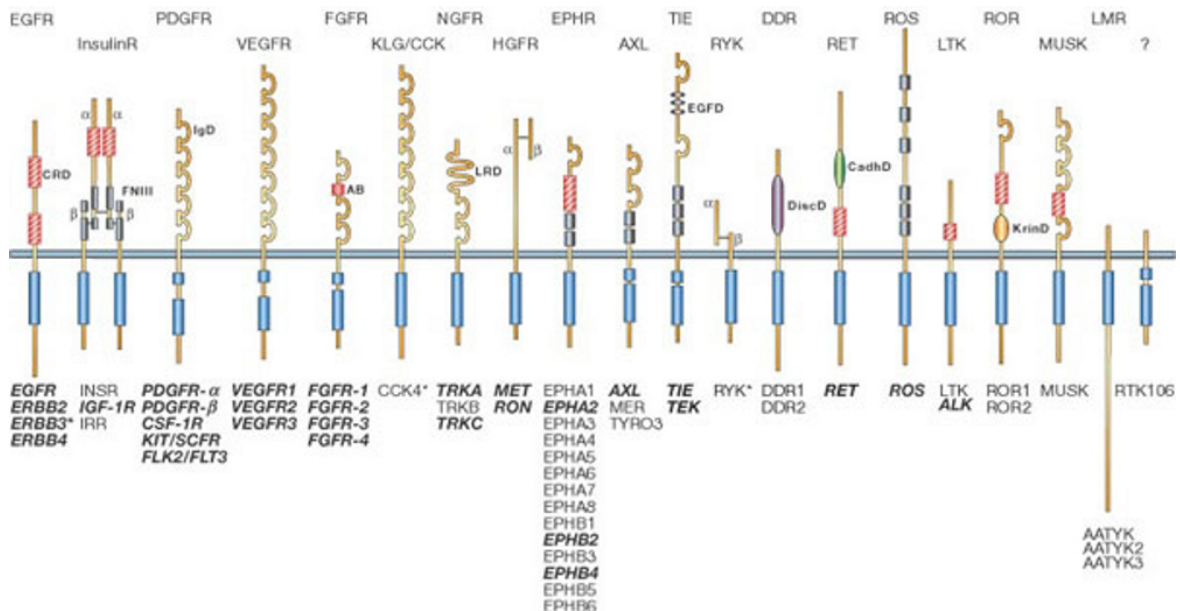


Fig. 2. Schematic classification of RTKs. The prototypic receptor for each family is indicated above the receptor, and the known members are listed below (Blume-Jensen et al., 2001⁹).

1.2. Protein kinases: function and structural elements

A protein kinase is an enzyme which transfers the γ -phosphate of a nucleoside triphosphate (usually ATP) to specific amino acids with free hydroxyl groups. Such residues can be serine and threonine (serine/threonine kinases)¹⁰, tyrosine (tyrosine kinases)¹ or histidine (histidine kinases)¹¹. Some kinases have dual activity on two types of amino acids (dual-specificity kinases)¹².

The kinase fold is composed of two domains: the smaller N-terminal lobe and the larger C-terminal lobe, connected by a flexible hinge region (Fig. 3a). The N-lobe is made up of a five stranded β -sheet and (usually) two α -helices, of which the α C-helix is the structurally most conserved, whereas the larger C-lobe is mainly α -helical. An ATP-molecule binds in the deep cleft between the two lobes. The adenosine base forms H-bonds with the kinase hinge region, the ribose moiety binds to the ribose binding pocket and the phosphate groups interact with the *glycine-rich loop* (GXGXXG), which is also called the phosphate-binding loop (P-loop). Usually, phosphorylation sites lie within unstructured regions of proteins, and structural studies using substrate peptides have shown that the C-lobe recognizes the sequence on either side of the phosphorylated residues¹³. In the N-lobe, a conserved *lysine-glutamate salt bridge* is typical in active kinases and formed by a glutamate located on the α C-helix and a lysine residue located on the β 3-strand. This lysine also interacts with the phosphates of ATP, stabilizing them in the correct position for catalysis¹⁴. The *HRD*- and *DFG*-motifs are other essential features of protein kinases. The aspartate residue of the HRD-motif acts as a catalyst that deprotonates the substrate, for that this region of the kinase is also named the catalytic loop¹⁵. The aspartate of the DFG-motif coordinates a magnesium ion that activates the γ -phosphate of the ATP-molecule, and this region of kinases is often called magnesium-binding loop. In active kinases, the DFG-motif adopts a DFG-in conformation and contacts the HRD-motif through a hydrophobic interaction. Displacement of these motifs into positions incompatible with catalysis is part of the regulatory mechanism of many kinases and can also be induced by inhibitors (Fig. 3b).

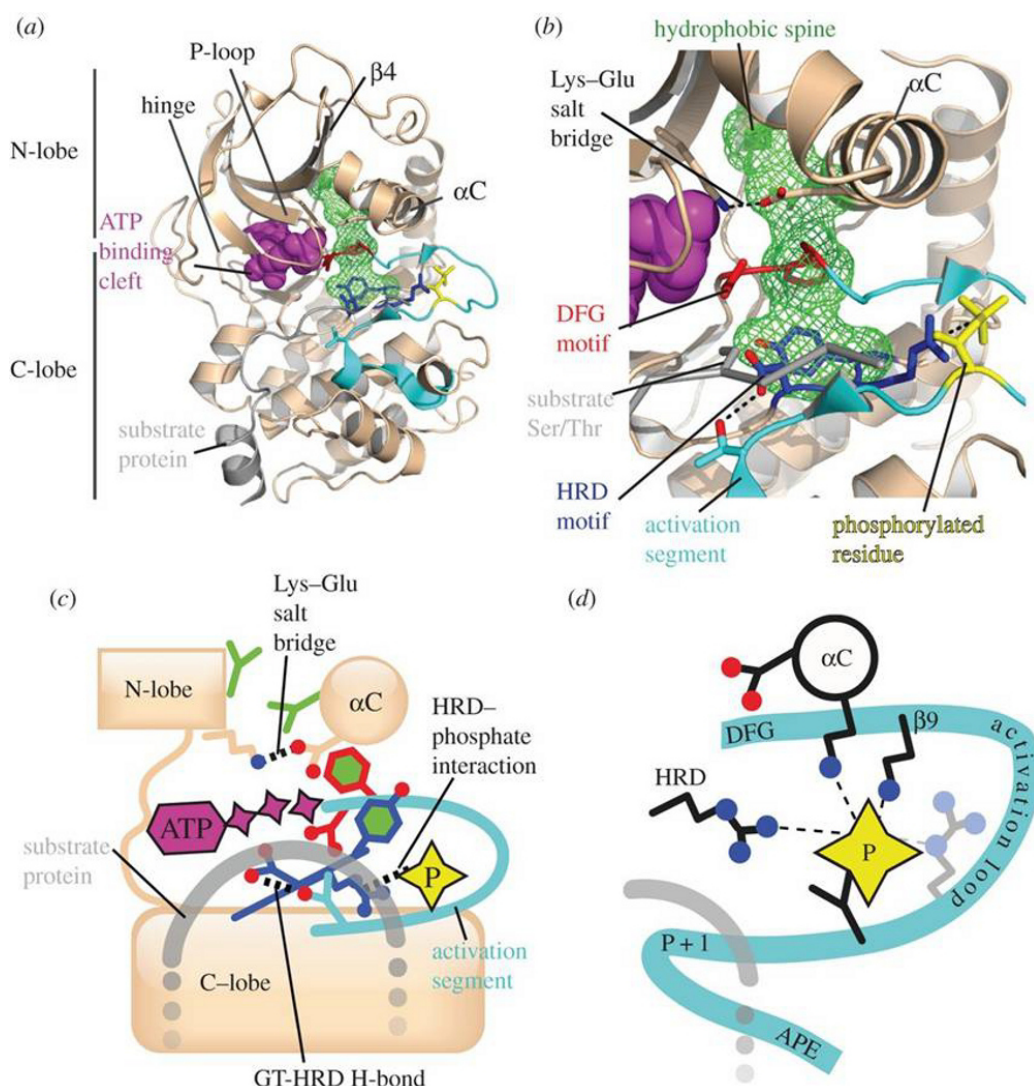


Fig. 3. Distinctive structural elements of kinases. Figure colour scheme: kinase main chain and Lys/Glu pair (beige), HRD-motif (blue), DFG-motif side chains (red), activation segment (cyan), phosphorylated residue on activation loop (yellow), hydrophobic spine (green), ATP (magenta), substrate protein (grey). **(a)** Architecture of the protein kinase A (PKA) catalytic domain in cartoon representation, with key residues shown as sticks and the ADP ligand shown as spheres (PDB code 1JBP). The side chains of the four residues that comprise the hydrophobic spine are surrounded by a wire mesh. **(b)** Close-up view of the active site. Key interactions between residues are indicated with dashed black lines; these do not necessarily imply hydrogen bonds. **(c)** Schematic summarizing the major features of the active kinase conformation. **(d)** Schematic illustrating the canonical interactions between the phosphorylated residue of the activation loop and basic residues in the α C-helix, β 9 and the HRD-motif (Bayliss et al., 2012¹⁶).

The *hydrophobic spine* is a typical feature of active kinase and is made up of the side chains of four hydrophobic residues: a first side chain comes from the β 4-strand in the N-lobe and the second one from the adjacent α C-helix; the side chains of the phenylalanine from the DFG-motif and the histidine from the HRD-motif¹⁷ complete the hydrophobic spine (Fig. 3b). The activation segment starts at the DFG-motif and ends at the *APE-motif*, another consensus sequence. The activation loop is variable in length and sequence and is the primary site of regulatory phosphorylation¹⁴. Usually, phosphorylated residue/s

on the activation loop is/are typically coordinated by the side chains of basic residues at specific positions: the residue preceding the glutamic acid on the α C-helix, the residue three positions C-terminal to the DFG-motif on the β 9-strand and the arginine residue of the HRD-motif (Fig. 3d). These three anchors help to stabilize the activation segment in the correct conformation for substrate binding¹⁶.

1.3. Conformational plasticity of kinases

Protein phosphorylation is a refined and complicated communication mechanism essential for cell life: 2% of the human genes encode protein kinase domains^{18,19}. The spatial and temporal control of phosphorylation events is crucial to cell growth and development, and this control depends on kinases. Deregulated kinase activity has usually catastrophic consequences, resulting in cell transformation and cancer. As shown in figure 4, a kinase can have different intermediate states: inactive and autoinhibited, partially in/active, active, active and phosphorylated. Some of these conformations are obtained *in vitro* with the help of inhibitors. Generally, when the enzyme is inactive the hinge cleft is open, the activation segment is folded so that the residues involved in phosphorylation are not accessible and the hydrophobic spine is disassembled. In some kinases the *DFG-in conformation* of the DFG-motif is differently oriented with respect the active state and the lysine-glutamate salt bridge is disengaged. An active kinase shows the opposite conformation, with mainly a strong rearrangement of the activation segment and a closure of the N-lobe to the C-lobe due to the ATP-binding. The DFG-motif is used as an indicator of another kinase inactive state: the *DFG-out conformation*. This one was discovered in the unliganded insulin receptor kinase²⁰, and later exploited for the development of specific inhibitors. This conformation is an inactive state of the enzyme and its main feature is the flip of phenylalanine of the DFG-motif with respect to the DFG-in conformation. As a consequence, also the activation segment changes its fold. The DFG-out conformation creates an additional hydrophobic pocket adjacent to the ATP-binding site, frequently referred to as the allosteric binding site. Because the amino acids surrounding this pocket are less conserved than those in the ATP binding pocket and therefore, the allosteric pocket is exploited in for gaining selectivity during inhibitor design. The tight regulation of kinase conformation is crucial for kinase activity and partially depends of the kinase itself and partner proteins.

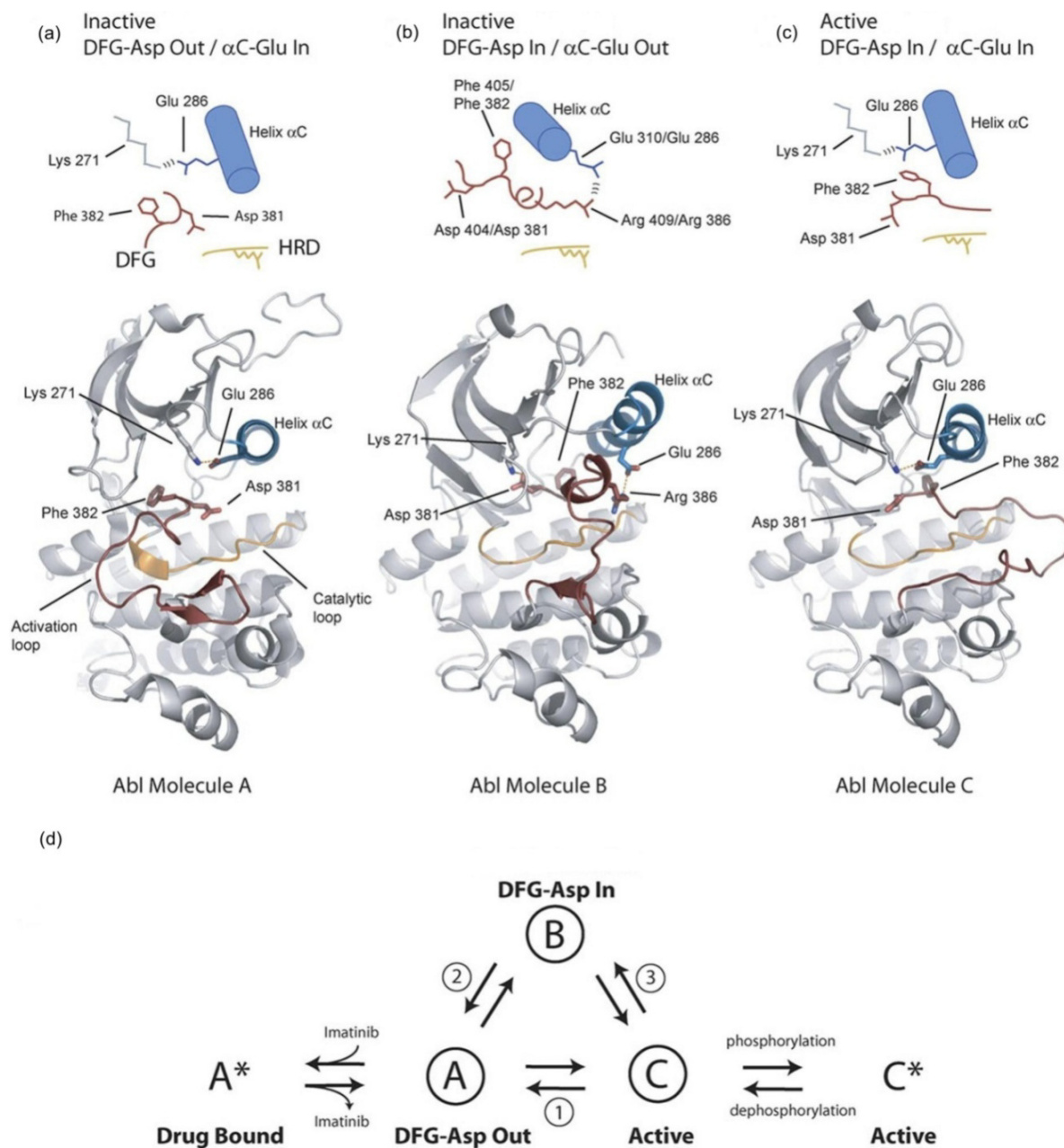


Fig. 4. Three key Abl kinase conformations. (a - c) at the top, a close-up representation of each state, detailing the conformations of the DFG-motif (red) and α C-helix (blue). Below, a crystal structure of each conformation is shown. The activation loop is colored red, the α C-helix blue, and the catalytic loop orange. (d) In cells the active conformation of Abl undergoes rapid autophosphorylation that is expected to trap the protein in the active conformation (indicated as C*). Similarly, the inhibitor imatinib only binds to Abl when the kinase domain adopts conformation A and forms a stable complex with the protein (A*). The interconversion between the different states of Abl is shown in the context of this competition (Figure adapted from Levinson et al., 2006²¹).

1.4. The fibroblast growth factor receptor family

Fibroblast growth factor receptors (FGFRs) act as mitogens in embryogenesis and development (cell morphogenesis, proliferation, differentiation and migration) and in adult organisms (organ, vascular and skeletal development)^{22–24}. The FGFRs 1–4 have an overall structure similar to most RTKs, containing three Ig-like extracellular sub-domains, a trans-membrane helix and an intracellular tyrosine kinase domain²³ (Fig.5a). FGFR5 is composed of an extracellular domain only, and might negatively regulate signaling^{25,26}.

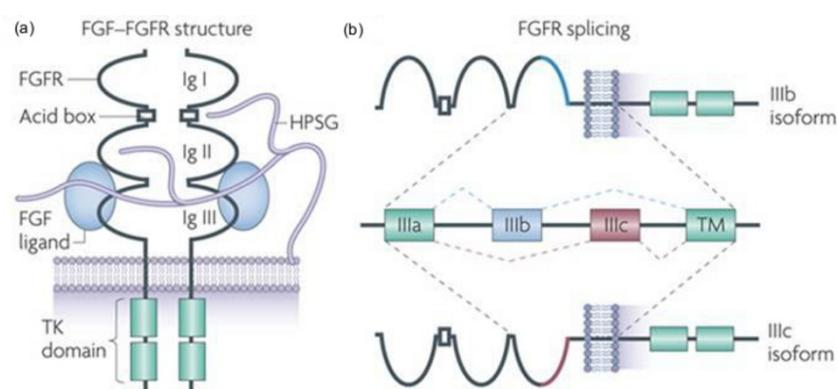


Fig. 5. Schematic representation of a FGFR protein. **(a)** FGF receptor (FGFR) complex comprises two receptor molecules, two FGFs and one heparan sulphate proteoglycan (HSPG) chain. FGFs bind with low affinity to cell surface HSPGs (purple) and with high affinity to specific FGFRs. The FGFRs consist of three extracellular immunoglobulin (Ig) domains, a single transmembrane helix and an intracellular tyrosine kinase domain. The II and III Ig domains form the ligand-binding pocket and have distinct domains that bind both FGFs and HSPGs. **(b)** Ligand-binding specificity is generated by alternative splicing of the Ig III domain. The first half of Ig III is encoded by an invariant exon (IIIa), which is spliced to either exon IIIb or IIIc, both of which splice to the exon that encodes the transmembrane (TM) region. Epithelial tissues predominantly express the IIIb isoform and mesenchymal tissues express IIIc. FGFR4 is expressed as a single isoform that is paralogous to FGFR-IIIc. Figure adapted from Turner et al., 2010²⁴.

In mammalia there are 18 different FGFs and each FGFR recognizes a unique subset of the fibroblast growth factors²⁷. The extracellular ligand-binding domain of an FGFR is composed of three immunoglobulin (Ig) like domains, labeled as D1–D3, a stretch of seven to eight acidic residues connecting D1 and D2 (the “acid box”) and a conserved positively charged region in D2 that serves as a binding site for heparin^{28,29} (Fig. 5a). The extracellular domain of FGFR was shown to undergo N-glycosylation and mutations that impaired glycosylation resulted in permanent activation of FGFR³⁰. The cytoplasmic domain contains the tyrosine kinase domain, with conserved regulatory sequences that undergo autophosphorylation and/or phosphorylation by heterologous protein kinases.

An important hallmark of the FGFR family is that a variety of FGFR isoforms are generated by alternative splicing of FGFR transcripts, usually affecting the Ig-like domains, profoundly alternating ligand-binding specificity^{31,32} (Fig. 5b). The alternative splicing in D3 exists in FGFR1, 2 and 3, but not in FGFR4. Other splicing forms include soluble secreted FGFRs forms^{33,34} and deletions in the kinase domain. Variants lacking parts of the kinase domain may act as down regulators of the signal, since they can heterodimerize with active forms, leading to non-functional dimeric receptors^{35,36}. Fibroblast growth factors (FGFs) bind to the extracellular domain and stimulate receptor dimerisation together with heparin-like molecules^{23,28,37}.

Cell surface heparin or heparan sulfate proteoglycans (HSPG) interact with FGFs to induce growth factor polymerization, binding to FGFR and subsequent dimerisation of FGFRs. HSPG absence limits dimerisation and activation of the FGF-FGFR complex^{37,28,29}. In activated oligomers, the kinase domain becomes active by displacement of its activation segment to an “open” position, allowing Mg-ATP binding and phosphotransfer events on signaling proteins²⁴. Dimerisation juxtaposes the kinase domains into close proximity, enabling them to *trans*-phosphorylate each other. The activity of FGFR-KDs is tightly regulated and connected with phosphorylation of up to seven specific tyrosines residues^{38,39}. Some of these are docking sites for downstream signaling molecules containing SH2 (Src homology 2) domains. For instance, the SH2 domain of PLC γ (phospholipase C γ) binds to a phosphotyrosine in the C-terminal tail of the activated receptors⁴⁰. The juxtamembrane domain of FGFRs is considerably longer than that of other receptor tyrosine kinases. This region contains a highly conserved sequence that serves as a binding site for the phosphotyrosine binding (PTB) domains of the two members of the FRS2 family of docking proteins FRS2a and FRS2b⁴¹ (Fig. 6a).

Activating mutations, gene amplification, altered splicing and chromosome translocation of FGFRs are responsible for some developmental syndromes and human cancers in different tissues^{23,42}.

Activating mutations and chromosomal translocations can result in ligand-independent dimerisation or constitutive activation of the kinase. Intragenic translocations generate fusion proteins⁴³, usually with the N-terminus of a transcription factor fused to the C-terminal FGFR kinase domain, resulting in dimerisation of the fusion protein and constitutive signalling⁴⁴. Receptor gene amplification, which results in receptor over expression, can also be accompanied by altered splicing, contributing to receptor

accumulation. In addition, tumour cells can stimulate stromal cells to release FGF ligands and increase the release of ligands from the extracellular matrix (paracrine loop), or FGF ligands are produced by a cancer cell (autocrine loop)²⁴. The autocrine loop can be established by FGFR expression out of context or by the increased expression of FGF ligands. FGF released from stromal cells or cancer cells can act on endothelial cells to promote angiogenesis (Fig 6b).

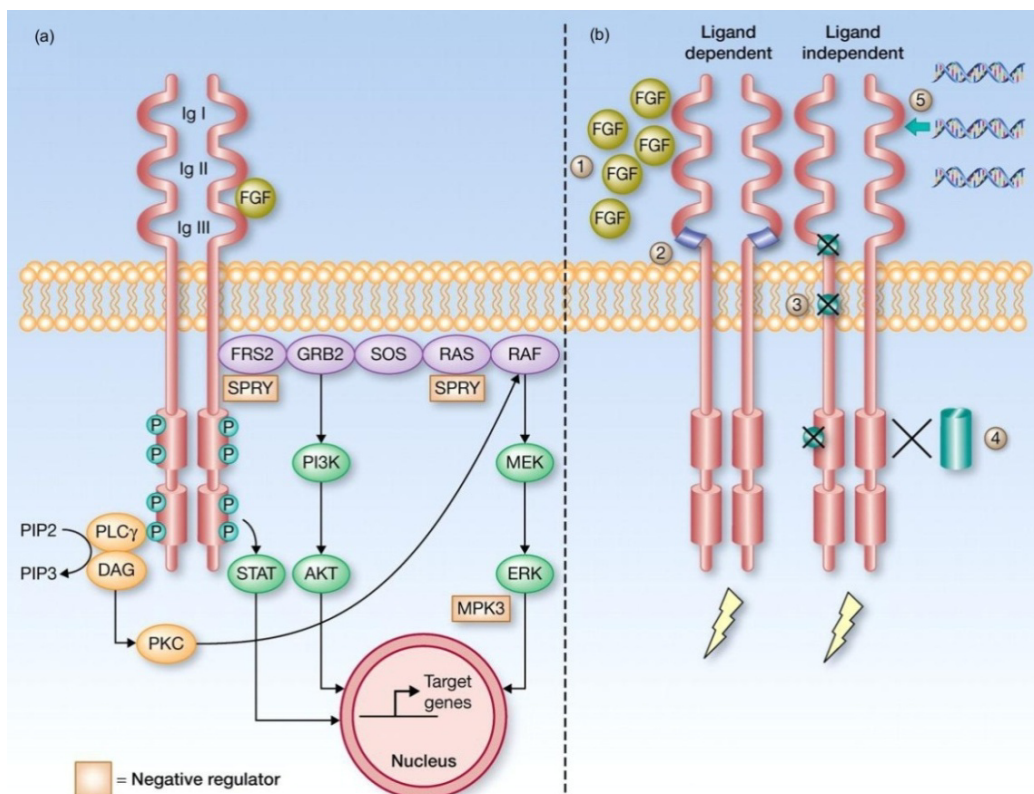


Fig. 6. FGFR structure, signaling, and dysregulation in cancer. **(a)** FGFR and downstream signalling: a complex is formed among the FGF ligand, heparan sulfate and FGFR to cause receptor dimerisation and *trans*-phosphorylation at several tyrosine residues in the kinase domain. Subsequent downstream signaling occurs through two main pathways: via the intracellular receptor substrates FRS2 and PLC γ or via STAT-dependent signaling. Negative regulation of the FGFR signaling pathway is mediated via FGF-regulated inhibitory factors such as SPRY and MKP3. **(b)** FGFR dysregulation in cancer. Ligand activation of FGFRs can be dysregulated when a cell overproduces FGF ligand (1) that activates a corresponding FGFR, or when a cell produces splice-variant FGFRs (2) that have altered specificity to endogenous FGF ligands. Ligand-independent dysregulation of FGFRs can occur when an FGFR becomes mutated (3), leading to receptor dimerization or constitutive activation of the kinase, or when a gene translocation occurs (4), whereby the FGFR fuses with a transcription factor or promoter region resulting in over expression or activation of the FGFR. A third mechanism is when a gene amplification for the receptor occurs (5), resulting in grossly exaggerated expression of the receptor. Other mechanisms of FGFR dysregulation include germline SNPs, which are associated with increased cancer risk or a poor prognosis, and impairment of the normal negative feedback mechanisms, such as reduced expression of the negative regulator SPRY. (Brooks et al., 2012⁴⁵).

Crystal structures of FGFR1 and FGFR2 kinase domains have been solved in different conformations: apo- and phosphorylated protein^{46,47,48,7,49} and in complex with small-molecule inhibitors^{50,51}. Moreover, different FGFR2 mutants have been structurally characterized⁴⁹. The structure of the FGFR3-K650E mutant was recently published as the first FGFR3 structure⁵².

In this study I describe four crystal structures of the FGFR4 kinase domain in different conformations. I have solved two different apo-FGFR4 structures with the DFG-motif adopting a conformation similar to an autoinhibitory conformation observed in the hepatocyte growth factor receptor kinase (c-MET), but different to the known structures of FGFR kinases. I have also solved crystal structures of FGFR4-KD in complex with the inhibitors Dovitinib and Ponatinib, revealing the first FGFR protein in a DFG-out conformation.

1.5. Therapeutic approaches

To date, three strategies have been developed to treat cancer related to FGFR activity: kinase inhibitors, monoclonal antibodies and FGF ligand trap. In this thesis, I focus on the first strategy and I show two structures of the FGFR4-KD in complex with two inhibitors.

Aberrant RTKs signaling may be inhibited through the development of small-molecule drugs that selectively interfere with the tyrosine kinase activity and thereby block receptor autophosphorylation and activation of downstream signal transducers^{53,54}.

In the case of FGFRs, tyrosine kinase inhibitors would be relevant therapeutics for the treatment of cancers over expressing FGFRs (breast cancer), displaying activating FGFR mutations (bladder and endometrial cancers), or expressing chimerical fusion proteins (rhabdomyosarcoma). Small-molecule tyrosine kinase inhibitors targeting FGFRs are currently in early phases of clinical trials, and most of them show broad specificity and target not only FGFRs, but also PDGFRs and/or VEGFRs due to high structural similarity of their kinase domains, increasing drug side effects^{24,27}. FGFR-specific tyrosine kinase inhibitors have shown promising results in cancer cells and mouse models. The broad expression of FGFRs throughout the body and the importance of the FGFRs in various physiological processes, as well as the high degree of homology between the kinase domains of FGFRs, have to be considered in the development and application of FGFR tyrosine kinase inhibitors for cancer treatment.

1.6. FGFR4 and its therapeutic value

FGFR4 is involved in myogenesis and muscle regeneration with a tissue-specific expression⁵⁵⁻⁵⁷. FGFR4 displays a unique affinity to certain FGF ligands^{55,58} and it has different splice variants, one concerning the kinase domain³⁶. FGFR4 has a reduced autophosphorylation and tyrosine kinase activity compared to other FGFRs^{22,57,59}. FGFR4 tyrosines phosphorylation has been reported to be glycosylation dependent and to require hetero-oligomerisation with other phosphorylated and active FGFRs³⁶. Both FGFR1 and FGFR4 are expressed during skeletal muscle growth, development and repair, as well during myogenesis of satellite cells^{55,60,61}. FGFR4 expression and function was also implicated in muscle regeneration *in vivo*⁶⁰. While FGFR1 seems to play a role in supporting myoblast proliferation, the specific role of FGFR4 during myogenesis remains an enigma^{36,61}.

Many mutations involve different FGFR4 domains and have been found in different cancer types (Fig. 7). For instance, the known polymorphism G388R is in the transmembrane region, occurs in approximately 50% of the population⁶² and has been associated with cancer development and progression⁶³⁻⁶⁶. The Y367C mutation identified FGFR4 as dominant oncogene and a driver of tumor growth in breast cancer⁶⁷. My work was on the FGFR4 kinase domain, which is also affected by tumorigenic mutations⁴². Especially, mutants N535K and V550E are activating mutations in rhabdomyosarcoma (RMS), a childhood cancer^{68,69}. In addition, Valine 550 has been found mutated to methionine in breast cancer⁷⁰. Other mutations of the kinase domain are known to result in lung, skeletal muscle and skin cancer⁴². It has been reported that chemotherapy resistance involves FGFR4 up-regulation⁷¹ and proposed to become a targetable regulator of drug resistance in colorectal cancer⁷². For these reasons and for its oncogenic potential, FGFR4 is a valid therapeutic target. Kinases domain ligands may inhibit wild-type up-regulated FGFR4, extracellular and transmembrane domains mutants and kinase domain mutants. Here, I show two FGFR4 structures in complex with inhibitors Ponatinib and Dovitinib, both in second phase of clinical trials. Recently, Ponatinib was identified as a potent FGFR4 inhibitor with an IC₅₀ in the nanomolar range⁷³. Especially, it has been shown that Ponatinib inhibits tumor growth in mice models expressing mutationally activated FGFR4⁷⁴. Dovitinib is an ATP-competitive tyrosine kinase inhibitor and shows activity against FGFRs1-4, VEGFRs1-3 and other RT-

kinases^{75,76}. Dovitinib induces both cytostatic and cytotoxic responses in vivo resulting in regression of FGFR3-expressing tumors⁷⁷.

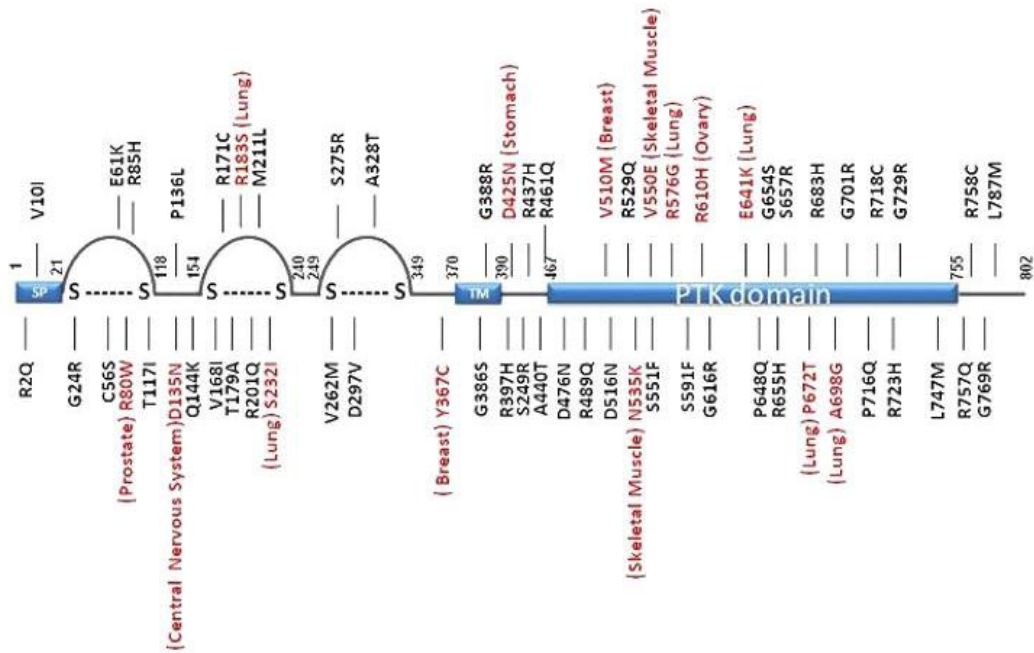


Fig. 7. Schematic view of the main FGFR4 mutations. Germline mutations are in black, somatic mutations are in red. In brackets it states cancer type connected to the relative mutation. Figure adapted from <http://atlasgeneticsoncology.org/Genes/FGFR4ID512ch5q35.html>.

1.7. Structure determination by X-ray crystallography

Protein function is deeply related to three-dimensional protein folding, and knowledge of the 3D-structure of a target protein is obviously essential for structure-based functional studies. To date, different methods have been developed to reach protein structural models at different resolution levels: Electron Microscopy (EM), Small-Angle Scattering of X-rays (SAXS) and Small-Angle Neutron Scattering (SANS), Atomic Force Microscopy (AFM), X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. X-ray free-electron lasers (XFELs) is one of the most recent method, developed to study GPCRs and membrane proteins crystals, which are usually very tiny⁷⁸.

X-ray crystallography has no size limitation for the molecule or complex under investigation, but displays only a “snapshot” of the macromolecule and X-ray diffracting crystals are required. The crystallization success is currently not predictable, making it a sometimes tedious and labor-intensive process. During data collection, the X-ray detector can only record intensities but not phases of the scattered electromagnetic waves. Each reflection on the diffraction pattern corresponds to a wave consisting of structure factor amplitude and phase angle. A structure factor can be represented as the vector F_{hkl} (a complex number, named Fourier coefficient) with the amplitude $|F_{hkl}|$ and the phase φ_{hkl} (Fig. 8a). Amplitudes are easily calculated by taking the square root of the intensity, but phases are lost during data collection.

The phase problem is solved using three main methods:

- molecular replacement (MR), where a model structure is present;
- SIR, MIR methods, exploiting scattering of heavy atoms;
- SAD, and MAD methods, exploiting anomalous scattering;
- direct methods, requiring very high resolution and small proteins for an *ab initio* phase determination.

FGFR4 was solved by molecular replacement and the theory of this method is briefly explained in the next paragraph.

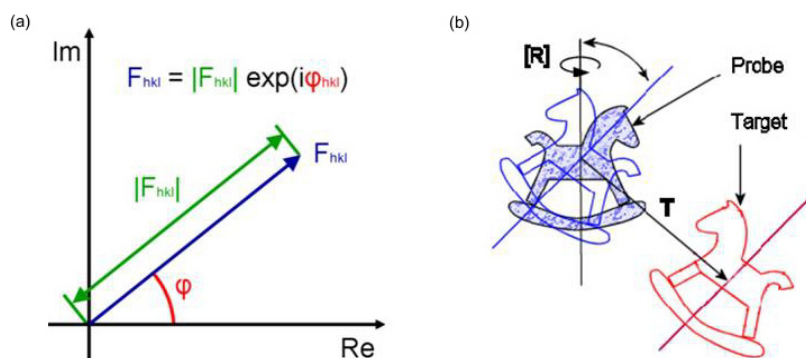


Fig. 8. Molecular Replacement. **(a)** Argand diagramme. A complex number can be represented as a pair of numbers (a, b) forming a vector on a diagram representing the complex plane. "Re" is the real axis, "Im" is the imaginary axis, and "i" is the imaginary unit. **(b)** The blue horse molecule with known structure (A) is rotated through the [R] operation and shifted through T to bring it over the position of the unknown red molecule (A'). The molecular replacement equation is simplified as $A' = A [R] + T$. Figure adapted from <http://www-structmed.cimr.cam.ac.uk/Course.html>.

1.8. Molecular replacement

Molecular Replacement (MR) is a phasing method that can be applied when a search model structurally similar (~40%) to a molecule in the target crystal structure is available. Nowadays, molecular replacement is extensively used due to the numerous structures available and dedicated software packages and servers (MolRep, Phaser, MrBump, AMORE)⁷⁹⁻⁸¹. To carry out molecular replacement, the model structure has to be correctly oriented in the unknown unit cell. This means specifying three rotation angles to place the molecule in the 3D space and three translational parameters to correctly orient the molecule in the unit cell (Fig. 8b). Identifying these parameters is relatively easy due to the Patterson function⁸². In contrast to the electron density map (Fourier transform), the Patterson map can be computed directly from intensities (amplitudes squared) without additional phases. In a Patterson function, all phases are set to 0, resulting in a map where the peaks correspond to distance vectors between individual atoms. For example, if there is a peak of electron density for atom 1 at position x_1 and a peak of electron density for atom 2 at position x_2 , then the Patterson map will have peaks at positions $x_2 - x_1$ and $x_1 - x_2$. So for each vector from atom x_1 to x_2 there is an oppositely oriented vector of the same length (from atom x_2 to x_1), therefore the Patterson map always has centrosymmetry. The vectors can be divided in two categories: intramolecular vectors (from one atom in the molecule to another atom in the same molecule) depending only on the orientation of the molecule and not on its position in the cell, so these can be exploited in the rotation function; intermolecular vectors depend

both on the orientation of the molecule and on its position so, once the orientation is known, these can be exploited in the translation function. In the Patterson map the height of the peaks is proportional to the product of the heights of the two peaks in the electron density map, especially the Patterson map has its maxima at the intermolecular vectors.

Once that the space group has been pinpointed, the selected software combines the model structure and the unknown one, applies algorithms of the rotation and translation functions and the result is an approximate initial set of phases of the Fourier coefficient of the unknown structure. For that purpose a Patterson map is calculated from the experimental intensities and compared to a theoretical Patterson map of the model structure coordinates.

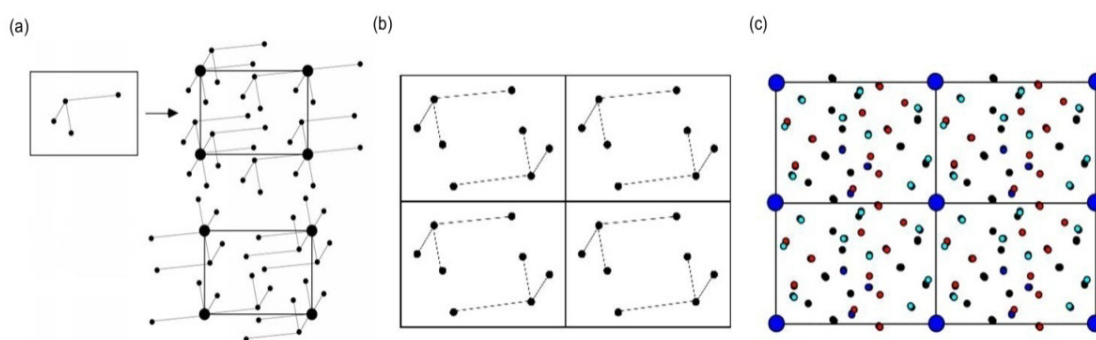


Fig. 9. Patterson map representation. **(a - b)** Rotation and translation function are represented in the unit cell **(a)** and in the crystal lattice **(b)**. **(c)** Here a entire Patterson map is represented: in light blue intramolecular vectors resulting from rotation function, while in blue and red intermolecular vectors, resulting from translation function. Figure adapted from <http://www-structmed.cimr.cam.ac.uk/Course.html>.

Once that the phases have been approximated, the new structure is rebuilt and refined during many cycles. Refining is achieved through statistical adjustment of the atomic parameters to better fit the experimental diffraction data. The R-value measures how much calculated amplitudes differ from the observed amplitudes. To prevent overfitting and getting a misleading level of agreement with the observed data, a fraction of the data is left out during refinement for cross-validation. These cross-validation data gives the R-free value, which is an unbiased indication of the quality of the structure. Two methods are widely used in refinement: maximum likelihood (REFMAC⁸³) and simulated annealing (CNS/Phenix⁸⁴). Both methods use restraints of bond lengths, angles and temperature factors (B-factors) during refinement of the atomic model. In maximum likelihood, the phases are adjusted to minimize the R-factor. In simulated annealing the potential energy of a model is perturbed and slowly

returned to equilibrium following a certain annealing schedule. Additional methods can improve the maps, for instance in this work TLS refinement was used. This is a macromolecular refinement in which the displacement of whole groups of atoms of the molecule (like domains) is parameterized as a combination of translation, libration and screw motion.

2. Materials and Methods

2.1. Molecular cloning

Standard polymerase chain reaction was performed in 50µl with 1x Phusion HF buffer, 200µM dNTPs, 0.5µM of forward and reverse primers, 3% DMSO, 50-100ng of DNA template and 1U Phusion DNA polymerase (New England BioLabs). The following program was used for amplification in an Eppendorf MasterCycler:

Initial denaturation: 98°C x 30sec

30 cycles: denaturation 98°C x 10sec – annealing 55°C x 1min – elongation 72°C x 30sec

Final elongation: 72°C x 10min

The PCR product was then purified using the QIAquick extraction kit (Qiagen). After restriction enzyme digestion using New England Biolabs enzymes and protocols, the fragments were separated and purified from agarose gels with the QIAquick extraction kit (Qiagen). The target vector was digested in the same way and also purified via gel-extraction. Ligation of the PCR product into the vector was carried out using T4 DNA Ligase (New England BioLabs) according to the manufacturer protocol. After transformation into *Escherichia coli* (*E. coli*) TOP10F' and growth at 37°C over night, plasmid DNA was prepared using the QIAcube (Qiagen) with the corresponding QIAprep Spin Miniprep Kit (Qiagen). The quality of the final plasmid vector containing the target insert was checked by restriction enzyme digestion and DNA sequencing. For agarose electrophoresis and visualization of DNA, typically 1% (w/v) agarose was dissolved by heating in 1x TAE buffer (Appllichem) and gels were prepared using a Sub-Cell GT electrophoresis system (Bio-Rad). After mixing with 6x DNA loading dye (Fermentas), the samples were loaded and electrophoresis was carried out at a constant voltage of 110V in 1x TAE buffer (Appllichem). Either a 1kb or 100bp DNA ladder (peqlab) was used as a size reference. DNA was visualized with SYBRR Safe DNA gel stain (Life Technologies), followed by imaging on a gel-doc system (Bio-Rad).

Human FGFR4 was obtained by GENEART, corresponding to human FGFR4 isoform 1 in UniProtKB P22455. The used vectors are pFastBacHTB (Life Technologies) and its derivatives generated at

Proteros Biostructures. Many FGFR4 length variants were cloned in vectors containing different protein-tags. A list is reported and explained in “Results” at page 31.

The successful clone coded for amino acids 447-753 of human FGFR4 and was cloned into pFastBacHTB vector containing a hexahistidine-tag and TEV (Tobacco Etch Virus) protease cleavage site. FGFR4-V550E mutant was generated with QuikChange II Site-Directed Mutagenesis Kit (Agilent).

2.2. Insect cell culture and heterologous protein expression

For the expression of target protein in insect cells, viral DNA was generated using the Baculovirus Expression Vector System (Life Technologies) following the instructions given by the supplier. For recombinant protein expression, *Spodoptera frugiperda* cells (Sf9) were grown in Grace's Insect Media (Gibco-Life, Technologies) supplemented with Fetal Bovin Serum (Sigma Life Science) and infected at a cell density of $2-2.6 \times 10^6$ cells/ml (vitality 95-98%) with an MOI of 2.5 (Virus Counter InDevR). 64h after infection, cells were harvested by centrifugation at 900g for 10min and shock-frozen at -80°C .

Initially, test expressions were performed in 100ml cultures which were analyzed using a NiNTa spin column kit (Qiagen) for HIS-tagged target proteins and quantified with SDS-page and western blot analysis. Test expressions were repeated until a good sub-clone was identified (in terms of protein yield) and different parameters were tested to improve expression yields (MOI, post-infection time, cell density).

2.3. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis analysis (SDS-PAGE)

To denature protein samples, they were boiled in 1% sodium dodecyl sulfate (SDS)-loading buffer for 5min at 95°C . Subsequently samples were loaded on gels purchased from Life Technologies (NuPAGER NovexR 10% Bis-Tris Midi Gel). Electrophoresis was performed at 190V until the running front reached the boarder of the gel. Prestained Protein Marker VI (AppliChem) was used as a size reference covering 10-245 KDa. The gel running buffer was either 1x NuPAGER MES SDS running buffer (Life Technologies). To visualize proteins, the SDS-PAGE gels were soaked in the Coomassie based staining solution InstantBlue™ (Biozol) at room temperature (RT) for 30min. InstantBlue™ protein-staining solution then was removed and substituted with water. Imaging was performed on a gel-doc system (Bio-Rad).

2.4. Western blot analysis

After SDS-PAGE analysis, Western Blotting was performed using the iBlot™ Dry Blotting System (Life Technologies) with the proper iBlot Transfer Stack, nitrocellulose (Life Technologies) according to the manual provided by the distributor. The anti-HIS₆-Peroxidase (Roche) was used as antibody and visualized using “BM Blue POD substrate, precipitating” (Roche).

2.5. Determination of DNA and protein concentration

Protein and DNA-concentrations were determined using a NanoDrop Spectrophotometer (ND-1000 spectrophotometer, peqlab). In order to calculate the concentrations of the protein samples of the different constructs, the respective extinction coefficients and theoretical molecular weights (MW) were calculated from primary proteins sequences using the ProtParam online tool⁸⁵.

The extinction coefficient of $39420 \text{ M}^{-1}\text{cm}^{-1}$ with Abs 0.1% (=1 g/l) equaling 1.149 was used for quantification.

2.6. Purification protocol

In general, all the purification steps were performed at room temperature (about 22°C). In case of protein precipitation events, the protein was cooled and kept on ice. All chromatography buffers were prepared and the pH adjusted at room temperature. Subsequently they were filtered (0.22 µm) and degassed. The purification steps were performed on chromatography systems and columns obtained from GE Healthcare (ÄKTA system). Before the protein was pooled, samples were analyzed on pre-cast SDS-gels (10%) obtained from Life Technologies. Many buffers systems and columns were tested, and the buffers recipes here reported refer to combination “purification-crystallization” which finally led to solve the structures.

Pellet was thawed in buffer A (50mM Hepes, pH=8, 300mM NaCl, 10% glycerol, 0.3mM CHAPS, 20mM imidazole, 1mM TCEP) supplemented with complete-EDTA free proteases inhibitors and DNase I (Roche). Cells were lysed using a homogenizer stirrer and soluble protein was recovered after centrifugation at 10000g for 30min at 10°C. 5mM NiSO₄ were added to the supernatant and the lysate was incubated for 2 hours with Ni-NTA Agarose beads (Qiagen). The beads were recovered by centrifugation at 500g for 2min, resuspended in Buffer A and packed in an XK16/20 column. The

column was washed with buffer A and bound protein was eluted with a gradient 0-100% of buffer B (50mM Hepes pH=8, 200mM NaCl, 10% glycerol, 0.1mM CHAPS, 300mM imidazole, 1mM TCEP). The His-tag was removed using TEV protease in a dialysis step run overnight at room temperature. The ratio of FGFR4/TEV (w/w) was 20:1. The next day a second passage over the Ni-NTA column was performed and the flow-through of the column was collected. FGFR4 protein was further purified on a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare) in buffer C (10mM Hepes pH=8, 100mM NaCl, 3% glycerol, 1mM TCEP) and concentrated for crystallization. To estimate the MW of proteins and complexes, the columns were calibrated with the Low and the High Molecular Weight Gel Filtration Calibration Kit (Amersham Pharmacia Biotech), which uses proteins in the range between 13.7 kDa and 669 kDa.

2.7. Phosphorylation and dephosphorylation – Mass spectrometry analysis (MALDI)

Dephosphorylated FGFR4 was generated after the second affinity chromatography by adding λ -phosphatase (produced in house) and 2mM $MnCl_2$. The ratio used FGFR4/ λ -phosphatase was 60:1. The reaction was incubated 1h at room temperature (about 22°) under mild agitation. λ -phosphatase was then removed from the reaction mixture by size-exclusion chromatography performed with HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare) in buffer C.

Double-phosphorylated FGFR4 was obtained after purification and before concentration through the incubation of the protein sample with 10mM ATP and 5mM $MgCl_2$ for 3h at 20°C at neutral pH and under mild agitation. Excess ATP was removed by size-exclusion chromatography performed with HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare) in buffer C.

The phosphorylation state was checked by Matrix-assisted laser desorption/ionization (MALDI), a mass spectrometry application at the Protein Core Facility of the Max Planck Institute of Biochemistry in Munich.

2.8. Addition of inhibitors

For the Ponatinib- and Dovitinib-FGFR4 structures, the buffers for the second affinity chromatography were supplemented with 5mM $MgCl_2$ and 1mM ATP.

After size-exclusion chromatography, Ponatinib (AP24534) was added to the protein mixture in a 1:1 molar ratio and the mixture was incubated for 10min at 4°C. Precipitated protein (app. 1/3rd of initial protein) was removed by centrifugation and the protein was concentrated for crystallization. For the Dovitinib-FGFR4 structure, after size-exclusion chromatography the protein was concentrated and incubated with 2mM Dovitinib (TKI258) for 10min at 4°C before crystallization.

2.9. Protein concentration steps

In order to concentrate protein samples after intermediate and final purification steps, centrifugal filter devices (Amicon® Ultra and Sartorius Vivaspin 500) with a 30KDa nominal molecular weight limit were used as described in the provided protocol. Concentration was conducted up to the desired volume for intermediate purification and up to the desired protein concentration in final concentration steps.

2.10. Phosphorylation western blot

Homogeneous non-phosphorylated FGFR4 was generated with λ -phosphatase produced in-house as described above and analyzed by mass spectrometry. The phosphorylation reaction was performed in the presence of 10mM MgCl₂ and 5mM ATP at constant 20°C. Samples were taken before the reaction and up to 120minutes. All the samples were from the same reaction and two samples were taken for every time point. The reaction was immediately stopped by heating the samples at 95°C for 15 minutes, and supplementing them with the addition of 2mM DDT and 3x Sample Loading Buffer for the next electrophoresis. We used antibodies against p-Tyr653 and p-Tyr654 of FGFR1 (Emelca Bioscience), exploiting the highly conserved sequence (98%) of the activation segment between FGFR4 and FGFR1. Tyr653 of FGFR1 corresponds to Tyr642 and Tyr654 to Tyr643 in FGFR4. The western blot was performed with PVDF membranes (Life Technologies) and BSA was used as background control, while non-phosphorylated and double-phosphorylated FGFR4s were used as negative and positive controls, respectively. The blot follows a standard protocol with a first phase with BSA incubation, a second phase with primary antibodies incubation and a third phase with secondary conjugated antibody incubation. The development phase was carried out and visualized with BM Blue POD substrate (Roche).

2.11. Thermofluor analysis

Thermofluor assay⁸⁶ was used to determine protein-unfolding curve of FGFR4 wild-type and mutant V550E. The final solution contains purified protein at 1mg/ml, 100x Sypro Orange. For the samples containing ATP-analogue, the solution was supplemented with 5mM AMP-PNP and 10mM MgCl₂. The gradient temperature is from 25 to 90°C with recorded signals every 0.5°C. Every condition was repeated twice.

2.12. Protein crystallization

Fresh FGFR4 kinase domain was crystallized at 4°C with a concentration between 6-25mg/ml from different conditions using various salts as precipitants, especially anions. Apo-1-FGFR4 in crystal form I crystallized from 0.6M NaH₂PO₄, 0.6M KH₂PO₄, 0.075M NaHEPES pH=7.5 and 24% glycerol. Apo-2-FGFR4 in crystal form II supplemented with 5mM AMP-PNP and 10mM MgCl₂ crystallized in 2.5M C₂H₃NaO₂ and 0.1M MES pH=6.5. The FGFR4-Ponatinib cocrystal was grown from 0.4M NH₄H₂PO₄ and FGFR4-Dovitinib crystal from 0.1M Citric acid pH=4.0, 1.6M (NH₄)₂SO₄, final pH=5. For the glycerol condition, the reservoir was directly used for cryo-protection. For the FGFR4-Dovitinib crystal, the cryo-solution was composed by reservoir with the addition of 20% glycerol, while for the other two crystallization conditions, cryo-solutions were prepared using fresh reservoir supplemented with 20% ethylene glycol. Crystals were transferred into the cryo-solution for a few seconds and shock-frozen in liquid nitrogen.

2.13. X-ray data collection, data processing and structures solution

Data sets were collected at beamline X06SA of the Swiss Light Source (Table 2 in Results) and processed with XDS⁸⁷ and XSCALE⁸⁷. A model of the FGFR4 kinase domain was generated in SWISS-MODEL^{88,89} and used to solve the first apo-FGFR4 structure with MOLREP⁸⁰. All structures were rebuilt manually in COOT^{90,91} and refined using REFMAC^{81,83}. All the structures miss the trans-activation loop, which is about from Pro568 to Pro583 residues in the C-lobe. Solvent/water molecules were modeled at stereochemically reasonable sites. Figures were generated in PYMOL⁹² and edited with GIMP⁹³.

2.14. PDB accession numbers

The coordinates and structure factors of the structures described have been deposited at the wwPDB under accession code: Apo-1-structure 4TYE, Apo-2-structure 4TYG, Dovitinib-structure 4TYI and Ponatinib-structure 4TYJ.

3. Results

3.1. Cloning and expression of FGFR4-KD gene

Different variants of the human fibroblast growth factor 4 kinase domain were cloned, amplified and expressed as described in Methods. Through mutagenesis, the most known rhabdomyosarcoma mutants N535K and V550E were also generated (Table1).

```

YVQVLKTADI NSSEVEVLYL RNVSAEDAGE YTCLAGNSIG LSYQSAWLTV 350
LPEEDPTWTA AAPEARYTDI ILYASGSLAL AVLLLLLAGLY RGQALHGRHP 400
RPPATVQKLS RFPLARQFSL ESGSSGKSSS SLVRGVRLSS SGPALLAGLV 450
SLDLPLDPLW EFPRDR LVLG KPLGEGCFGQ VVRAEAFGMD PARPDQASTV 500
AVKMLKDNAS DKDLADLVSE MEVMKLIGRH KNIINLLGVC TQEGPLYVIV 550
ECAAKGNLRE FLRARRPPGP DLSPDGPRSS EGPLSFPVLV SCAYQVARGM 600
QYLESRKCIEH RDLAARNVLV TEDNVMKIAD FGLARGVHHI DYYKKTISNGR 650
LPVKWMAPEA LFDRVYTHQS DVWSFGILLW EIFTLGGSPY PGIPVEELFS 700
LLREGHRMDR PPHCPPELYG LMRECWHAAP SQRPTFKQLV EALDKVLLAV 750
SEEYLDLRLT FGPYSPSGGD ASSTCSSSDS VFSDHPLPLG SSSFPPFGSGV 800
QT 802

```

Fig. 10. FGFR4 amino acid sequence. The kinase domain sequence is highlighted in yellow by UniProt (P22455). In bold red is marked the successful clone 304, from residue 447 to 753.

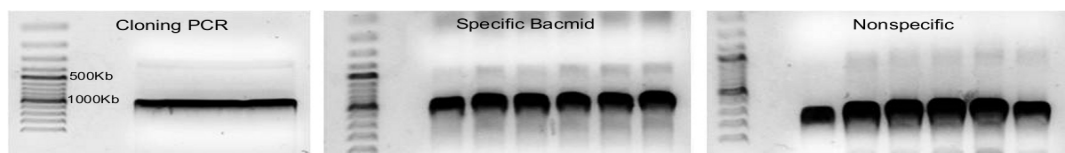


Fig. 11. Agarose gel electrophoresis of FGFR4 DNA, clone H. On the left, the initial PCR before digestion. The specific bacmid PCR is done with the primers of the initial PCR, while the nonspecific one is done with one specific primer set and a primer set that binds in the baculovirus genome. The result of the two bacmid PCRs must be coherent.

As listed below, at least 10 variant of FGFR4-KD were prepared and, when it was possible, tested until crystallization trials. In addition, to increase protein solubility and stability, protein tags such as GST or NusA were used with unsatisfying results. The best combination was achieved with the expression protocol reported in “Methods” and the shortest variant H in combination with the 6x histidine-tag, which speeded up the purification steps and significantly increased protein stability, eliminating protein precipitate. For these reasons, the next paragraph refers to variant H and its relative mutated clone J (Table 1).

Table 1. Human FGFR4-KD variants tested.

<i>Clone</i>	<i>FGFR4 residues*</i>	<i>Baculovirus Vector</i>	<i>Result</i>
Clone A	Hs. 447-762	pFastBac-6HIS-NUS-TEV-hFGFR4	Poor expression
Clone B	Hs. 447-802	pFastBac-6HIS-NUS-TEV-hFGFR4	Poor expression
Clone C	Hs. 442-753	pFastBac-6HIS-GST-TEV-hFGFR4	Contaminants could not be removed
Clone D	Hs. 442-802	pFastBac-6HIS-GST-TEV-hFGFR4	Contaminants could not be removed
Clone E	Hs. 391-802	pFastBac-6HIS-NUS-TEV-hFGFR4	Insoluble
Clone F	Hs. 442-802	pFastBac-6HIS-NUS-TEV-hFGFR4	Insoluble
Clone G	Hs. 391-802	pFastBac-6HIS-TEV-hFGFR4	Did not crystallized
Clone H	Hs. 447-753	pFastBac-6HIS-TEV-hFGFR4	Different structures solved
Clone I	Hs. 447-753/N535K	pFastBac-6HIS-TEV-hFGFR4	Not correct clone
Clone J	Hs. 447-753/V550E	pFastBac-6HIS-TEV-hFGFR4	Crystals require optimization

*refers to *Homo sapiens*. The kinase domain sequence is reported in Fig. 10.

6HIS = 6-histidine; TEV = Tobacco Etch Virus nuclear inclusion; GST = glutathione S-transferase; NUS = NusA, antitermination protein.

3.2. Purification of FGFR4-KD-wt and FGFR4-KD-V550E mutant

FGFR4-KD was purified in three column steps. A first affinity chromatography exploited the interaction between the His-tag of the construct and the Nickel ions of the resin and results in a pool of proteins able to bind the Nickel-matrix. A second affinity chromatography, called “negative” because in this case the target protein without the His-tag is contained in the flow through separates the target from proteins interacting with Nickel ions. In the third step, a size exclusion column separates proteins by their size (Fig. 12). Many trials pointed out different essential details to obtain high protein yield and purity. Some involved buffers optimization, especially the final buffer in which the protein is conserved before crystallization screenings. Adding an anion-exchange column to the purification procedure did not improve the final yields. The main improvement was the use of Ni-NTA Agarose beads (Qiagen) instead of Ni-Sepharose High Performance (GeHealthcare) and the incubation of the cell lysate for at least 2 hours with Nickel-beads supplemented with NiSO₄, preventing that proteins remove NiSO₄ from the beads. To reduce the presence of contaminants, it made a difference working at room temperature (about

22°C) instead of the usual 4°C: in fact higher temperature does not affect FGFR4-KD stability or solubility but decreases contaminants stability for wild-type protein. On the contrary, the few purifications of mutant FGFR4-KD-V550E (active site mutant) at room temperature showed a higher FGFR4-KD-V550E instability, as later confirmed by the thermofluor experiments.

The FGFR4-KD variant H has a molecular weight about 34.7kDa without the HIS-TEV sequence and 36.2kDa including the tag and protease cleavage site, and variant J has similar weights (absolute values counted using the human sequence plus cloning vector residues - ExPASy Compute pI/Mw tool⁸⁵ and mass spectrometry analysis). In size-exclusion chromatography, FGFR4-KD elutes as a monomer however, traces of dimers were found in SDS-page gels and on phosphorylation western blots. Dimers could depend either on the presence of external cysteine which form disulfide bonds or the physiological ability of the kinase to form dimers during autocatalytic activation/phosphorylation (Fig. 12c). To increase purity, fractions corresponding to the centre of the chromatographic peak of the size-exclusion chromatography were collected.

Although the protein purity was almost the same during every purification, the yield changed depending on expression culture quality. In general, from 5 liters of culture about 20mg of FGFR4-KD-wt are recovered with purity about 95%. However, FGFR4-KD is not homogenous and contains approximately 60% of non-phosphorylated protein mixed with single and double phosphorylated forms as explained later (Fig. 13a).

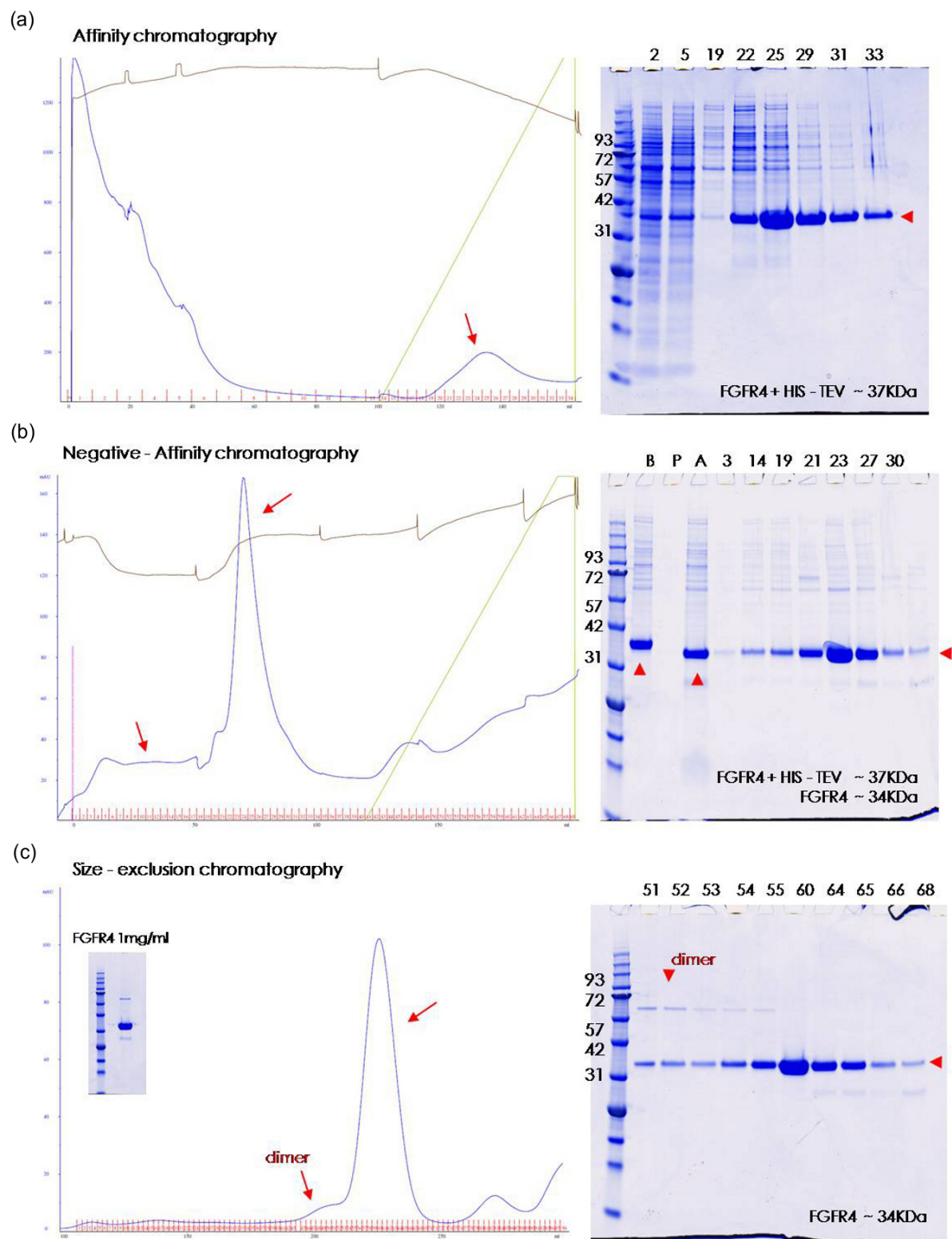


Fig. 12. FGFR4-KD-wt purification profiles and their relative SDS-page analysis. **(a)** First affinity chromatography. **(b)** Negative - affinity chromatography. (B: before tag-digestion; P: precipitate after tag-digestion; A: after dialysis). **(c)** size-exclusion chromatography.

3.3. Phosphorylation states and mass spectrometry analysis

At the beginning of the crystallization experiments, FGFR4-KD variant H did not crystallize even after many screenings. To collect more information on the protein and potentially resolve the problem, the homogeneity of the sample was analyzed by mass spectrometry.

It is known that FGFR-KDs have up to 7 tyrosines which undergo to auto/phosphorylation during their activation⁹⁴. Two of them are in the active site and are responsible for autoinhibition/auto activation mechanisms, in FGFR4-KD these are Tyr642 and Tyr643. In mass spectrometry analysis, we found that after the purification, FGFR4-KD-wt was pure with respect to SDS-PAGE analysis, but most of the protein was non-phosphorylated and parts of it in single- and double-phosphorylation states (Fig. 13a). It is possible to hypothesize that these two phosphorylation events occur at positions Tyr642 and Tyr643.

It had been reported that for the crystallization of FGFR2 with and without ATP the micro-homogeneity of the sample and its phosphorylation state⁹⁵ was essential, therefore a fully dephosphorylated FGFR4-KD was generated with λ -phosphatase/MnCl₂ (Fig 13b). In addition, experiments to produce a fully phosphorylated FGFR4-KD by incubation with ATP/MgCl₂ were performed. However, phosphorylating FGFR4-KD before the size-exclusion chromatography for 1h resulted in another heterogeneous sample, where the most dominant form is the single-phosphorylated one. Surprisingly, this form was more prone to crystallize, especially in combination with inhibitors. It is possible to reach an almost homogenous double-phosphorylated form with an incubation of 3h at 20°C. Interestingly, the presence of the double phosphorylated dimeric form in this sample was observed (Fig. 13c). This last form was never tested in crystallization screening but was used in western blot analysis.

It has been tried to phosphorylate FGFR4-KD-V550E, but the protein precipitates quickly even at lower temperatures, therefore, these experiments were stopped.

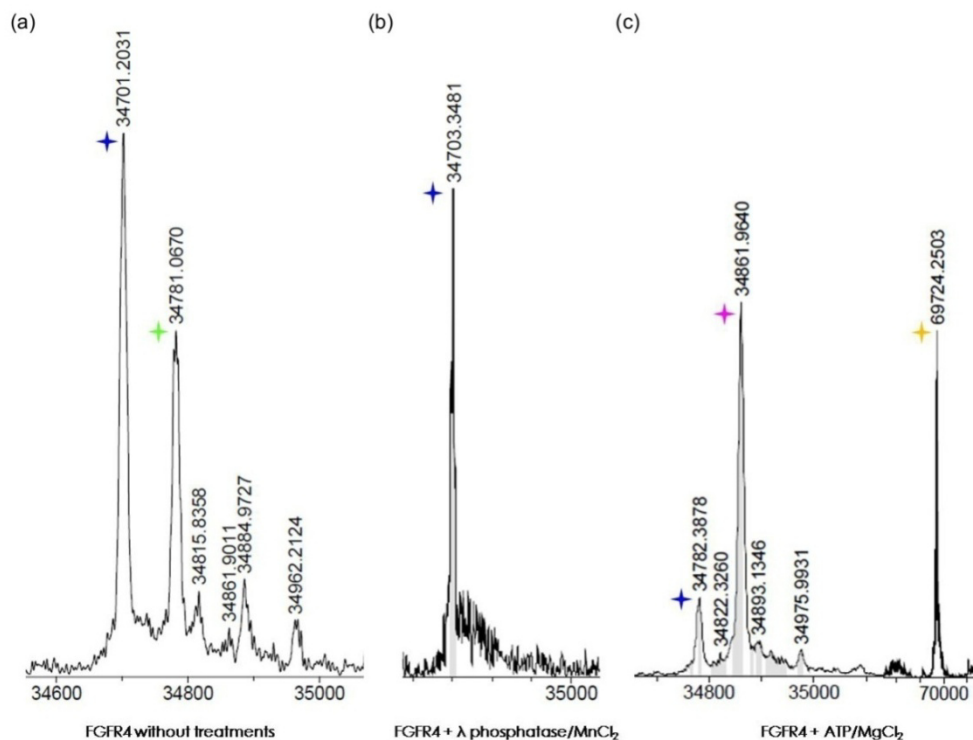


Fig. 13. Mass spectrometry analysis of FGFR4-KD-wt. **(a)** Heterogeneous sample after purification. **(b)** Homogenous sample after treatment with λ -phosphatase/ MnCl_2 . **(c)** Heterogeneous sample obtained after incubation with ATP/ MgCl_2 . Star colors legend: blue, non-phosphorylated; green, single phosphorylated; pink, double phosphorylated; yellow, double phosphorylated dimeric form.

3.4. Crystallization screenings

A large set of crystallization screenings were tested for initial screenings, containing combinations of salts, polyethylene glycol (PEG), 2-methyl-2, 4-pentandiol (MPD), alcohols and detergents. As described above, different constructs and also differently phosphorylated samples of these variants were tested. Concerning ligands (AMP-PNP, Dovitinib, Ponatinib and other here not mentioned) were tested in co-crystallization experiments.

Crystallization temperatures were tested (such as 4, 12 and 20°C) but positive hits happened only at 4°C, rarely at 12°C and precipitation was too fast at 20°C. Using frozen protein dramatically reduces chances to get crystals. Crystals of FGFR4-KD-wt in complex with AMP-PNP (ATP-analogue) were obtained but they are intergrown with multiple layers and could so far not be optimized. Also FGFR4-KD-V550E crystallized in tiny intergrown needles not yet optimized.

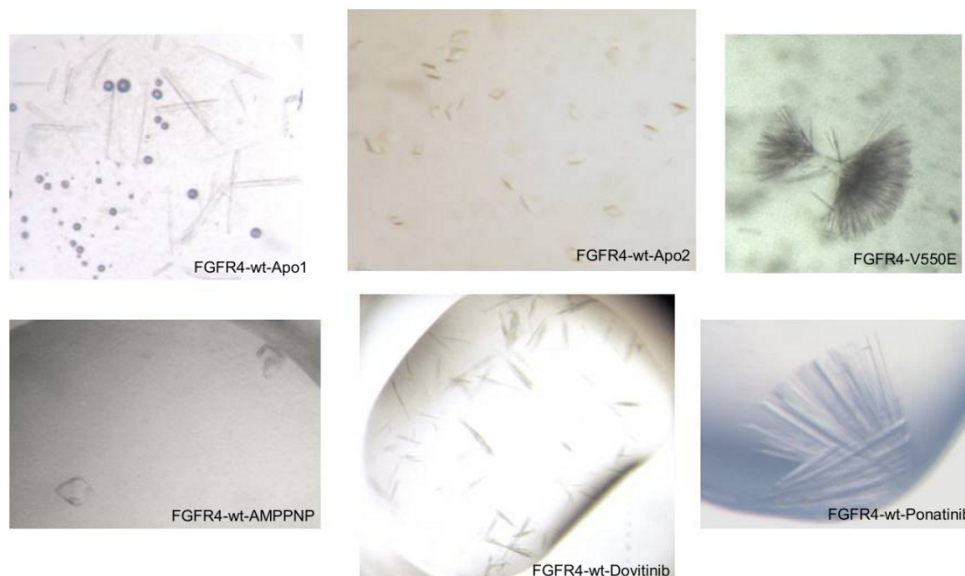


Fig. 14. Images of FGFR4-KD crystals.

3.5. Apo-FGFR4-KD structures

The apo-FGFR4-KD kinase domain crystallized from a mixture of mostly (about 60%) non-phosphorylated and singly and doubly phosphorylated protein as confirmed by mass spectrometry analysis. Two apo-FGFR4-KD structures in different crystal forms have been solved with molecular replacement and they share their main features (referenced as Apo-1 and Apo-2, respectively). As the Apo-1-FGFR4-KD structure crystallized in the same crystal form as the Dovitinib- and Ponatinib-structures (Table 2), the structural description is mainly of Apo-1.

Overall, FGFR4-KD-KD adopts the typical protein kinase structure composed by a smaller N-terminal and a larger C-terminal domain (Fig. 15). Both apo-FGFR4-KD structures are in DFG-in conformation and lack electron density indicative of phosphorylation of any tyrosine residue. The structures reveal an unusual autoinhibited conformation that has not been reported for members of the FGFR family so far: C-terminal to the DFG-motif, the activation segment (residues 630 – 659) protrudes upwards and reaches the tip of the P-loop in the N-lobe (Figs. 15 and 16). This contact involves Tyr642 and Tyr643, which become phosphorylated during activation of FGFR4 kinase⁹⁴.

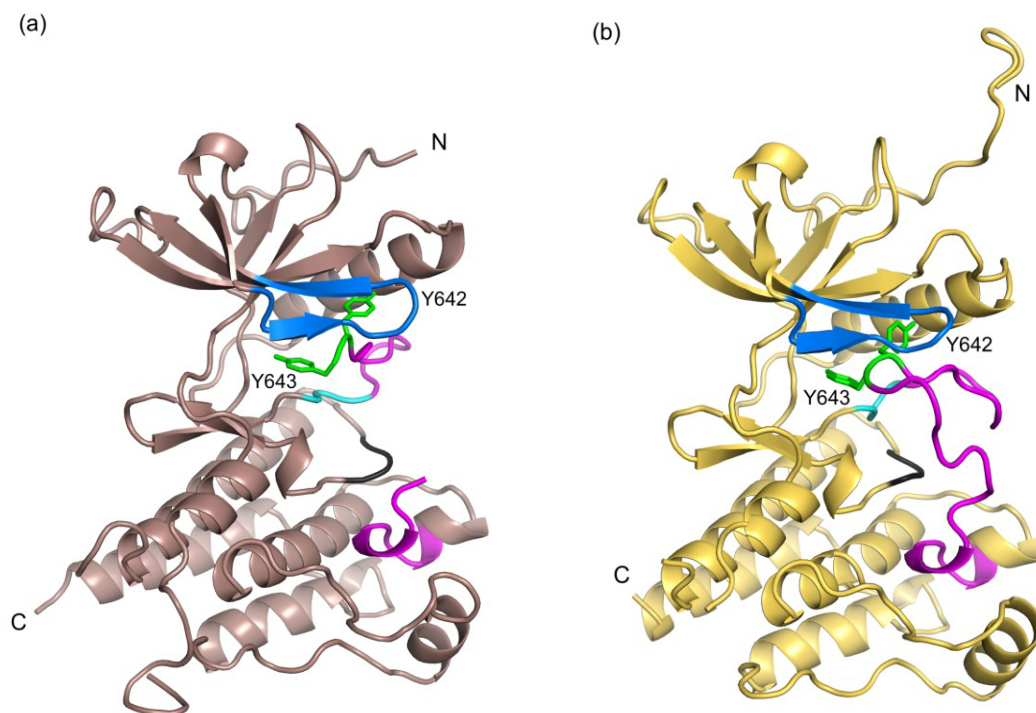


Fig. 15. Human apo-FGFR4-KD structures. **(a)** Apo-1-FGFR4-KD. **(b)** Apo-2-FGFR4-KD. The P-loop is in blue and the activation segment in magenta. The DFG-motif is colored in cyan and the HRD-motif in black. Tyr642 and Tyr643 are represented in sticks and highlighted in light green. The figures were generated after superimposition of the C-lobe and both the apo-structures are in the same orientation.

In the Apo-1-FGFR4-KD structure, the electron density for the activation segment is incomplete, starting from residue 644, thus allowing modeling of tyrosines 642-643. The side chain of Tyr643 is partially exposed to the solvent while the side chain of Tyr642 points towards the α C-helix and is coordinated by Phe478 from the P-loop and residues from the β 3-strand and the α C-helix which shield it from phosphorylation. The side-chains of Tyr643 and Asp641 are located in the ATP-binding pocket which, together with the displacement of the P-loop, prevents ATP binding (Fig. 16a).

The Apo-2-FGFR4-KD structure shows a slightly different conformation of the N-lobe (Fig. 16b), and superposition of the C-lobes of the two apo-structures indicates a rotational N-terminal domain motion of about 11.4 degrees (Fig. 17a). However the Apo-1 and Apo-2-structures belong to two different space groups and the rotation is about a different axis compared to the Apo-1/Dovitinib/Ponatinib-structures (see below). In the Apo-2-FGFR4-KD structure, the activation segment is completely defined by electron density and positioned closer to the P-loop. Here, Asp641 is out of the ATP-binding pocket, which is occluded by Tyr643 and partially by Tyr 642 (Fig 16b).

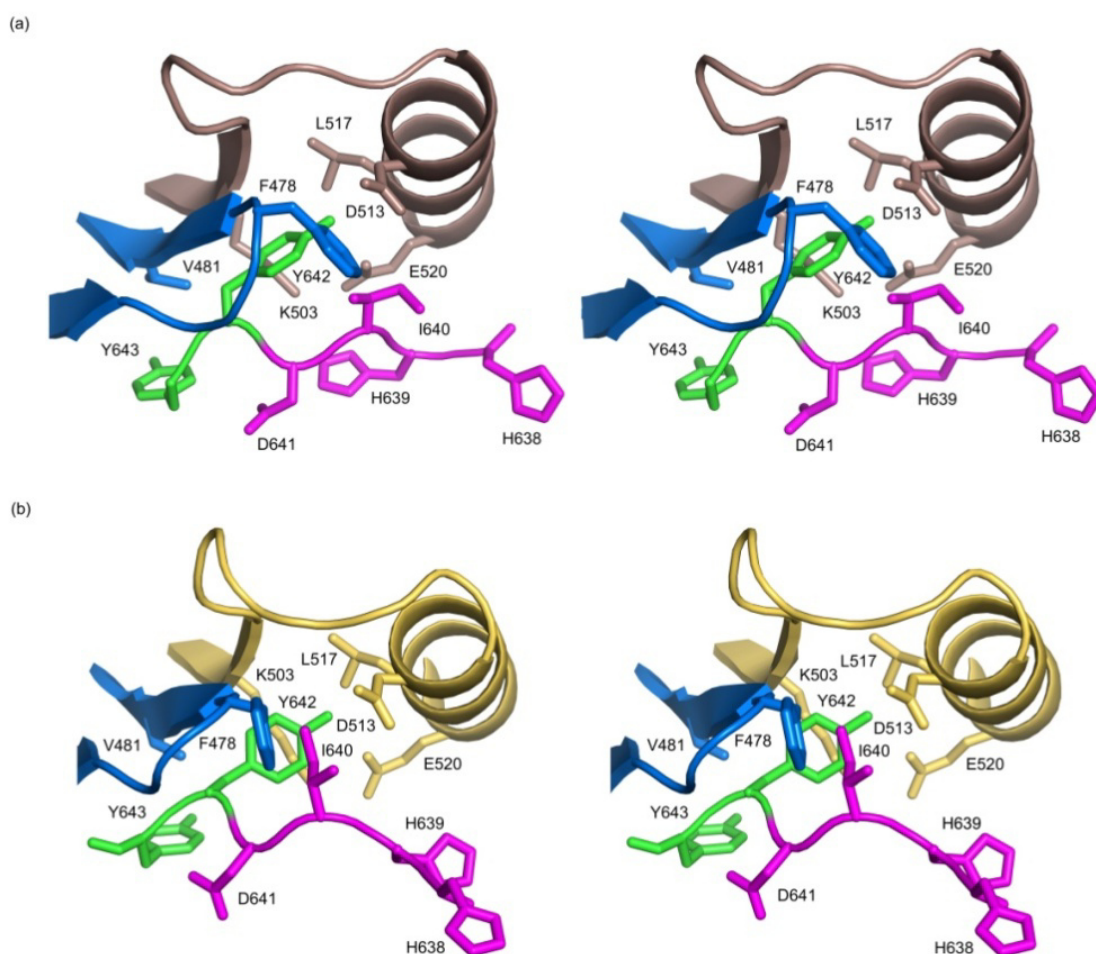


Fig. 16. Stereo close-up at FGFR4-KD N-lobe. **(a)** N-lobe of Apo-1-FGFR4-KD. **(b)** N-lobe of Apo-2-FGFR4-KD. The P-loop is in blue and the activation-segment in magenta. Tyr642 and Tyr643 are highlighted in light green. The figures were generated after superimposition of the C-lobe and both the N-lobes are in the same orientation.

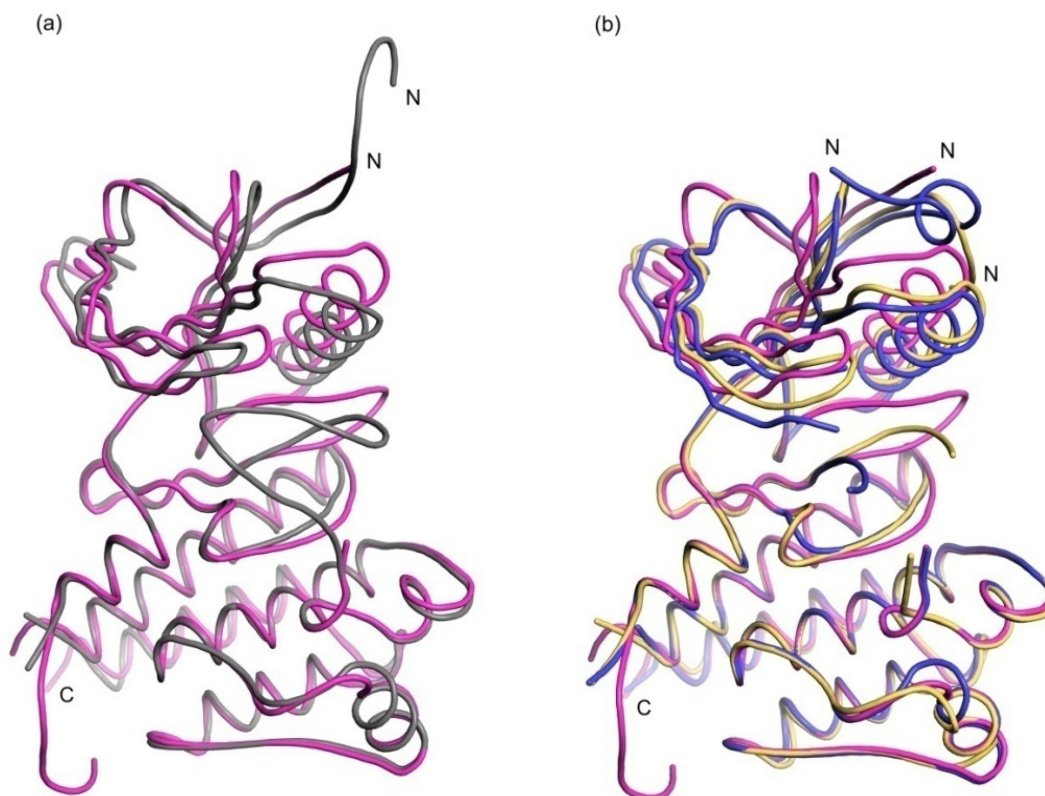


Fig. 17. Domain motions in FGFR4-KD structures. **(a)** Superposition of the C-lobe of Apo-1 and Apo-2-FGFR4-KD. Apo-1 is colored in magenta and Apo-2 in dark grey; they show a rotation of 11.4° . **(b)** Superposition of the C-lobe of Apo-1-, Dovitinib- and Ponatinib-FGFR4-KD structures. Dovitinib-structure is in yellow, while Ponatinib-structure in blue. Both inhibitors affect the N-lobe rotation leading to its closure upon the C-lobe. The rotation of Apo-1 relative to the Dovitinib-structure is 8.8° and 11° to the Ponatinib-structure. The figures were generated after superimposition of the backbone of the C-lobe (residues 659-749), ligands are removed and both (a) and (b) figures are in the same orientations.

A conformation of the activation segment similar to the apo-FGFR4-KD structures was observed for non-phosphorylated inactive c-MET and attributed to autoinhibition⁹⁶ (Fig. 18 - Table 3). However, our FGFR4-KD-structures differ from c-MET by the presence of the salt bridge between Lys503 and Glu520 from the α C-helix (Fig. 16), generally conserved in active kinase.

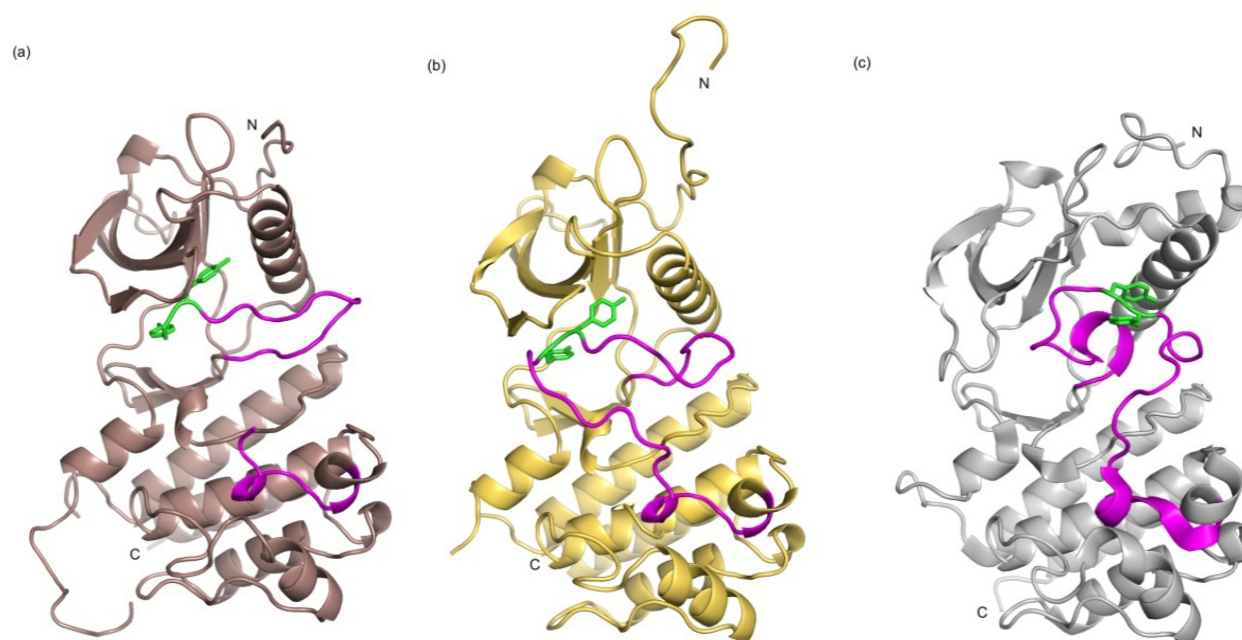


Fig. 18. Comparison of human apo-FGFR4-KD structures with the c-MET kinase structure. **(a)** Apo-1-FGFR4-KD. **(b)** Apo-2-FGFR4-KD. **(c)** c-MET kinase. The activation-segment in magenta while Tyr642 and Tyr643 are in represented in stick and highlighted in light green. The figures were generated after superimposition of the C-lobe.

Table 2. Crystallographic parameters for human FGFR4-KD structures

	<i>Apo-1-</i> <i>FGFR4-KD</i>	<i>Apo-2-</i> <i>FGFR4-KD</i>	<i>Dovitinib-</i> <i>FGFR4-KD</i>	<i>Ponatinib-</i> <i>FGFR4-KD</i>
<i>Data collection</i>				
Space group	P4 ₃ 2 ₁ 2	H32	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Resolution (Å)	58.74 – 2.8 (3.02 – 2.8) ^a	79.7 – 2.4 (2.56 – 2.4)	61.7 – 3.4 (3.69 – 3.4)	62.03 – 2.4 (2.65 – 2.4)
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.9 62.9 184.9	159.4 159.4 92.2	65.9 65.9 182.2	66.2 66.2 176.0
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90 90 90	90 90 120	90 90 90	90 90 90
R _{merge}	12.7 (85.4)	8.9 (73.8)	8.7 (83.1)	9.1 (85)
<i>I</i> / <i>σ</i> <i>I</i>	12.5 (2.3)	12.1 (2.7)	10.4 (1.7)	17.34 (3.0)
Completeness (%)	99.6 (99.2)	98.7 (99.6)	96.4 (96.8)	99.9 (100)
Total observations	58,867	74,629	15,344	158,760
Total unique observations	9,467	17,454	5,806	15,173
<i>Refinement</i>				
R _{cryst} /R _{free}	22.3 / 27.8	20.2 / 23.8	27.7 / 34.0	21.4 / 24.7
N.of reflection ^b	8742 (725)	16365 (1092)	5427 (379)	14070 (1101)
RMS bonds (°)	0.007	0.007	0.007	0.007
RMS angles (°)	1.272	1.033	1.025	1.227
No. atoms				
Protein ^c /water	2302 / 25	2324 / 80	2176 / 0	2184 / 56
Ligand/solvent	0 / 10	0 / 10	29 / 0	39 / 14
Average B-factors				
Protein/water	60.2 / 38.7	60.8 / 53.9	110.380 / 0	53.3 / 55
Ligand			127.7	44.7
Ramachandran plot				
Most favored geometry (%)	89.7	93.2	86.7	92.7
Additionally allowed (%)	9.9	6.4	12.0	6.9

^aValues in brackets refer to the highest resolution shell; ^bValues in brackets refer to free R-value test set;

^cElectron density gaps are listed in the methods.

Table 3. List of structures mentioned

<i>Protein structure</i>	<i>PDB code</i>	<i>Reference</i>
Apo-1-FGFR4-KD	4TYE	
Apo-2-FGFR4-KD	4TYG	
Dovitinib - FGFR4-KD	4TYI	
Ponatinib - FGFR4-KD	4TYJ	
Non-phosphorylated - c-MET	2G15	Wang W. et al., 2006
Non-phosphorylated - FGFR1*	1FGK	Mohammadi M. et al., 1996
Phosphorylated - FGFR1	3GQI	Bae J.H. et al., 2009
Non-phosphorylated - FGFR2*	2PSQ	Chen H et al., 2007
Phosphorylated - FGFR2	2PVF	Chen H et al., 2007
Non-phosphorylated - FGFR3 - K650E	4K33	Huang Z. et al., 2013
Ponatinib - BCR-ABL	3OXZ	Zhou T. et al., 2011
BCR-ABL - T315I	3IK3	O'Hare T. et al., 2009
Ponatinib - DDR1	3ZOS	Canning P. et al., 2014
Ponatinib - RIPK2	4C8B	Canning P. et al., unpublished
4-(aminoalkylamino)-3-benzimidazole -quinolinones - CHK1	2GDO	Ni Z.J. et al., 2006

*Since there are many FGFR1 and 2 structures, we limited our comparison to selected examples.

3.6. Phosphorylation western blot and thermofluor analysis

To get more information on the connection between FGFR4-KD apo-structures, autoinhibition, and activation, we determined the sequence of the phosphorylation events in the activation segment of FGFR4-KD. Therefore, a phosphorylation western blot of tyrosines 642 and 643 was performed. We used antibodies against p-Tyr653 (Tyr642 in FGFR4-KD) and p-Tyrs654 (Tyr643 in FGFR4-KD) of FGFR1, exploiting the highly conserved sequence (98%) of the activation segment between FGFR4 and FGFR1. Homogeneous non-phosphorylated FGFR-KD was generated with λ -phosphatase and analyzed by mass spectrometry. The phosphorylation reaction was followed for up to 2 hours in presence of ATP and $MgCl_2$. The result of the western blot indicates that Tyr643 and Tyr642 are phosphorylated at similar rates, with bands appearing after 3 and 5 minutes respectively, indicating a slightly faster rate for Tyr643 (Fig.19).

To characterize our FGFR4-KD variants further, we used thermofluor analysis⁸⁶. Additionally, it has been reported that mutant V550E is an activating mutation in cells⁶⁸. We expressed and purified a variant of the same length as the wild-type carrying the Glu550 mutation. Thermofluor assay was used

to determine protein-unfolding of FGFR4-KD wild-type and mutant V550E with and without ATP-analogue (Fig. 20). FGFR4-KD wild-type with and without AMP-PNP/MgCl₂ shows unfolding at 46.5° and 47.5°C respectively, whereas FGFR4-KD-V550E unfolds at 38.5°C in absence and 43.5°C in presence of AMP-PNP/MgCl₂, indicating ATP-binding potential. Higher concentrations of AMP-PNP up to 25mM have been tested and increase the melting temperature for both variants (wt and mutant) by additional 2-3 degrees.

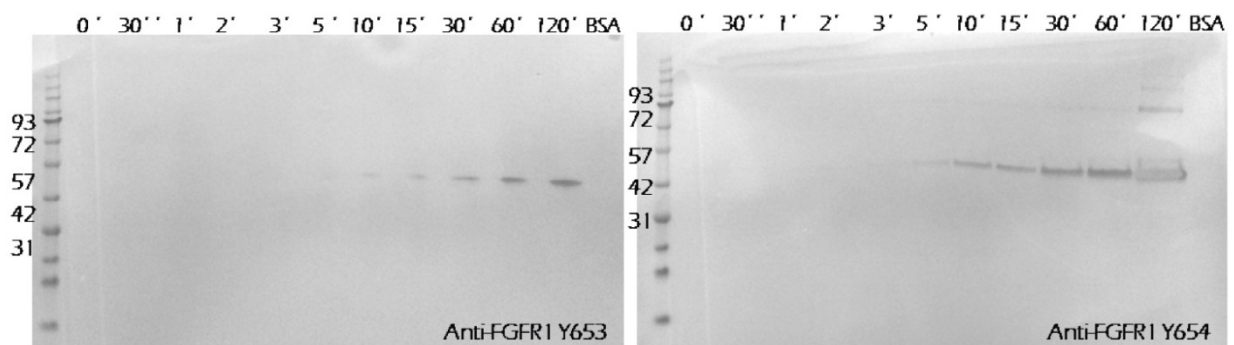


Fig. 19. Western blot showing a time course of the phosphorylation of FGFR4 Tyr642 and Tyr643.

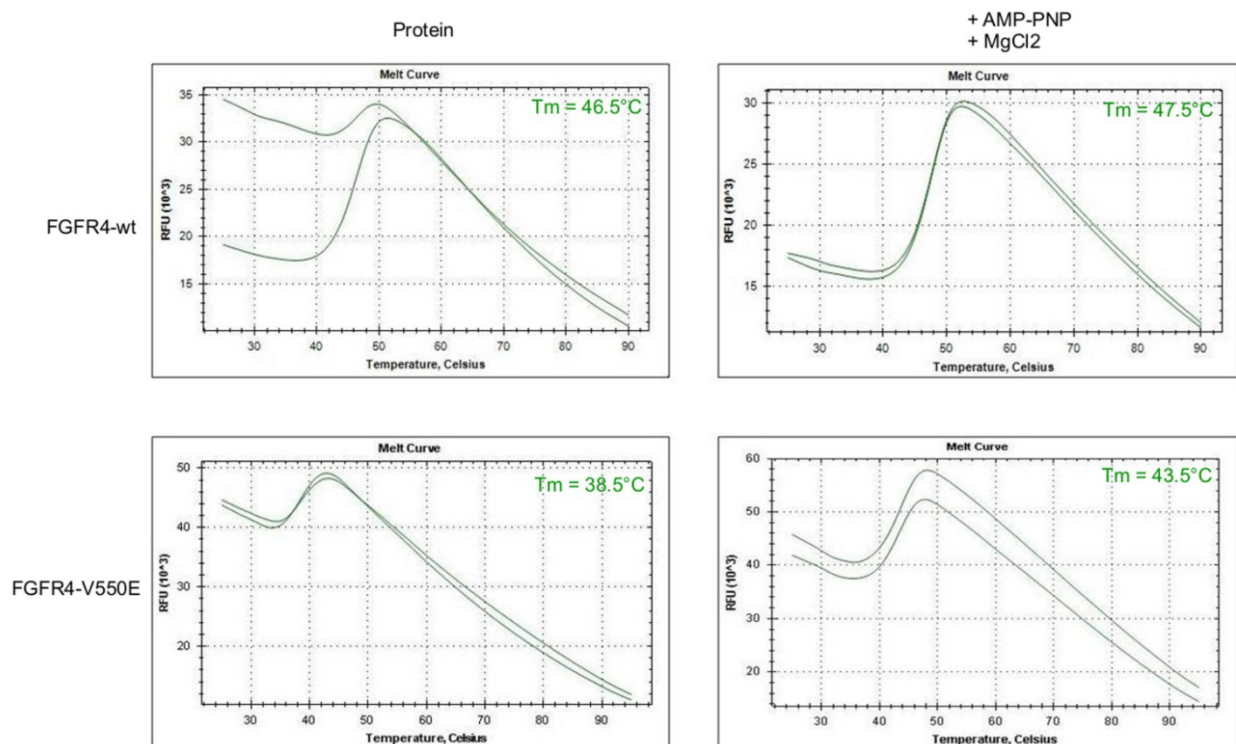


Fig. 20. FGFR4-KD unfolding curves. Thermofluor analysis was used to calculate FGFR4-wt and V550E mutant melting temperature, with and without ATP analogue (AMP-PNP/MgCl₂). Celsius degrees are reported on the X-axis and relative fluorescence units (RFU) are on the Y-axis.

3.7. Dovitinib-FGFR4-KD structure

As in the apo-forms, the DFG-in conformation of FGFR4-KD is seen in complex with the ATP-competitive tyrosine kinase inhibitor Dovitinib. Dovitinib shows activity against FGFRs1-4, VEGFRs1-3 and other RT-kinases^{75,76}. The Dovitinib-FGFR4-KD structure has been solved at intermediate resolution (3.4Å - Table 2) and the piperazine group of the inhibitor is not clearly defined by electron density. However, it is possible to build a Fem-map of the inhibitor through the PHENIX application FEM (Feature Enhanced Maps), resulting in an optimized density map. (Fig. 21c).

The binding mode agrees well with the structure of the CHK1 kinase in complex with 4-(aminoalkylamino)-3-benzimidazole-quinolinone⁹⁷, a compound that has a similar core-structure (Table 3). The quinolone group of Dovitinib binds in the ATP-binding pocket of FGFR4-KD and is stabilized by hydrophobic contacts with the gatekeeper residues Val550, Val481 and Leu619 (Fig. 20). The backbone atoms of Ala553 form hydrogen bonds with the carbonyl oxygen of the quinolone group and the nitrogen of the benzimidazole group. FGFR4-KD Dovitinib crystallized in the same space group as Apo-1- and Ponatinib structures, and a superposition of their C-lobes highlights inhibitor dependent domain motions (Fig. 17b). Dovitinib causes N-lobe closure to the C-lobe with a rotation of 8.8 degree compared with Apo-1-FGFR4-KD structure. The P-loop covers the active site and the activation segment is displaced from its autoinhibitory position by the ligand (the activation segment is not defined by electron density from Gly637 to Arg650).

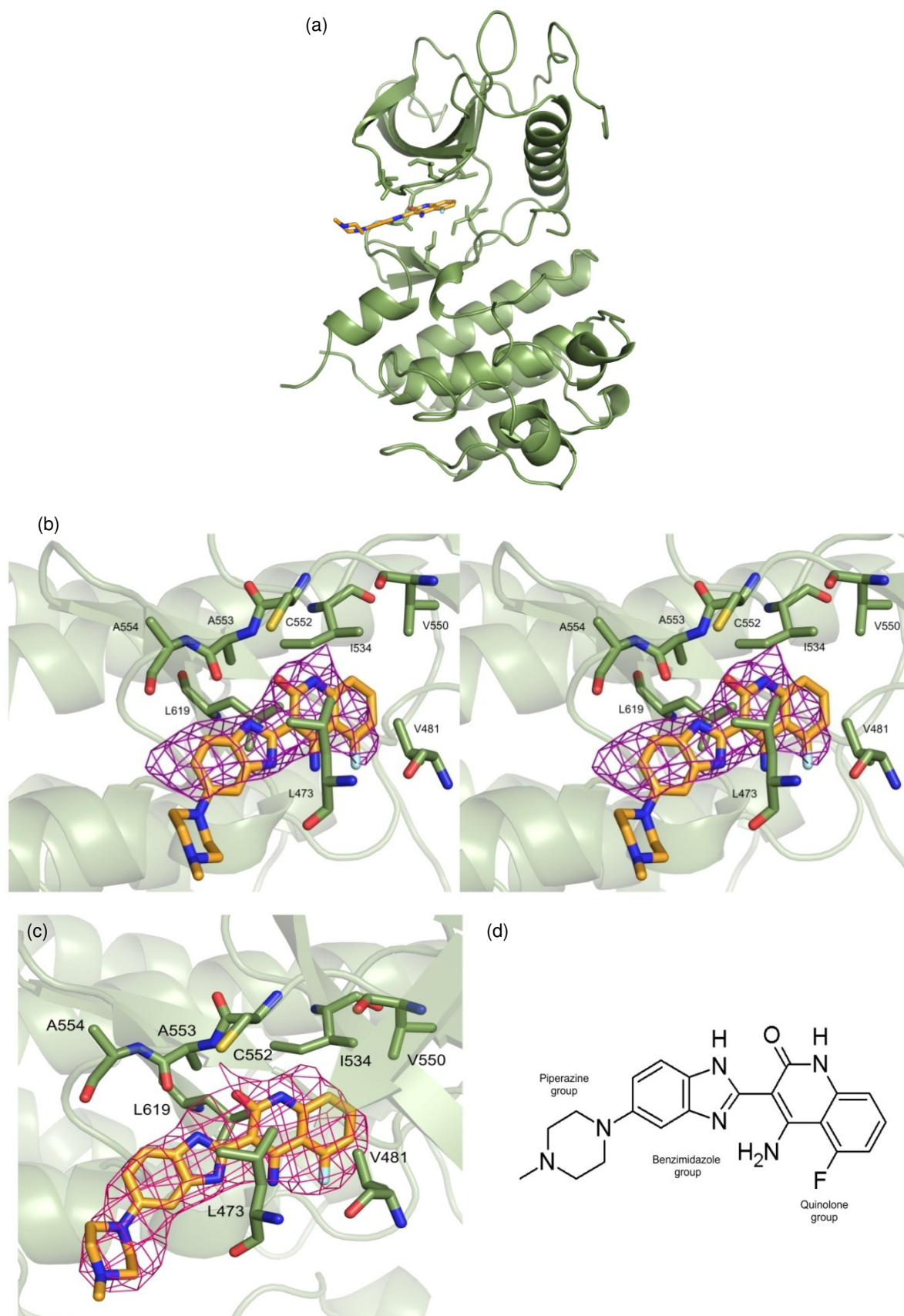


Fig. 21. Dovitininib-FGFR4-KD structure. **(a)** Overall structure of Dovitininib-FGFR4-KD. **(b)** Stereo close-up of FGFR4-KD active-site with Dovitininib bound. Fo-Fc omit map of Dovitininib FGFR4-KD complex with a contour level of 2.5σ. **(c)** Fem-map of Dovitininib FGFR4 complex at 1.5σ, generated with Phenix-FEM. **(d)** Drawing of Dovitininib inhibitor. Atoms are colored according to atom type: carbon, orange; nitrogen, blue; oxygen, red; fluor, cyan.

3.8. Ponatinib-FGFR4-KD structure

We also cocrystallized FGFR4-KD in complex with Ponatinib, a potent inhibitor of FGFR-kinases^{73,98,74}. Ponatinib belongs to the third generation of kinase inhibitors and had been designed to overcome resistance of the BCR-ABL gatekeeper mutant T315I⁹⁹. FGFR4-KD-Ponatinib cocrystals diffract to 2.4Å. With respect to the other FGFR4 structures, the N-terminus from Ser446 to Asp453 forms an additional α -helix. The FGFR4 kinase-Ponatinib complex exhibits the DFG-out conformation, a specific inactive conformation that has not been observed in other structures of FGFR kinase domains (Fig. 22) so far. Ponatinib binds to the hinge region of FGFR4-KD in the same way as in DDR1¹⁰⁰ (Table 3) and stretches along the ATP-binding cleft into the so-called allosteric binding pocket generated by a flip-out of Phe631 located in the DFG-motif. The imidazo-pyridazine group makes apolar contacts with the hinge region (Cys552) and the FGFR4-KD-gatekeeper Val550 is stabilized by the benzimide group. Bisarylamide derivatives, as Ponatinib, form a typical double H bond in the active site: in this case, it is established between Glu520 from the α C-helix and the benzimide nitrogen and carbonic oxygen and Asp630 from the DFG-motif. The trifluoro-amine group is involved in apolar interactions within the deep pocket of FGFR4-KD and the terminal piperazine moiety of Ponatinib is exposed to solvent. As mentioned above, Ponatinib induces a domain motion of the N-lobe, and superposition of the C-lobe shows a rotation of 11 degrees with respect to the Apo-1-FGFR4-KD structure (Fig. 17b).

Crystal structures of RIPK2, DDR1 and ABL in complex with Ponatinib (Table 3) are available, and RIPK2 and DDR1 superpose well with our FGFR4-KD structure, whereas ABL and its mutant T315I shows a slightly distorted inhibitor conformation, especially around the trifluoro-amine group. Moreover, an interesting difference with FGFR4-KD is that ABL-Tyr253, located in the P-loop, together with Phe317 and Phe382 locks Ponatinib in a hydrophobic network which has been reported to be essential for Ponatinib efficacy. In DDR1 Arg789, from the activation segment, is in the same position as Tyr253 of ABL, indicating a similar locking function. However, in FGFR4-KD the P-loop has a more elongated conformation with respect to the ABL-T315I mutant and Arg635, corresponding to DDR1 Arg789, is far away from Ponatinib.

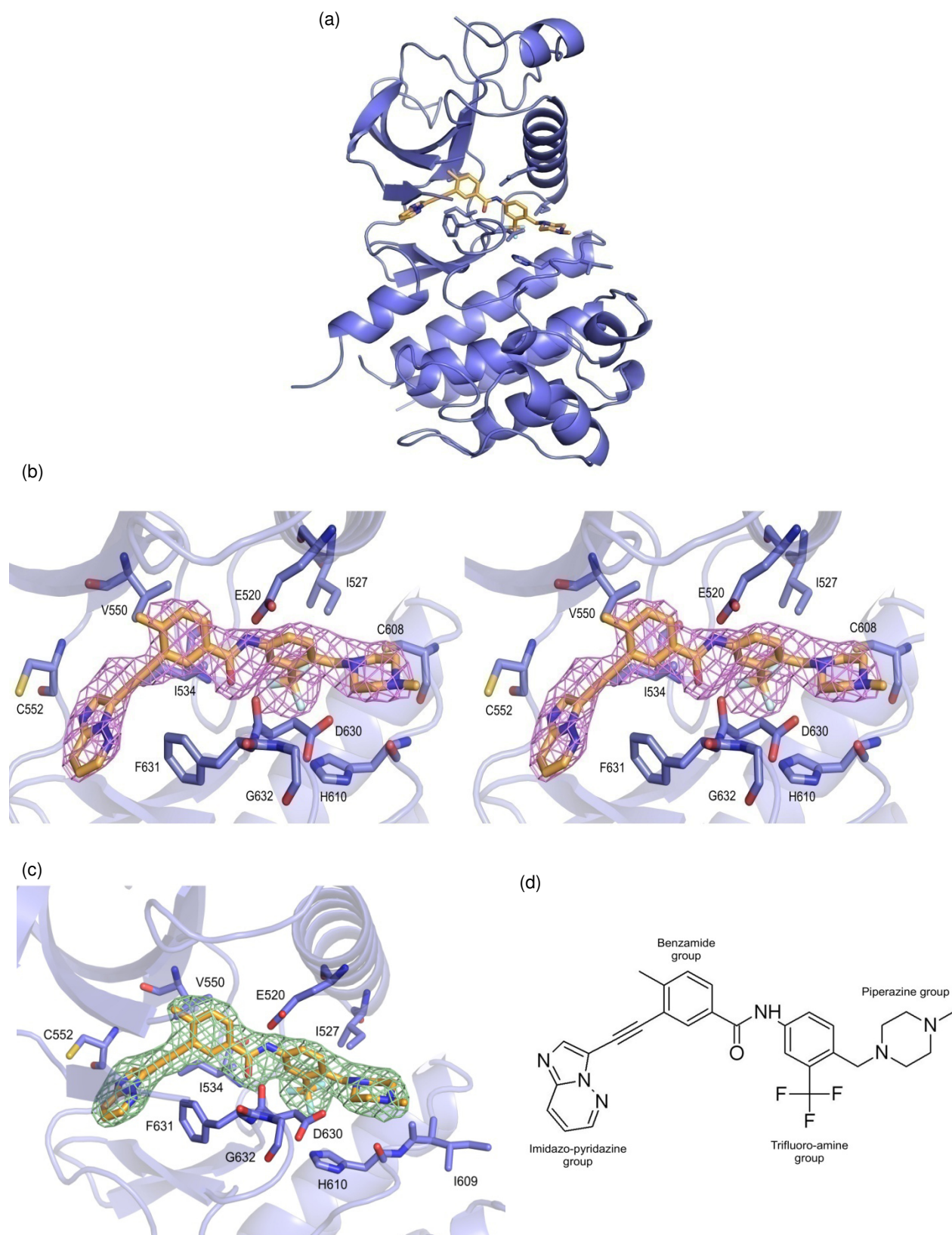


Fig. 22. Ponatinib-FGFR4-KD structure. **(a)** Overall structure of Ponatinib-FGFR4-KD. **(b)** Stereo close-up of FGFR4-KD active-site with Ponatinib bound. Fo-Fc omit map of Ponatinib FGFR4 complex with a contour level of 2.5σ . **(c)** Refined density map of FGFR4-KD in complex with Ponatinib. **(d)** Drawing of Ponatinib inhibitor. Atoms are colored according to atom type: carbon, orange; nitrogen, blue; oxygen, red; fluor, cyan.

4. Discussion

4.1. Autoinhibition in FGFR4-KD

FGFR4 has a main role in skeletal muscles development, with a peculiar autophosphorylation rate and tyrosine kinase activity compared to other FGFRs^{57,22,36}. In addition, it has been suggested that the FGFR4 kinase domain undergoes heterologous phosphorylation by other FGFRs, and its truncated variant appears to have an autoinhibitory role³⁶.

The two FGFR4-KD apo-structures described in this thesis both show an uncommon orientation of the activation segment that is stabilized by interactions with different elements of the N-lobe. In addition, Tyr642 that becomes phosphorylated during FGFR4-KD activation is trapped in a hydrophobic network between β 3-strand and α C-helix, whereas the second tyrosine in the activation segment (Tyr643) is in two different positions in the two apo-structures. The absence of phosphates bound and the orientation of the activation segment could indicate an autoinhibitory conformation of the kinase domain, similar to autoinhibitory strategies already reported for other receptor tyrosine kinases. For example, IRK, TrkA and Ror2 have a tyrosine in the substrate-binding site preventing its phosphorylation^{20,101}. A comparison among the Apo-1-, Apo-2- and Dovitinib-structures and active/phosphorylated FGFR1-2 (Table 3) shows that FGFR4-KD DFG-motif is oriented in an autoinhibited conformation, even though Dovitinib induces a displacement of the activation segment.

In accordance with our structural analysis, we suggest an autoinhibition “dual switch” mechanism for FGFR4-KD, as formulated for c-MET¹⁰². Accordingly, the kinase can reach a less autoinhibited/partially active conformation when both tyrosines of the activation segment are successively phosphorylated and stepwise displaced. The activation segment of FGFR4-KD is similar to the one reported for c-Met and, with Tyr642 buried deeply inside the protein, it appears logical that Tyr643, which is more easily accessible, should be the first residue to be phosphorylated. Upon this initial phosphorylation event, the interaction of the activation segment with the rest of the protein could be weakened to allow for a larger structural reorganization, enabling phosphorylation of Tyr642. Indeed, this has been demonstrated for the activation process of c-MET, by screening with phosphorylation blot analysis of the tyrosines of the

activations segment (Tyr1234 and 1235)¹⁰². Our phosphorylation analysis had indicated that Tyr642 is phosphorylated after Tyr643, like in c-MET, in accordance with the structural data.

This result points out that FGFR4 behaves differently from other FGFRs as had been shown by other experiments^{22,36,57,59}. Specifically, it was reported for FGFR1^{38,39}, that the internal tyrosine 653 is phosphorylated before the external tyrosine 654.

4.2. FGFR4-KD in cancer

Over the last years, FGFR4 has been identified as oncogene and a candidate target for cancer diagnosis and therapy^{103,68,74}. Mutations in the FGFR4 kinase domain have been shown to be associated with lung and rhabdomyosarcoma cancer (RMS)^{68,42}.

Mutations of Arg616 and Glu681 appear to be associated with lung cancer^{104,42}. A mutation of Arg616 to glycine disengages Asp612 of the HRD-motif and thereby reduces the catalytic activity of the kinase. Similar effects are observed when Glu681 is mutated to lysine: Glu681 in FGFR4 forms hydrogen bonds with W655 and A615 which are also involved in Asp612 coordination.

RMS is an aggressive childhood cancer originating from skeletal muscles and is linked with 4 residues: Asn535, Val550, Ala554 and Gly576. FGFR4 N535K/D corresponds to the equivalent mutation N549K in FGFR2, which is responsible for the disengagement of the molecular brake^{49,98}. Ala554 is located in the hinge region at the entrance of the ATP-binding site, and its mutation to valine could result in an interaction with Leu473 in the P-loop. Deduced from our Dovitinib-structure, a valine at position 554 could interact with the hydrophobic part of the benzimidazole group, suggesting that molecules similar to Dovitinib could also inhibit the mutated version of FGFR4-KD. Gly576 is located on the trans-activation loop that is disordered and not visible in our structures. However, residue 575 is an aspartate and the G576D mutation may lead to a distorted conformation of this segment.

Three clinical mutations involve the gatekeeper residue Val550, which has been found mutated to leucine/glutamate in RMS and to methionine in breast cancer^{70,42}. *In vivo* and *in vitro* studies are available only for V550E variant, which is an activating mutant^{68,74}. A mutation at the gatekeeper position in FGFR4-KD could result in a clash with Tyr643 enforcing an open conformation of the activation segment. Our thermofluor analysis prove that FGFR4-KD-V550E is less stable than the wild-

type in absence of ATP-analogue but, stabilized by ATP (Fig. 5), reaching a melting temperature closer to FGFR4-KD-wt.

In vitro studies confirmed that Ponatinib inhibits wild-type and both N535K and V550E mutants, while *in vivo* it has higher activity against the mutant forms⁷⁴. From our Ponatinib-structure it is possible to hypothesize that a mutation to a charged residue strengthens inhibitor binding, as Glu550 could interact with the nitrogen of the benzimide group and induce a small shift of Ponatinib with respect to its position in the wild-type-structure. In contrast to the glutamate-mutation described for FGFR4-KD, mutation of the gatekeeper in FGFR2 to isoleucine abolishes Ponatinib-binding which is attributed to a steric clash of two residues with the inhibitor, the gatekeeper (Val564) and Ile548, which is located on the β -strand next to the gatekeeper⁹⁸.

4.3. FGFR4-KD and drug design

FGFR proteins have been known to be involved in cancer for many years and the synthesis of target-specific inhibitors is a main goal in drug discovery. Our structures of the FGFR4-KD in complex with Dovitinib and Ponatinib reveal amino acids specific for FGFR4-KD which can be exploited within the drug discovery process. Dovitinib is smaller than Ponatinib, and its quinolone group binds closer to the gatekeeper Val550 (Fig. 21). Dovitinib derivatives may be considered as a strategy for FGFR4-KD gatekeeper mutants.

Interestingly, Cys552 is absent in other FGFRs, DDR1, RIPK2 and ABL kinases, and is located at the entrance of the ATP-binding site. In both complex-structures, Cys552 is in direct contact with the inhibitor and offers opportunity for a covalent anchor with Ponatinib derivatives (Fig. 7c). Ponatinib binds FGFR4-KD in a DFG-out conformation, documenting the selectivity of the binding pocket. In contrast with the other FGFRs, FGFR4 has valine 523 instead of a methionine at the equivalent position on the α C-helix, and Val523 contacts the piperazine group of Ponatinib. Comparing the FGFR4-KD Ponatinib-structure with other Ponatinib structures available, we identify some more differences. For instance, the gatekeeper residue is a valine only in the FGFR family, while in ABL, DDR1 and RIPK2 it is a threonine. In addition, the hydrophobic residue in the β -strand under the gatekeeper is different among these kinases and this can modify the binding affinity. Cys608 is in contact with the piperazine

group of Ponatinib and conserved within the FGFR family but substituted in the other kinases. Moreover, the ABL Ponatinib-structure clearly shows an interaction among the P-loop, the activation segment and the inhibitor, while this is not the case in FGFR4-KD.

Taken together, the crystal structures shown in this thesis identify structural elements that can be exploited for the design of selective FGFR-ligands.

Part II

Expression and purification of the human calcium channel ORAI1

Summary

The plasma membrane defines and separates cells and organelles. The membrane is a lipid bilayer in which cholesterol, glycolipids and membrane proteins are located.

Orai1 is a plasma membrane protein with four predicted transmembrane segments and its N- and C-termini are located in the cytosol. Orai channels exhibit distinct inactivation profiles, permeability properties, and 2-aminoethoxydiphenyl borate (2-APB) sensitivity¹⁰⁵. They assemble to a tetrameric channel with a unique structure that is unrelated to any other known ion-channel family. Severe combined immunodeficiency SCID patients are homozygous for a single missense mutation in Orai1 (R91W). Children with Orai1 mutations suffer from ectodermal dysplasia, anhydrosis, congenital non-progressive myopathy, and slight mental retardation, in addition to defective T-cell function.

Orai1 has known to be a tetrameric protein complex with at least two states: open and closed. In this thesis I present some strategies I explored to solve some main issues: reaching a minimum level of protein yield and purity for structural studies, reducing contaminant proteins, getting a homogenous sample.

1. Introduction

1.1. Membrane proteins functions

Cells and organelles are defined and separated by the external space through a lipid bilayer, the membrane. Initially, electron microscopy experiments showed that the membrane consists of a lipid bilayer to which numerous membrane proteins are attached¹⁰⁶. Nowadays, we describe the membrane as a phospholipid bilayer in which proteins, cholesterol and glycolipids are inserted and can move through it (*fluid mosaic model*¹⁰⁷). In the membrane, the hydrophilic heads of phospholipids are exposed to the external sides of the bilayer (extracellular space and cytoplasm), while the hydrophobic tails are protected in the inner core of the bilayer. In the bilayer Lipid molecules continually move in a horizontal direction, but do not move between the inner and outer sides of the membrane except under special conditions (*lipid flip-flop*). Lipid composition differs between the inner and outer sides of the membrane, and in animal cells glycolipids are known to be found more on extracellular side, working as recognition signals.

Membrane functions are mostly carried out by proteins. Membrane proteins are synthesized, folded and assembled by mutual action of the ribosome and the translocon complex on the endoplasmic reticulum (co-translational translocation). The entire process is driven by specific recognition sequences, which are recognized by ribosome, translocon and scaffold proteins.

Membrane proteins anchor extracellular proteins (collagen, fibronectin, laminin, etc.) and help them in holding up the extracellular matrix or lymphocyte adhesion/rolling. Membrane proteins are also involved in the respiratory chains of mitochondria and bacteria, where oxidation of substrates is coupled to the transfer of electric charges across the mitochondrial or bacterial membrane. In photosynthesis, the photosynthetic reaction centers use the energy of light for the primary charge separation and transport of electrons across the photosynthetic membrane. The lipid-hydrophilic nature of the membrane obstacles an easy diffusion of substances and generates a voltage difference on the two sides. Destabilization of the membrane potential is used by neurons to transmit signals, being the fastest way of communication compared to chemical messengers. Passive transport is driven by channels, porins and permeases, while active transport is catalyzed by pumps exploiting energy-rich molecules like ATP molecules. Besides

channels and transporters, the pharmacologically most attractive membrane proteins are the G-protein coupled receptors (GPCRs). Binding of an agonist to the receptor leads to an activation of trimeric G-proteins and the start of a signal transduction cascade.

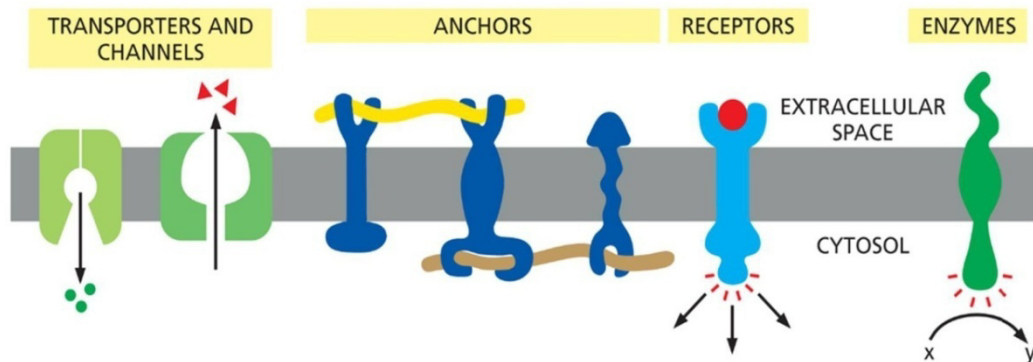


Fig. 23. Schematic representation of integral membrane protein functions (Essential Cell Biology, Alberts¹⁰⁸).

1.2. Calcium signaling

Calcium (Ca^{2+}) signaling is one of the most widespread signaling systems in signal transduction pathways. The concentration of calcium ions in the cytoplasm in most mammalian cells is of the order of 100-200nM, while extracellular calcium ion concentration is maintained in the low millimolar range. Thus, perturbations in this wide gradient can produce rapid increase in intracellular calcium. Calcium-dependent proteins (such as calmodulin) can evoke a myriad of cellular responses, ranging from the extremely rapid responses of neurosecretory cells and skeletal muscle to the long-term responses involving modulation of cell division and differentiation. Calcium ions may come from two sources: through their influx across the plasma membrane or from the endoplasmic reticulum.

Ca^{2+} signaling mechanism can regulate many processes due to its extremely versatility in terms of speed, amplitude and spatio-temporal patterning. This versatility emerges from the use of an extensive molecular repertoire of signaling components, which are assembled in combinations to create signals with widely different spatial and temporal profiles. More variations are achieved through the interactions that Ca^{2+} makes with other signaling pathways (crosstalk)¹⁰⁹. The calcium signaling network can be divided into four functional units:

- Signaling is triggered by a stimulus that generates various Ca^{2+} -mobilizing signals.
- The latter activate the ON mechanisms that move calcium into the cytoplasm.

- Calcium functions as a messenger to stimulate numerous Ca^{2+} -sensitive processes.
- Finally, the OFF mechanisms, composed of pumps and exchangers, remove calcium from the cytoplasm to restore the resting state.

1.3. Calcium Release-Activated Channels (CRAC)

Calcium Release-Activated Channels (CRAC) are specialized plasma membrane calcium ion channels. When calcium ions are depleted from the endoplasmic reticulum of mammalian cells, the CRAC channel is activated to slowly replenish the level of calcium in the endoplasmic reticulum (ER).

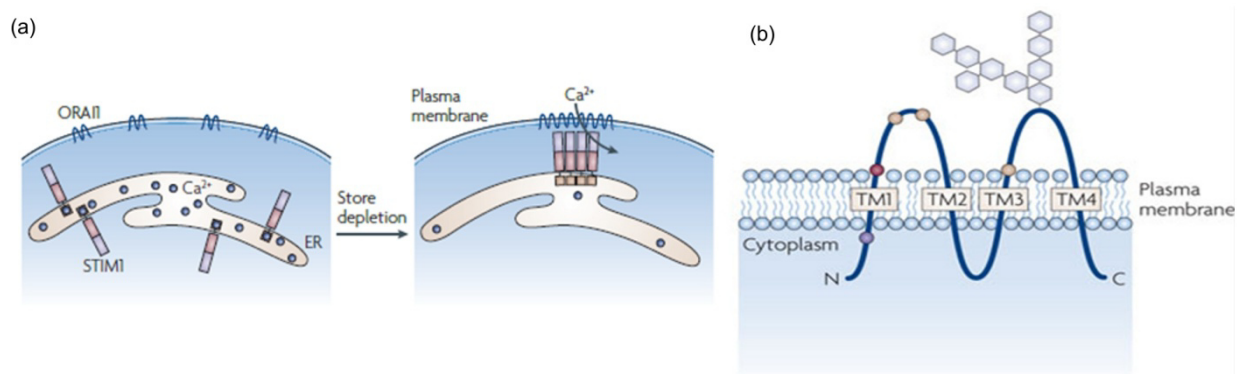


Fig. 24. (a) In cells with full Ca^{2+} stores, stromal interaction molecule 1 (STIM1) is homogeneously distributed in the endoplasmic reticulum (ER) with its EF hand occupied with Ca^{2+} . ORAI1 is dispersed throughout the plasma membrane. Upon stimulation with agonists that increase the levels of cytoplasmic inositol-1,4,5-trisphosphate, the Ca^{2+} content of the store falls. Ca^{2+} dissociates from STIM1 and this results in STIM1 oligomerisation and subsequent migration to ER–plasma membrane junctions. At these sites, STIM1 captures diffusing ORAI1 channels. Interaction between the amino and carboxyl termini of ORAI1 with the CRAC activating domain (CAD) on STIM1 leads to CRAC channel opening. **(b)** Predicted topology of the plasma membrane protein ORAI1. ORAI1 has four transmembrane domains (TM1–TM4), with intracellular amino and carboxyl termini. The purple amino acid represents the single point mutation (R91W) that is seen in CRAC channel-deficient severe combined immunodeficient patients. The red amino acid is glutamate 106. The yellow amino acids represent aspartate 112 and 114 and glutamate 190. Mutation of these latter amino acids affect selectivity, but they might not line the pore. Hexagonal structures represent sugar residues attached to an N-linked glycosylation site N223 (Parekh A.B., 2010¹¹⁰).

Two families of small proteins form the CRAC complex. The first of the two protein families, which functions in sensing calcium ion levels in the ER, comprises the Stromal Interacting Molecules-(STIM)-1 and -2. STIM proteins are single transmembrane proteins that reside in the ER; their N-termini are oriented toward the lumen and contain an EF-hand calcium binding motif¹¹¹. Depletion of calcium from the ER causes calcium ions to dissociate from STIM, resulting in a conformational change that promotes the clustering and migration of STIM molecules¹¹² to closely reach ER–plasma membrane junctions. At

these junctions, STIM oligomers interact with proteins that represent the second of the two families, the ORAI family, consisting of pore-forming ORAI subunits (isoforms 1, 2 and 3- Fig. 23a)^{113–115}.

1.4. Orai1, the pore-forming subunit

Orai1 is a plasma membrane protein with four predicted transmembrane segments and N- and C-termini are located in the cytosol (Fig. 24b). The extracellular loop between the third and fourth transmembrane segments contains an N-glycosylation site (Arg223)¹¹⁵. The Orai family consists further of Orai2 and Orai3 which all can be activated following store depletion via a coupling to STIM1. The overall domain structure of Orai2/3 is similar to that of Orai1, but missing the proline-rich region in the N-terminus. Orai channels exhibit distinct inactivation profiles, permeability properties, and 2-aminoethoxydiphenyl borate (2-APB) sensitivity¹⁰⁵. All three Orai proteins assemble to a tetrameric channel with a unique structure that is unrelated to any other known ion-channel family¹¹⁶. The proline- and arginine-rich N-terminal cytoplasmic sequences have been suggested to play an essential role in Orai1 assembly, as its N-terminus acts in a dominant negative manner on the CRAC complex¹¹⁷. Deletion of the whole N-terminus of Orai1 abolishes CRAC-function, while its partial truncation up to amino acid 73 retains Orai1 channel activity¹¹⁸.

Orai1 exists in two forms in similar quantities: a longer form (Orai1 α) of approximately 33 kDa, and a shorter form (Orai1 β) of approximately 23 kDa. The Orai1 β form derives from alternative translation initiation from a methionine at position 64. In the lost sequence there is a poly-arginine sequence involved in interaction of Orai1 with plasma membrane phosphatidylinositol 4,5-bisphosphate. The loss of this phospholipid binding domain would be expected to influence the mobility of Orai1 protein in the plasma membrane¹¹⁹.

Biochemical studies and electron cryo microscopy reported human Orai1 to be homotetrameric^{116,120} (Fig. 25), however the subunit stoichiometry of Orai1 alone has not been clearly understood, although several reports suggest Orai1 exists as an oligomeric form^{114,121}. Nevertheless, the crystal structure of Orai from *Drosophila melanogaster* has recently been solved and reveals that the calcium channel in insects is composed of a hexameric assembly of Orai subunits arranged around a central ion pore¹²² (Fig. 26).

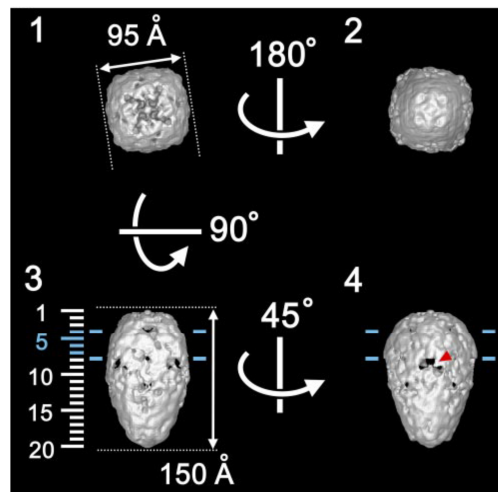


Fig. 25. Three dimensional structure of human Orai1 by electron cryo microscopy. Surface representations of Orai1 viewed from four different Euler angles (α , β , γ): 1 (0, 180, -45), 2 (0, 0, -45), 3 (0, 90, -45), and 4 (0, 90, 0). The molecular mass enclosed by the isosurface is 210 kDa, corresponding to 149% of the tetrameric Orai1 protein. Protein is displayed in bright shades. Two blue lines, ~ 30 Å apart in panels 3 and 4, indicate the putative position of the lipid bilayer. A red arrowhead in panel 4 indicates one of inverted-V-shaped orifices in the cytoplasmic domain (Maruyama Y., 2010¹¹⁶)

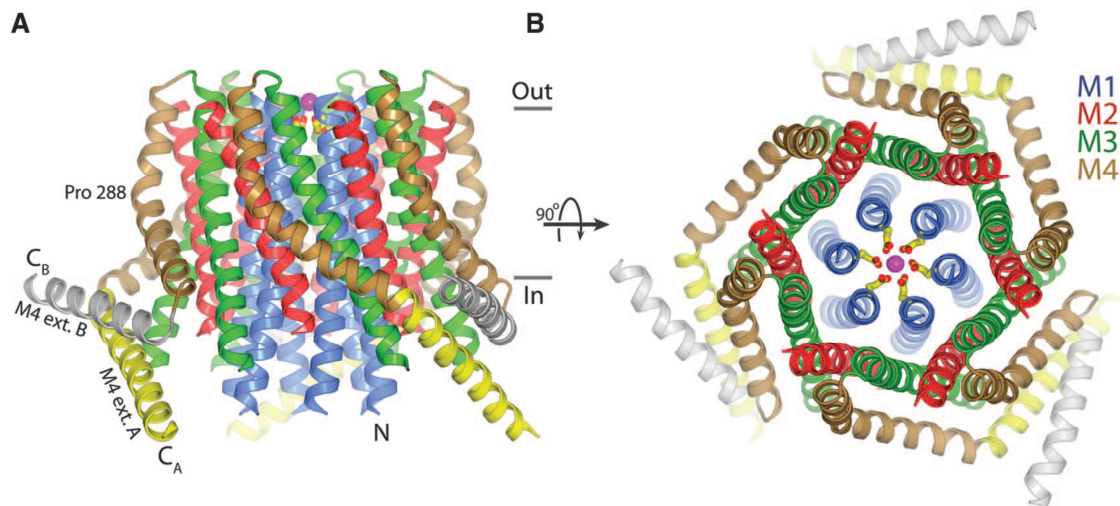


Fig. 26. Architecture of Orai1 from *Drosophila melanogaster*. **(A)** Ribbon representation showing the tertiary structure of the channel from the side. The helices are colored: M1 (blue), M2 (red), M3 (green), M4 (brown), M4 extension (yellow in subunit A and gray in subunit B). Also shown are a Ca^{2+} ion (magenta sphere) and the nearby Glu178 residues (yellow sticks). Based on the hydrophobic region of the channel's surface, horizontal lines (~ 30 Å apart) suggest approximate boundaries of the inner (In) and outer (Out) leaflets of the membrane. **(B)** Orthogonal view of the channel from the extracellular side (Hou X., 2012¹²²).

Severe combined immunodeficiency SCID patients are homozygous for a single missense mutation in Orai1 (R91W), leading to the loss of CRAC-function¹²³. SCID represents a subgroup of primary immune deficiencies causing serious infections, such as pneumonia, meningitis or bloodstream infections¹²⁴. It is currently known that defective T-cell signaling in SCID patients can arise from mutations in different genes including the point mutation R91W in Orai1. Children with Orai1 mutations suffer from ectodermal dysplasia, anhydrosis, congenital non-progressive myopathy, and slight mental retardation, in addition to defective T-cell function¹²⁵. The arginine at position 91 in wild-type Orai1 is positively charged and highly hydrophilic, while the tryptophan in the SCID mutant is neutral and hydrophobic. Electrophysiological examination of various point mutants has revealed that negatively charged or polar residues largely maintain functional Orai1 channels with unaltered calcium ions selectivity and similar current densities. However, hydrophobic substitutions result in the loss of Orai1 channel function with the coupling to STIM1 retained. This increase in hydrophobicity of the amino acid at position 91 may change Orai1 channel conformation by altering the orientation of the first transmembrane helix in the plasma membrane (Fig. 27), thereby leading to the loss of channel function¹²⁶.

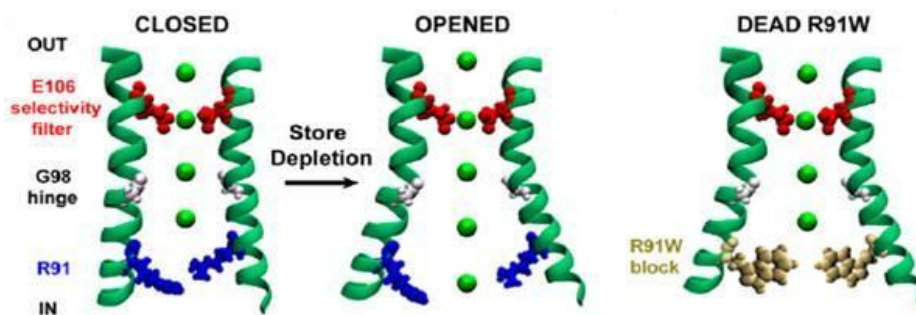


Fig. 27. Schematic models of Orai1 channel gating. Two of the pore-lining Orai1 TM1 segments are shown, with E106, G98, and R91 highlighted from outside to inside. E106 sites near the outside control Ca^{2+} selectivity. Near the inside, the R91 sites prevent ion flow in the closed state of the channel. Located in the middle of TM1, the G98 sites permit a conformational change that opens the channel upon store depletion followed by STIM1–Orai1 interaction. R91W Orai1, irreversibly closed by a greasy tryptophan plug and hence a “dead” channel (Zhang S.L. et al., 2011¹²⁶).

1.5. STIMs: stromal interaction molecules

STIM proteins are dynamic sensors of Ca^{2+} stored within the endoplasmic reticulum and are triggered by Ca^{2+} store depletion to self-activate, aggregate and translocate to endoplasmic reticulum. There STIM proteins attach to the plasma membrane and activate the highly Ca^{2+} selective Orai family of plasma membrane Ca^{2+} entry channels.

The amino-terminal ER-luminal segments of STIM proteins are the Ca^{2+} sensing domains by specific EF-hand domain Ca^{2+} binding sites, which are highly tuned to detect small changes in luminal calcium levels. The detection of decreased luminal Ca^{2+} concentrations triggers a conformational change that leads to aggregation and translocation of STIM proteins.

Cytoplasmic carboxy-terminal STIM protein domains contain a uniquely configured group of α -helical sequences that form an activating domain able to bind to and gate the Orai channel in the plasma membrane. The helical domains can both mediate and regulate the Orai channel-coupling process.

The STIM-induced ER-plasma membrane junctional domains contain a number of important regulatory protein sequences for scaffold proteins, such as: CRAC regulatory protein 2A (CRACR2A), junctate, store-operated Ca^{2+} entry (SOCE)-associated regulatory factor (SARAF), golli and partner of STIM1 (POST). These proteins turn on and off the coupling between STIM proteins and the Ca^{2+} signaling target proteins that function within junctions.

New information reveals STIM proteins as sensors not only of decreased ER Ca^{2+} but also of oxidative stress, temperature increases, hypoxic stress and acidosis. STIM proteins also target proteins other than Orai channels, including voltage-operated $\text{CaV}1.2$ channels, transient receptor potential channels and plasma membrane Ca^{2+} ATPase and SERCA Ca^{2+} pumps.

The function of STIM proteins is to generate long-term spatially discrete Ca^{2+} entry signals that specifically turn on transcriptional events. The triggered Ca^{2+} entry also replenishes stores, and this is required to maintain Ca^{2+} oscillations and to protect the integrity of the ER from potentially damaging decreases in endoplasmic reticulum Ca^{2+} levels that could compromise the protein trafficking and assembly functions of endoplasmic reticulum¹²⁷.

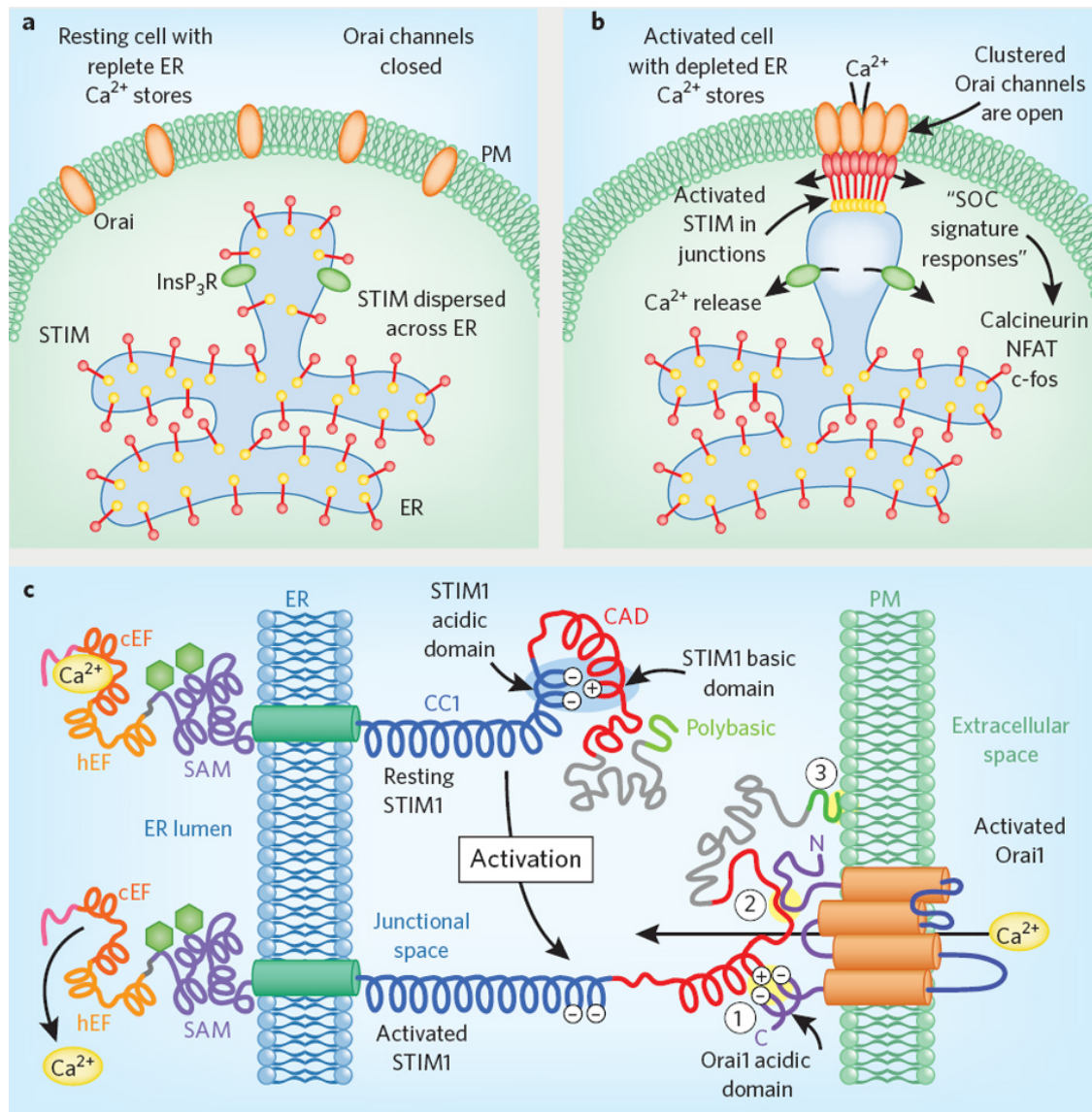


Fig. 28. (a) STIM proteins in the resting state. Under normal conditions, InsP₃R Ca²⁺ release channels are closed, ER Ca²⁺ stores are full and STIM proteins with Ca²⁺ bound to their EF-hands are distributed throughout the ER. There is no coupling with Orai channels in ER-PM junctions, and inactive Orai channels are distributed evenly within the PM. (b) STIM proteins in activated cells. Receptor-induced InsP₃ production activates Ca²⁺ release channels (InsP₃Rs). In areas of Ca²⁺ depletion, Ca²⁺ dissociation from STIM proteins causes their activation, leading to aggregation and translocation into ER-PM junctions, where they become trapped through interactions with PM lipids and Orai proteins. (c) STIM activation and coupling to Orai channels. STIM proteins sense small changes in luminal Ca²⁺ and function as molecular switches to couple with and activate Orai channels (Soboloff J., 2011¹²⁸).

2. Materials and Methods

2.1. Molecular cloning

Standard polymerase chain reaction was performed in 50µl with 1x Phusion HF buffer, 200µM dNTPs, 0.5µM of forward and reverse primers, 3% DMSO, 50-100ng of DNA template and 1U Phusion DNA polymerase (New England BioLabs). The following program was used for amplification in a Eppendorf MasterCycler:

Initial denaturation: 98°C x 30sec

30 cycles: denaturation 98°C x 10sec – annealing 55°C x 1min – elongation 72°C x 30sec

Final elongation: 72°C x 10min

The PCR product was then purified using the QIAquick extraction kit (Qiagen). After restriction enzyme digestion using New England Biolabs enzymes and protocols, the fragments were separated and purified from agarose gels with the QIAquick extraction kit (Qiagen). The target vector was digested in the same way and also purified via gel-extraction. Ligation of the PCR product into the vector was carried out using T4 DNA Ligase (New England BioLabs) according to the manufacturer protocol. After transformation into *Escherichia coli* (*E. coli*) TOP10F' and growth at 37°C over night, plasmid DNA was prepared using the QIAcube (Qiagen) with the corresponding QIAprep Spin Miniprep Kit (Qiagen). The quality of the final plasmid vector containing the target insert was checked by restriction enzyme digestion and DNA sequencing. For agarose electrophoresis and visualization of DNA typically, 1% (w/v) agarose was dissolved by heating in 50x TAE buffer (Applichem) and gels were prepared using a Sub-Cell GT electrophoresis system (Bio-Rad). After mixing with 6x DNA loading dye (Fermentas), the samples were loaded and electrophoresis was carried out at a constant voltage of 110V in 50x TAE buffer (Applichem). Either a 1kb or 100bp DNA ladder (peqlab) was used as a size reference. DNA was visualized with SYBRR Safe DNA gel stain (Life Technologies), followed by imaging on a gel-doc system (Bio-Rad).

Human ORAI1 was obtained by GENEART, corresponding to human ORAI1 isoform alpha in UniProtKB Q96D31. The used vectors are pFastBachHTA (Life Technologies). Diverse ORAI1 length

variants were cloned in vectors containing two main protein-tags (His- and Flag- tags) at N- and C-termini. A list is reported and explained in “Results” at page 65.

The successful clone coded the full length amino acids sequence 1-301 of human ORAI1 cloned into pFastBacHTA vector, containing 10xHis-tag and mutated at the potential glycosylation site N223D. This and other ORAI1 mutants were generated with QuikChange II Site-Directed Mutagenesis Kit (Agilent).

2.2. Insect cell culture and heterologous protein expression

For the expression of target protein in insect cells, viral DNA was generated using the Baculovirus Expression Vector System (Life Technologies) following the instructions given by the supplier (Fig 9). For recombinant protein expression, High Five Cell Line derived from *Trichoplusia ni* were grown in Grace's Insect Media (Gibco-Life Technologies) and infected at a cell density of $1.5-2 \times 10^6$ cells/ml (vitality 95-98%) with an MOI of 2 (Virus Counter InDevR) 64h after infection, cells were harvested by centrifugation at 900g for 10min and shock-frozen at -80°C .

Initially, test expressions were performed in 100ml cultures which were analyzed using a NiNTA spin column kit (Qiagen) for HIS-tagged target proteins and quantified with SDS-page and western blot analysis. Test expressions were repeated until a good sub-clone was identified (in terms of protein yield) and different parameters were tested to improve expression yields (MOI, post-infection time, cell density).

2.3. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis analysis (SDS-PAGE)

To denature protein samples, they were boiled in 1% sodium dodecyl sulfate (SDS)-loading buffer for 10min at 65°C . Subsequently samples were loaded on gels purchased from Life Technologies (NuPAGER NovexR 10% Bis-Tris Midi Gel). Electrophoresis was performed at 190V until the running front reached the boarder of the gel. Prestained Protein Marker VI (AppliChem) was used as a size reference covering 10-245 KDa. The gel running buffer was either 1x NuPAGER MES SDS running buffer (Life Technologies). To visualize proteins, the SDS-PAGE gels were soaked in the Coomassie based staining solution InstantBlue™ (Biozol) at room temperature (RT) for 30min. InstantBlue™

protein-staining solution then was removed and substituted with water. Imaging was performed on a gel-doc system (Bio-Rad).

2.4. Western Blot Analysis

After SDS-PAGE analysis, Western Blotting was performed using the iBlot™ Dry Blotting System (Life Technologies) with the proper iBlot Transfer Stack, nitrocellulose (Life Technologies) basically according to the manual provided by the distributor. The anti-HIS₆-Peroxidase and anti-FLAG-Peroxidase (Roche) were used as anti-body and visualized using BM Blue POD substrate, precipitating (Roche).

2.5. Determination of DNA and protein concentration

Protein and DNA-concentrations were determined using a NanoDrop Spectrophotometer (ND-1000 spectrophotometer, peqlab). Protein concentration was also estimated with the bicinchoninic acid (BCA) assay through the BCA Protein Assay Kit (Thermo Scientific Pierce), using albumin to generate the reference curve.

2.6. Membrane protein isolation and preparation

Frozen pellet was thawed and resuspended in Hypotonic Lysis buffer (10mM HEPES pH=7.5, 20 CaCl₂, 10mM MgCl₂, 10% glycerol, 1mM TCEP) supplemented with complete-EDTA free proteases inhibitors and DNase I (Roche). Resuspended pellet was centrifuged at 50000g for 30min at 10°C and then resuspended again in the Hypotonic Lysis buffer and treated with a homogenizer stirrer for 6min at medium speed. Centrifugation and homogenization steps were repeated three times. The third time the pellet was resuspended in the Solubilization buffer (20mM Tris/HCl pH=8.5, 100mM NaCl, 20mM arginine, 20mM glutamate, 10% glycerol, 0,3% DDM – 0,09% CHS, 1mM TCEP) and incubated on a magnetic stirrer at the slowest speed for 3h at 4°C.

2.7. Purification protocol

In general, all the purification steps were performed at 4°C, all chromatography buffers were prepared and the pH adjusted at room temperature. Subsequently they were filtered (0.22 µm) and degassed. The purification steps were performed on chromatography systems and columns obtained from GE

Healthcare (ÄKTA system) using low flow speed to prevent protein releasing after binding to the resins. Before the protein was pooled, samples were analyzed on pre-cast SDS-gels (10%) obtained from Life Technologies. Many buffers systems and columns were tested, the protocol here reported refers to the one which gives the best protein purity and yield.

After 3 hours, membrane proteins have been solubilized and the resuspended pellet was clarified with centrifugation at 75000g per for 30min at 10°C. The harvested supernatant-lysate was supplemented with 100mM NiSO₄ and incubated 5min at °C on a magnet stirrer. Then Ni Sepharose High Performance (GE Healthcare) resuspended in buffer A (20mM Tris/HCl pH=8.5, 100mM NaCl, 20mM arginine, 20mM glutamate, 40mM imidazole, 5% glycerol, 0,1% DDM – 0,03% CHS, 1mM TCEP) were added and the mixture was incubated over night at 4°C on the magnetic stirrer. The next day, the XK16/20 column is packed. The column was washed with buffer A and bound protein was eluted with buffer B (20mM Tris/HCl pH=8.5, 100mM NaCl, 20mM arginine, 20mM glutamate, 300mM imidazole, 5% glycerol, 0,1% DDM – 0,03% CHS, 1mM TCEP). ORAI1 protein was further purified on a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare) in buffer C (10mM Tris/HCl pH=8.5, 8mM Nonylmaltoside, 8mM Nonylglucopyranoside), during which the detergent was substituted. Then, the protein was concentrated for crystallization. To estimate the MW of proteins and complexes, the columns were calibrated with the Low and the High Molecular Weight Gel Filtration Calibration Kit (Amersham Pharmacia Biotech), which uses proteins in the range between 13.7 kDa and 669 kDa.

2.8. Proteins concentration steps

In order to concentrate protein samples after intermediate and final purification steps centrifugal filter devices (Amiconr Ultra and Sartorius Vivaspin 500) with 30KDa nominal molecular weight limit were used as described in the provided protocol. Concentration was conducted up to the desired volume for intermediate purification and up to the desired protein concentration in final concentration steps.

2.9. Detergents screenings

The screening was performed with NiNTa spin column kit (Qiagen) for HIS-tagged target proteins and quantified with SDS-page and western blot analysis. Small aliquots of resuspended pellet were

incubated with solubilization buffer containing 2 times the estimated CMC (Critical Micelle Concentration) value of each detergent. The detergents tested are reported in “Results”, page 63.

2.10. Protein crystallization

Crystallization screenings were set up with Phoenix robot (Art Robbins Instruments) with vapor diffusion – sitting drop method. Thousand of conditions were screened belonging to commercial screenings (Hampton Research, Qiagen). Different protein concentrations (5-8mg/ml) and rates protein/reservoir (1/1 or 1/2) were tested. Plates were incubated both at 4° and 20°C.

Lipid Cupid Phase trials were carried out using NeXtal CubicPhase kit by Qiagen, composed by crystallization plates pre-covered with monoolein lipid and crystallization solutions Suite I and II.

Crystals were analyzed with the SilverQuest Silver Staining Kit (Life Technologies).

3. Results and Discussion

In the second part of my thesis I present some strategies I explored to reach a basal level of protein yield and purity for structural studies, reduce contaminant proteins and get a homogenous sample.

3.1. Cloning and expression of Orai1 variants

Different variants of the human Orai1 channel were cloned, amplified and expressed as described in “Methods”. Through mutagenesis, the predicted glycosylation site at asparagine 223 was removed in some variants to simplify the recombinant synthesis in insect cells and substituted with aspartatic acid. The severe combined immunodeficiency (SCID) mutated variant R91W was also generated. Mutant R91W leads to the abrogation of CRAC channel function^{123,129} and, even though it has the same expression level and STIM1 binding capacity as the wild-type channel, it is blocked in an open conformation due to the apolar hydrophobic interaction among tryptophan residues¹²⁶. My idea was to use R91W mutant to get a homogenous sample with channel molecules blocked in an open state. In later experiments, a variant with a deletion of 30 residues in the long cytoplasmatic loop (Δ 204-234) was synthesized to facilitate protein crystallization (Fig.29 – Table 4).

MHPEPAPPPS	RSSPELPPSG	GSTTSGSRRS	RRRSGDGEPP	GAPPPPPSAV	50
TYPDWIGQSY	SEVMSLNEHS	MQALSWRKLY	LSRAKLGASS	R TSALLSGFA	100
MVAMVEVQLD	ADHDYPPGLL	IAFSACTTVL	VAVHLFALMI	STCILPNIEA	150
VSNVHNLNSV	KESPHERMHR	HIELAWAFST	VIGTLLFLAE	VVLLCWVKFL	200
PLK K QPGQPR	PTSKPPASGA	AA N VSTSGIT	PGQ A AAIAST	TIMVPPGLIF	250
IVFAVHFYRS	LVSHKTRDQF	QELNELAEFA	RLQDQLDHRG	DHPLTPGSHY	300
A					301

Fig. 29. ORAI1 amino acid sequence (Q96D31). Red letters mark residues 91 and 223, while green letters mark the deletion 204-234.

Below there are listed the Orai1 variants cloned and prepared. Most of them had a very low level of protein expression and were dropped out after test expression (Table 4). At the beginning, I worked on the full length variant A (corresponding to human isoform Orai1- α), but the long N-terminus together with the long cytoplasmatic loop strongly prevent channel crystallization. The short variant D (corresponding to the truncated human isoform Orai1- β) was left when I realized that variant E carrying the tryptophan 91 mutation was more expressed and stable, despite the same sequence length. Finally,

variants H and I were cloned following the publication on Dm-Orai1¹²² however, despite their good expression level in the test expression, their scale-up never worked properly, resulting in a low yield with many contaminant proteins.

Table 4. Human ORAI1 variants tested.

<i>Clone</i>	<i>ORAI1 residues*</i>	<i>Baculovirus Vector</i>	<i>Result</i>
Clone A	Hs. 1-301/N223D	pFastBacHTA-FLAG-hORAI1-10HIS	Crystallization never worked
Clone B	Hs. 1-301/N223D	pFastBacHTA-FLAG-hORAI1-FLAG	Poor expression
Clone C	Hs. 1-301	pFastBacHTA-FLAG-hORAI1-10HIS	Poor expression
Clone D	Hs. 64-301/N223D	pFastBacHTA-hORAI1-10HIS	Few protein yield
Clone E	Hs. 64-301/N223D – R91W	pFastBacHTA-hORAI1-10HIS	Crystallization never worked
Clone F	Hs. 64-301	pFastBacHTA-hORAI1-10HIS	Poor expression
Clone G	Hs. 35-301	pFastBacHTA-hORAI1-10HIS	Poor expression
Clone H	Hs. 49-301/N223D – Δ 204-234	pFastBacHTA-10HIS-hORAI1	Low protein yield
Clone I	Hs. 49-301/N223D – R91W – Δ204-234	pFastBacHTA-10HIS-hORAI1	Low protein yield

*refers to *Homo sapiens*. The channel sequence is reported in Fig. 25.

In green are highlighted the main variants I worked with. 10HIS = 10-histidine; FLAG = Flag-tag, sequence DYKDDDDK.

Orai1 was initially expressed both in Sf9 (*Spodoptera frugiperda*) and Hi5 (*Trichoplusia ni*) insect cell lines. Test expressions and later purification identified Hi5 as the best choice (Fig. 30), because of the higher protein yield and better purification compared to Sf9. The final protocols were achieved with variant A and then applied to variants E, H and I, for that I refer to variant A when not specified.

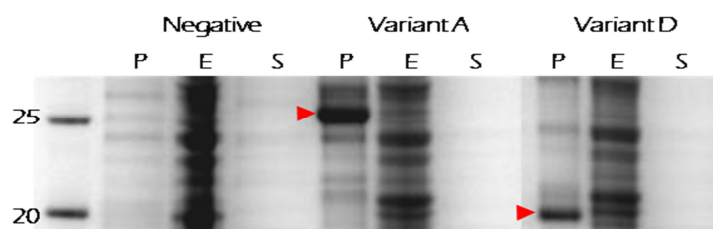


Fig. 30. Test expression analysis in Hi5 cells of ORAI1 channel variants. Variant A has a molecular weight of approximately 36kDa and variant D of approximately 26kDa.

3.2. Detergents screening

N-Dodecil- β -D-maltoside (DDM) was used as initial detergents, but further purifications required optimisation and different detergents were tested:

A = n-Octyl- β -D-glucopyranoside (OG)

B = n-Nonyl- β -D-glucopyranoside (NG)

C = n-Nonyl- β -D-maltoside (NM)

D = n-Decyl- β -D-maltoside (DM)

E = n-Undecyl- β -D-maltoside (UM)

F = n-Dodecil- β -D-maltoside (DDM)

G = n-Tridecyl- β -maltoside (TDM)

H = n-Octyl-1-thio- β -D-glucopyranoside (OTG)

I = n-Nonanoyl-N-methylglucamin (MEGA-9)

L = n-Decanoyl-n-methylglucmin (MEGA-10)

M = 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS)

N = Decyl β -D-maltopyranoside

O = n,n'-bis-(3-D-Gluconamidopropyl)Deoxycholamide (Deoxy Big CHAP)

P = n-Tetradecyl- β -D-maltoside (TTM)

Q = Dodecyl Octaethylene Glycol Ether (C12E8)



Fig. 31. Western blot of the elution fractions after detergent screening.

DDM is commonly used as a starter detergent and it appears to be not the best choice. NG and NM seem to solubilize more protein molecules than DDM and were used to purify the Dm-Orai1 channel variant¹²². For that the size-exclusion buffer C has been changed to a buffer containing a mixture of NG/NM detergents.

3.3. Oligomeric states of Orai1

After size-exclusion chromatography Orai1 elutes partially as tetramer and partially as an apparently higher molecular weight protein-detergent complex (PDC), where it is unclear whether this is composed of aggregated protein or oligomeric states of the channel mixed with detergent. Moreover, Orai1 appears on the SDS-page gel mostly with a double band of the monomeric form (~36kDa variant A), perhaps depending of the lipids surrounding the monomer. To confirm that both those bands correspond to Orai1 protein and not to a contaminant, a double western blot was performed using both anti-His and anti-

FLAG antibodies. In figure 32 the monomer is split in two bands and other oligomeric states are visible, despite the use of denaturing agents. Monomer, dimer, trimer and tetramer were analyzed and confirmed by mass spectrometry.

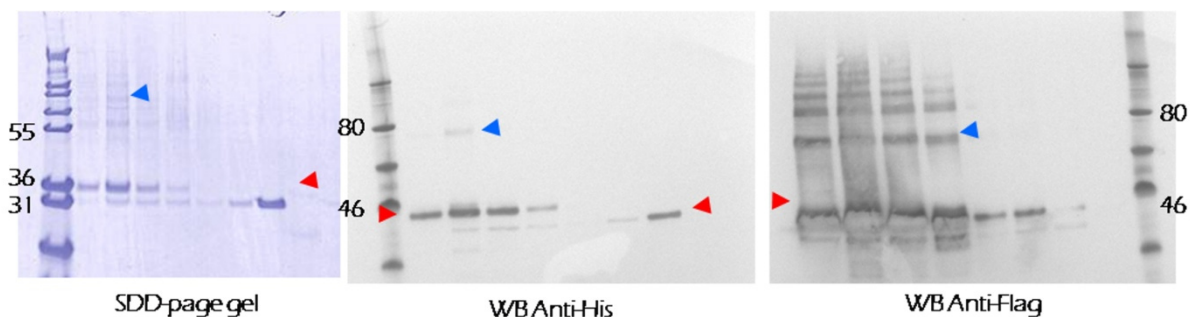


Fig. 32. Elution fractions of Orai1 after size-exclusion chromatography (SEC –left). Western blots with anti-His and anti-FLAG antibodies. Red arrows indicate the monomer and blue arrows the weak signal of the dimeric form.

3.4. Optimization of Orai1 purification

The initial Orai1 purification protocol was composed by an affinity chromatography exploiting the His-tag and a size-exclusion chromatography. The initial buffers were quite simple similar: buffer A (20mM Tris/HCl pH=8.5, 100mM NaCl, 40mM imidazole, 5% glycerol, 0,2% DDM, 1mM TCEP), buffer B (20mM Tris/HCl pH=8.5, 100mM NaCl, 300mM imidazole, 5% glycerol, 0,2% DDM, 1mM TCEP) and buffer C: 20mM Tris/HCl pH=8.5, 500mM NaCl, 5% glycerol, 0,2% DDM, 1mM TCEP. Due to the low protein yield, the size-exclusion chromatography was substituted with a reverse anionic exchange in which the buffer system was Bis-Tris at pH = 6. The anionic exchange separated Orai1 from most of the contaminants, eluting Orai1 in the flow through (Fig. 33) with a higher yield (8mg/ml in 100 μ l).

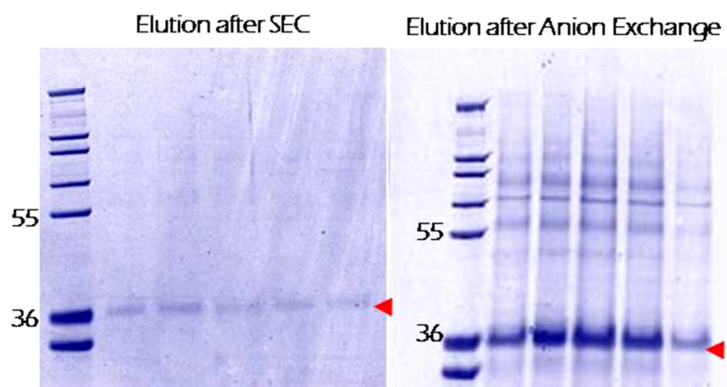


Fig. 33. Elution fractions of Orai1 after size-exclusion chromatography (SEC – left) and after Anion Exchange (right).

Exploiting the Flag-tag, another kind of Orai1 purification was performed in batch with dedicated beads (Anti-Flag M2 Affinity Gel Sigma-Aldrich), using the Flag peptide for the competitive elution. This purification never resulted in highly pure protein, even in combination with a nickel affinity chromatography (Fig. 34).

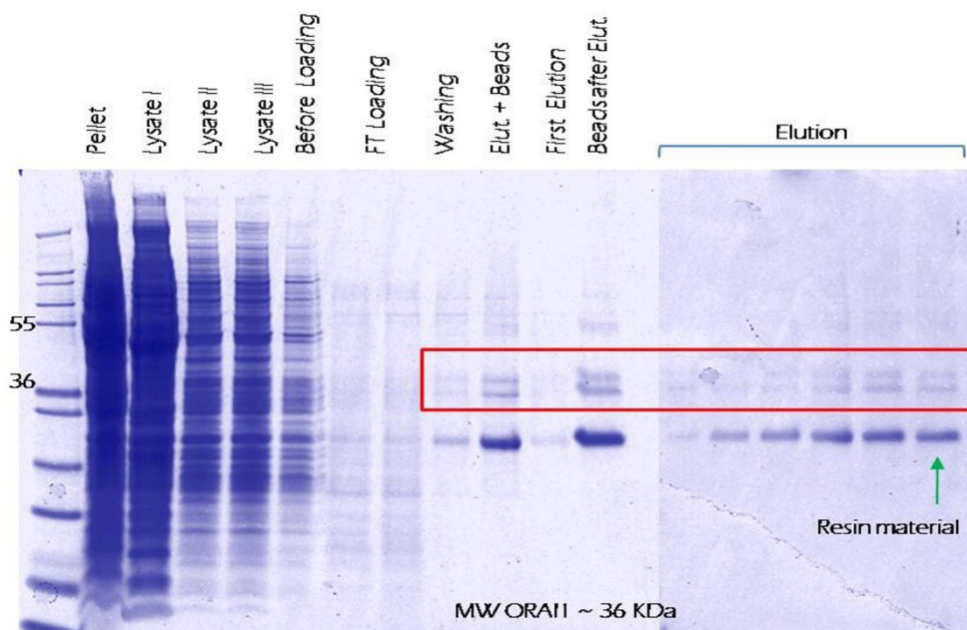


Fig. 34. Elution fractions of Orai1 after size-exclusion chromatography (SEC – left) and after Anion Exchange (right).

Literature about GPCR proteins often recommends supplementing the detergent with cholesteryl hemisuccinate (CHS) in order to better resemble the plasma membrane environment. Moreover, the employment of hypotonic buffers in the membrane preparation is becoming a common method, usually containing salts like $MgCl_2$ or $CaCl_2$. I combined these suggestions with different pH ranges from 7 to 9 and tested n-Undecyl- β -D-maltoside instead of DDM detergent, because of its promising behavior in the detergent screening (Fig. 32). Finally, I restored the size-exclusion chromatography instead of the anionic exchange, with buffer C containing UM detergent (20mM Tris/HCl pH=8.5, 500mM NaCl, 5% glycerol, 0,1% UM, 0,5mM TCEP). Combining these ideas, the best result achieved is reported in figure 30. Here, Orai1 appears to be split in the tetrameric and protein-detergent complex forms, which look identical on the SDS-page gel.

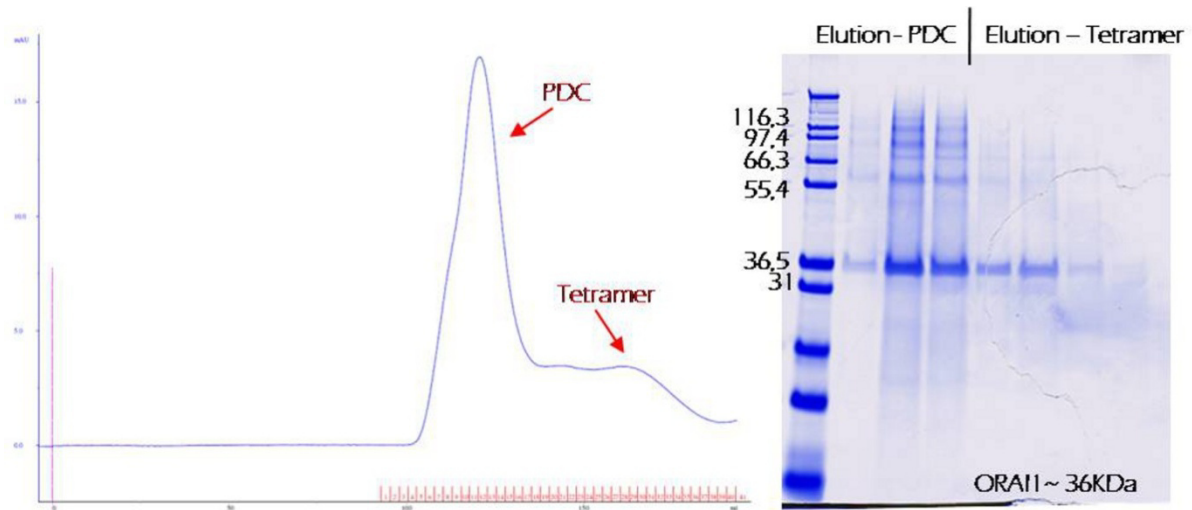


Fig 35. Size-exclusion chromatography of Orai1. In the SDS-page gel, the monomer is the main form, but also other oligomeric states are visible.

After the publication of the homologous Orai1 from *Drosophila melanogaster*¹²², variant H and I were synthesized. They carry two deletions, at N-terminus and in a cytoplasmatic loop. Although the expression of Dm-Orai1 is described in yeast (*Pichia pastoris*)¹²², the solubilization and purification steps are similar to the ones developed here for human Orai1. The main difference is the employment of a lipid mixture together with the DDM detergent, while I was using cholesteryl hemisuccinate, and of n-Nonyl- β -D-maltopyranoside/n-Nonyl- β -Dglucopyranoside mixture in the gel filtration buffer. For that, I modified buffer C to 10mM Tris/HCl pH=8.5, 8mM Nonylmaltoside, 8mM Nonylglucopyranoside. With this size-exclusion chromatography it was possible to obtain a tetrameric form of Orai1 (Fig. 36). However the protein yields of variant H and I were always too low to be used in crystallization screenings.

As mentioned in the “Introduction”, my data agree with other publications^{114,116} demonstrating that human Orai1 is likely tetrameric, despite the *Drosophila melanogaster* isoform is hexameric.

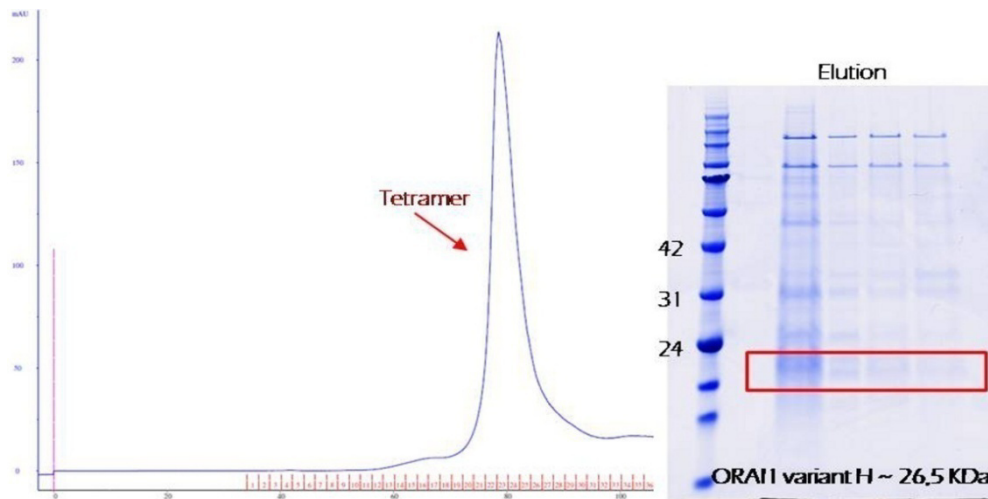


Fig 36. Size-exclusion chromatography of Orai1. In the SDS-page gel, the monomer is the main form, but also other oligomeric states are visible.

3.5. Purification of the CRAC complex: Orai1 with STIM1-CAD domain

Another strategy to purify Orai1 channel is to exploit its binding partner STIM1. For that, a human STIM1-CAD domain (Fig. 28) construct was synthesized (residues 344-485) with standard cloning method. The construct was supplemented with GST-tag (glutathione S-transferase protein) and precession cleavage site. STIM1-CAD was co-expressed with Orai1 variant A. Since it is known that tetrameric STIM1 binds tetrameric Orai1¹³⁰, during expression and membrane preparation STIM1-CAD should interact with Orai1 and form the CRAC complex.

The first tested strategy was to bind STIM1-CAD to the GST-resin together with Orai1. However, this interaction was never clearly visible on SDS-page gel and western blot.

Some more trials were performed exploiting the Flag-tag of Orai1, but again the purification was unsatisfying. The main problem was the low expression of STIM1-CAD and its autoprotolysis.

In order to form the CRAC complex and stabilize Orai1, a separate expression of STIM1-CAD and Orai1 was also tested. As describe above, Orai1 had already a well established expression and purification protocol, while several ways were explored to express and purify STIM1-CAD, such as: *E. coli* and insect cells expression, GST- and MBP-tag (maltose binding protein) combined with mutations to increase protein solubility. None of them resulted in protein enough stable to proceed in further experiments.

3.6. Crystallization of Orai1

Efforts to crystallize Orai1 channel variants A and E screened many conditions as detailed in “Methods”. Ambiguous crystals-like object were obtained in different conditions (Fig. 37), but none of them ever showed diffraction. After classical crystallization method, such as vapor diffusion, lipid cubic phase experiments were performed with variant A. Most of them resulted in crystals-objects of figure 33a without diffraction. They were dissolved in water and analyzed with silver stain electrophoresis (Fig. 37d). Two main bands appeared on the silver stain gel and which has a similar profile compare to the SDS-page gel obtained after size-exclusion chromatography (monomeric form 36kDa).

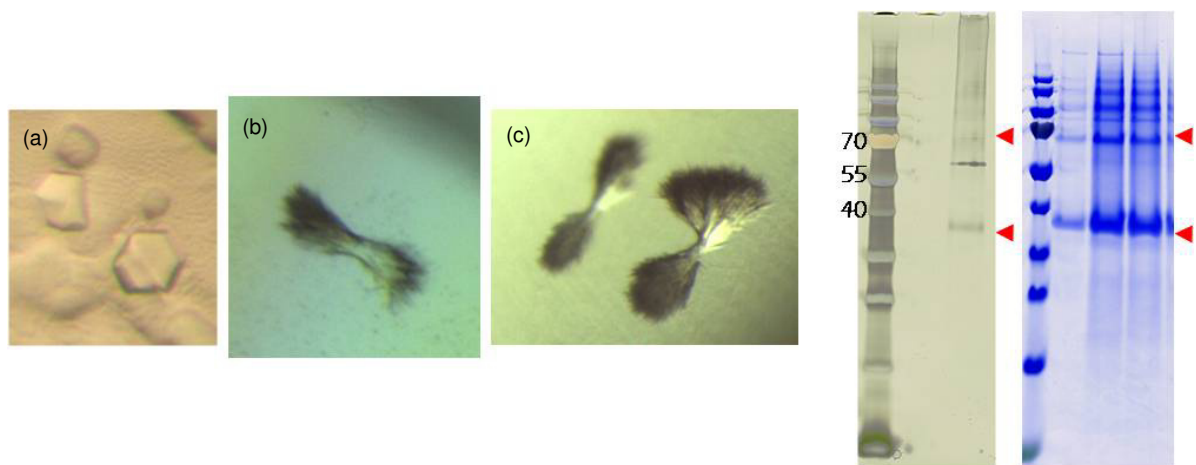


Fig. 37. Orai1 crystals-like objects. **(a-c)** Three different examples of 3D crystal of variant A. **(d)** Silver stain electrophoresis of crystals in photo (a). A typical SDS-page profile of Orai1 is reported on the side.

3.7. Future outlook

Should this project be continued, I would work more on the stability and homogeneity of the Orai1 channel and spend more efforts on variant E and some of its mutants, with and without the R91W mutation and without the N223 glycosylation site. Variant E and its derivatives appear promising for crystallization because flexible loops have been removed. During the project, it became evident that a careful characterization of the recombinant protein is even more important for a membrane protein than for a soluble protein, so another important step is testing different liposomes and finding the best lipid-combination for Orai1. This includes screening different detergents combined with lipids/phospholipids and checks the outcomes with fluorescence-detection size-exclusion chromatography (FSEC) or transmission electron microscopy (TEM). Moreover, a biophysical characterization of the channel

would give an evidence of the physiological profile for a specific variant, and methods such as patch clamp might be exploited.

Concerning crystallization, standard methods might be considered together with lipid cubic phase and mesophase experiments, however I am convinced that a better control of the lipid environment and an high homogeneity of the sample could make a difference. Even if this project did not end up in a crystallographic structure, it has been an intense trial and error with small victories along the way. For that, the expression and purification protocols of Orai1 are used as baseline to build up other membrane protein protocols for structural studies at Proteros Biostructures.

References

1. Hubbard, S.R. & Till, J.H. (2000). Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **69**, 373–398
2. Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225
3. Lemmon, M.A. & Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141**, 1117–1134
4. Maruyama, I. (2014). Mechanisms of Activation of Receptor Tyrosine Kinases: Monomers or Dimers. *Cells* **3**, 304–330
5. Hubbard, S.R., Mohammadi, M. & Schlessinger, J. (1998). Autoregulatory Mechanisms in Protein-tyrosine Kinases. *J. Biol. Chem.* **273**, 11987–11990
6. Ullrich, A. & Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203–212
7. Chen, H. et al. (2008). A crystallographic snapshot of tyrosine trans-phosphorylation in action. *Proc. Natl. Acad. Sci.* **105**, 19660–19665
8. Reiser, J., Sever, S. & Faul, C. (2014). Signal transduction in podocytes[mdash]spotlight on receptor tyrosine kinases. *Nat Rev Nephrol* **10**, 104–115
9. Blume-Jensen, P. & Hunter, T. (2001). Oncogenic kinase signalling. *Nature* **411**, 355–365
10. Roskoski, R. (2010). RAF protein-serine/threonine kinases: Structure and regulation. *Biochem. Biophys. Res. Commun.* **399**, 313–317
11. Besant, P.G., Tan, E. & Attwood, P.V. (2003). Mammalian protein histidine kinases. *Int. J. Biochem. Cell Biol.* **35**, 297–309
12. Caunt, C.J. & Keyse, S.M. (2013). Dual-specificity MAP kinase phosphatases (MKPs). *FEBS J.* **280**, 489–504
13. Lowe, E.D. et al. (1997). The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition. *EMBO J.* **16**, 6646–6658
14. Johnson, L.N., Noble, M.E. & Owen, D.J. (1996). Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149–158
15. Gibbs, C. & Zoller, M. (1991). Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. *J. Biol. Chem.* **266**, 8923–8931
16. Bayliss, R., Fry, A., Haq, T. & Yeoh, S. (2012). On the molecular mechanisms of mitotic kinase activation. *Open Biol.* **2**, 120136–120136
17. Kornev, A.P., Haste, N.M., Taylor, S.S. & Ten Eyck, L.F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci.* **103**, 17783–17788
18. Rubin, G. et al. (2000). Comparative Genomics of the Eukaryotes. *Science* **287**, 2204–2215
19. Manning, G., Whyte, D., Martinez, R., Hunter, T. & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* **298**, 1912–1934
20. Hubbard, S.R., Wei, L. & Hendrickson, W.A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754
21. Levinson, N. et al. (2006). A Src-Like Inactive Conformation in the Abl Tyrosine Kinase Domain. *PLoS Biol* **4**, e144
22. Wang, J.-K., Gao, G. & Goldfarb, M. (1994). Fibroblast growth factor receptors have different signaling and mitogenic potentials. *Mol. Cell. Biol.* **14**, 181–188

23. Eswarakumar, V.P., Lax, I. & Schlessinger, J. (2005). Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev.* **16**, 139–149
24. Turner, N. & Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. *Nat. Rev. Cancer* **10**, 116–129
25. Sleeman, M. et al. (2001). Identification of a new fibroblast growth factor receptor, FGFR5. *Gene* **271**, 171–182
26. Steinberg, F. et al. (2010). The FGFR1 receptor is shed from cell membranes, binds fibroblast growth factors (FGFs), and antagonizes FGF signaling in *Xenopus* embryos. *J. Biol. Chem.* **285**, 2193–2202
27. Beenken, A. & Mohammadi, M. (2009). The FGF family: biology, pathophysiology and therapy. *Nat Rev Drug Discov* **8**, 235–253
28. Schlessinger, J. et al. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **6**, 743–750
29. Mohammadi, M., Olsen, S.K. & Ibrahim, O.A. (2005). Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev.* **16**, 107–137
30. 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. *J. Biol. Chem.* **281**, 27178–27189
31. Tomlinson, D.C., L'Hôte, C.G., Kennedy, W., Pitt, E. & Knowles, M.A. (2005). Alternative Splicing of Fibroblast Growth Factor Receptor 3 Produces a Secreted Isoform That Inhibits Fibroblast Growth Factor-Induced Proliferation and Is Repressed in Urothelial Carcinoma Cell Lines. *Cancer Res.* **65**, 10441–10449
32. Holzmann, K. et al. (2012). Alternative Splicing of Fibroblast Growth Factor Receptor IgIII Loops in Cancer. *J. Nucleic Acids* **2012**, 1–12
33. Jang, J. (2002). Identification and characterization of soluble isoform of fibroblast growth factor receptor 3 in human SaOS-2 osteosarcoma cells. *Biochem. Biophys. Res. Commun.* **292**, 378–382
34. Sturla, L., Merrick, A. & Burchill, S. (2003). FGFR3IIS: a novel soluble FGFR3 spliced variant that modulates growth is frequently expressed in tumour cells. *Br J Cancer* **89**, 1276–1284
35. Corso, T.D. et al. (2005). Transfection of tyrosine kinase deleted FGF receptor-1 into rat brain substantia nigra reduces the number of tyrosine hydroxylase expressing neurons and decreases concentration levels of striatal dopamine. *Mol. Brain Res.* **139**, 361–366
36. Kwiatkowski, B.A., Kirillova, I., Richard, R.E., Israeli, D. & Yablonka-Reuveni, Z. (2008). FGFR4 and its novel splice form in myogenic cells: Interplay of glycosylation and tyrosine phosphorylation. *J. Cell. Physiol.* **215**, 803–817
37. Spivak-Kroizman, T. et al. (1994). Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell* **79**, 1015–1024
38. Furdui, C.M., Lew, E.D., Schlessinger, J. & Anderson, K.S. (2006). Autophosphorylation of FGFR1 Kinase Is Mediated by a Sequential and Precisely Ordered Reaction. *Mol. Cell* **21**, 711–717
39. Lew, E.D., Furdui, C.M., Anderson, K.S. & Schlessinger, J. (2009). The Precise Sequence of FGF Receptor Autophosphorylation Is Kinetically Driven and Is Disrupted by Oncogenic Mutations. *Sci. Signal.* **2**, ra6
40. Bae, J.H. et al. (2009). The Selectivity of Receptor Tyrosine Kinase Signaling Is Controlled by a Secondary SH2 Domain Binding Site. *Cell* **138**, 514–524
41. Ong, S. et al. (2000). FRS2 proteins recruit intracellular signaling pathways by binding to diverse targets on fibroblast growth factor and nerve growth factor receptors. *Mol. Cell. Biol.* **20**, 979–989

42. Wesche, J., Haglund, K. & Haugsten, E.M. (2011). Fibroblast growth factors and their receptors in cancer. *Biochem. J.* **437**, 199–213
43. Wu, Y.-M. et al. (2013). Identification of Targetable FGFR Gene Fusions in Diverse Cancers. *Cancer Discov.* **3**, 636–647
44. Chesi, M. et al. (1997). Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Nat Genet* **16**, 260–264
45. Brooks, A.N., Kilgour, E. & Smith, P.D. (2012). Molecular Pathways: Fibroblast Growth Factor Signaling: A New Therapeutic Opportunity in Cancer. *Clin. Cancer Res.* **18**, 1855–1862
46. Mohammadi, M., Schlessinger, J. & Hubbard, S.R. (1996). Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell* **86**, 577–587
47. Bae, J.H. et al. (2010). Asymmetric receptor contact is required for tyrosine autophosphorylation of fibroblast growth factor receptor in living cells. *Proc. Natl. Acad. Sci.* **107**, 2866–2871
48. Lew, E.D., Bae, J.H., Rohmann, E., Wollnik, B. & Schlessinger, J. (2007). Structural basis for reduced FGFR2 activity in LADD syndrome: Implications for FGFR autoinhibition and activation. *Proc. Natl. Acad. Sci.* **104**, 19802–19807
49. Chen, H. et al. (2013). Cracking the Molecular Origin of Intrinsic Tyrosine Kinase Activity through Analysis of Pathogenic Gain-of-Function Mutations. *Cell Rep.* **4**, 376–384
50. Mohammadi, M. (1997). Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors. *Science* **276**, 955–960
51. Eathiraj, S. et al. (2011). A Novel Mode of Protein Kinase Inhibition Exploiting Hydrophobic Motifs of Autoinhibited Kinases: DISCOVERY OF ATP-INDEPENDENT INHIBITORS OF FIBROBLAST GROWTH FACTOR RECEPTOR. *J. Biol. Chem.* **286**, 20677–20687
52. Huang, Z. et al. (2013). Structural Mimicry of A-Loop Tyrosine Phosphorylation by a Pathogenic FGF Receptor 3 Mutation. *Structure* **21**, 1889–1896
53. Levitzki, A. (1999). Protein Tyrosine Kinase Inhibitors as Novel Therapeutic Agents. *Pharmacol. Ther.* **82**, 231–239
54. Noble, M., Endicott, J. & Johnson, L. (2004). Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **303**, 1800–1805
55. Stark, K.L., McMahon, J.A. & McMahon, A.P. (1991). FGFR-4, a new member of the fibroblast growth factor receptor family, expressed in the definitive endoderm and skeletal muscle lineages of the mouse. *Development* **113**, 641–65
56. Partanen, J. et al. (1991). FGFR-4, a novel acidic fibroblast growth factor receptor with a distinct expression pattern. *EMBO J.* **10**, 1347
57. Vainikka, S. et al. (1994). Signal transduction by fibroblast growth factor receptor-4 (FGFR-4). Comparison with FGFR-1. *J. Biol. Chem.* **269**, 18320–18326
58. Rappolee, D., Basilico, C., Patel, Y. & Werb, Z. (1994). Expression and function of FGF-4 in perimplantation development in mouse embryos. *Dev. Camb. Engl.* **120**, 2259–2269
59. Raffioni, S., Thomas, D., Foehr, E.D., Thompson, L.M. & Bradshaw, R.A. (1999). Comparison of the intracellular signaling responses by three chimeric fibroblast growth factor receptors in PC12 cells. *Proc. Natl. Acad. Sci.* **96**, 7178–7183
60. Zhao, P. et al. (2006). Fgfr4 is required for effective muscle regeneration in vivo. Delineation of a MyoD-Tead2-Fgfr4 transcriptional pathway. *J. Biol. Chem.* **281**, 429–438
61. Marics, I., Padilla, F., Guillemot, J.-F., Scaal, M. & Marcelle, C. (2002). FGFR4 signaling is a necessary step in limb muscle differentiation. *Development* **129**, 4559–4569

62. Bange, J. et al. (2002). Cancer Progression and Tumor Cell Motility Are Associated with the FGFR4 Arg388 Allele. *Cancer Res.* **62**, 840–847
63. Xu, B. et al. (2011). FGFR4 Gly388Arg polymorphism contributes to prostate cancer development and progression: a meta-analysis of 2618 cases and 2305 controls. *BMC Cancer* **11**, 84
64. Serra, S. et al. (2012). The FGFR4-G388R Single-Nucleotide Polymorphism Alters Pancreatic Neuroendocrine Tumor Progression and Response to mTOR Inhibition Therapy. *Cancer Res.* **72**, 5683–5691
65. Sugiyama, N. et al. (2010). Fibroblast Growth Factor Receptor 4 Regulates Tumor Invasion by Coupling Fibroblast Growth Factor Signaling to Extracellular Matrix Degradation. *Cancer Res.* **70**, 7851–7861
66. Sugiyama, N. et al. (2010). FGF receptor-4 (FGFR4) polymorphism acts as an activity switch of a membrane type 1 matrix metalloproteinase-FGFR4 complex. *Proc. Natl. Acad. Sci.* **107**, 15786–15791
67. Roidl, A. et al. (2010). The FGFR4 Y367C mutant is a dominant oncogene in MDA-MB453 breast cancer cells. *Oncogene* **29**, 1543–1552
68. Vi, J.G.T. et al. (2009). Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. *J. Clin. Invest.* doi:10.1172/JCI39703
69. Crose, L.E.S. et al. (2012). FGFR4 Blockade Exerts Distinct Antitumorogenic Effects in Human Embryonal versus Alveolar Rhabdomyosarcoma. *Clin. Cancer Res.* **18**, 3780–3790
70. Stephens, P. et al. (2005). A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet* **37**, 590–592
71. Roidl, A. et al. (2009). Resistance to Chemotherapy Is Associated with Fibroblast Growth Factor Receptor 4 Up-Regulation. *Clin. Cancer Res.* **15**, 2058–2066
72. Turkington, R.C. et al. (2014). Fibroblast growth factor receptor 4 (FGFR4): a targetable regulator of drug resistance in colorectal cancer. *Cell Death Dis* **5**, e1046
73. Gozgit, J.M. et al. (2012). Ponatinib (AP24534), a Multitargeted Pan-FGFR Inhibitor with Activity in Multiple FGFR-Amplified or Mutated Cancer Models. *Mol. Cancer Ther.* **11**, 690–699
74. Li, S.Q. et al. (2013). Targeting Wild-Type and Mutationally Activated FGFR4 in Rhabdomyosarcoma with the Inhibitor Ponatinib (AP24534). *PLoS ONE* **8**, e76551
75. Renhowe, P.A. et al. (2008). 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. *J. Med. Chem.* **52**, 278–292
76. Kim, K.B. et al. (2011). Phase I/II and Pharmacodynamic Study of Dovitinib (TKI258), an Inhibitor of Fibroblast Growth Factor Receptors and VEGF Receptors, in Patients with Advanced Melanoma. *Clin. Cancer Res.* **17**, 7451–7461
77. Liao, R.G. et al. (2013). Inhibitor-sensitive FGFR2 and FGFR3 mutations in lung squamous cell carcinoma. *Cancer Res.* doi:10.1158/0008-5472.CAN-12-3950
78. Marx, V. (2014). Structural biology: “seeing” crystals the XFEL way. *Nat Meth* **11**, 903–908
79. Ccp4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
80. Vagin, A. & Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
81. Winn, M. et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235–242

82. Patterson, A. L. (1934). A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Phys. Rev.* **46**, 372–6
83. Murshudov, G., Vagin, A. & Dodson, E. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D* **53**, 240–255
84. Adams, P. et al. (2010). *{\it PHENIX}*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221
85. Gasteiger, E. et al. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 at <<http://dx.doi.org/10.1385/1-59259-890-0:571>>
86. Pantoliano, M. et al. (2001). High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J Biomol Screen* **6**, 429–440
87. Kabsch, W. (2010). Integration, scaling, space-group assignment and post refinement. 133–144.
88. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201
89. Bordoli, L. et al. (2008). Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**, 1–13
90. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
91. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D* **60**, 2256–2268
92. Delano, W. (2002). The PyMOL Molecular Graphics System. at <<http://www.pymol.org>>
93. <http://www.gimp.org/>.
94. Mohammadi, M. et al. (1996). Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Mol. Cell. Biol.* **16**, 977–989
95. Chen, H. et al. (2007). A Molecular Brake in the Kinase Hinge Region Regulates the Activity of Receptor Tyrosine Kinases. *Mol. Cell* **27**, 717–730
96. Wang, W. et al. (2006). Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3563–3568
97. Ni, Z.-J. et al. (2006). 4-(Aminoalkylamino)-3-benzimidazole-quinolinones as potent CHK-1 inhibitors. *Bioorg. Med. Chem. Lett.* **16**, 3121–3124
98. Byron, S.A. et al. (2013). The N550K/H mutations in FGFR2 confer differential resistance to PD173074, dovitinib, and ponatinib ATP-competitive inhibitors. *Neoplasia N. Y. NY* **15**, 975
99. Zhou, T. et al. (2011). Structural Mechanism of the Pan-BCR-ABL Inhibitor Ponatinib (AP24534): Lessons for Overcoming Kinase Inhibitor Resistance: Structural Mechanism of Ponatinib. *Chem. Biol. Drug Des.* **77**, 1–11
100. Canning, P. et al. (2014). Structural Mechanisms Determining Inhibition of the Collagen Receptor DDR1 by Selective and Multi-Targeted Type II Kinase Inhibitors. *J. Mol. Biol.* **426**, 2457–2470
101. Artim, S.C., Mendrola, J.M. & Lemmon, M.A. (2012). Assessing the range of kinase autoinhibition mechanisms in the insulin receptor family. *Biochem. J.* **448**, 213–220
102. Chiara, F. (2003). Mutations in the met Oncogene Unveil a “Dual Switch” Mechanism Controlling Tyrosine Kinase Activity. *J. Biol. Chem.* **278**, 29352–29358
103. Ezzat, S., Zheng, L., Zhu, X.-F., Wu, G.E. & Asa, S.L. (2002). Targeted expression of a human pituitary tumor-derived isoform of FGF receptor-4 recapitulates pituitary tumorigenesis. *J. Clin. Invest.* **109**, 69–78

104. Ding, L. et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075
105. Lis, A. et al. (2007). CRACM1, CRACM2, and CRACM3 Are Store-Operated Ca²⁺ Channels with Distinct Functional Properties. *Curr. Biol.* **17**, 794–800
106. Eskandari, S., Wright, E.M., Kreman, M., Starace, D.M. & Zampighi, G.A. (1998). Structural analysis of cloned plasma membrane proteins by freeze-fracture electron microscopy. *Proc. Natl. Acad. Sci.* **95**, 11235–11240
107. Singer, S. & Nicolson, G. (1972). The fluid mosaic model of the structure of cell membranes. *Science* **175**, 720–731
108. Alberts, B. et al. (Garland Science: 2009). *Essential Cell Biology*. at <<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0815341296>>
109. Berridge, M.J., Lipp, P. & Bootman, M.D. (2000). The versatility and universality of calcium signalling. *Nat Rev Mol Cell Biol* **1**, 11–21
110. Parekh, A.B. (2010). Store-operated CRAC channels: function in health and disease. *Nat. Rev. Drug Discov.* **9**, 399–410
111. Dziadek, M.A. & Johnstone, L.S. (2007). Biochemical properties and cellular localisation of STIM proteins. *New Dev. Ca2 Influx Channels Their Regul.* **42**, 123–132
112. Liou, J., Fivaz, M., Inoue, T. & Meyer, T. (2007). Live-cell imaging reveals sequential oligomerization and local plasma membrane targeting of stromal interaction molecule 1 after Ca²⁺ store depletion. *Proc. Natl. Acad. Sci.* **104**, 9301–9306
113. Putney, J.W. (2010). Pharmacology of store-operated calcium channels. *Mol. Interv.* **10**, 209
114. Penna, A. et al. (2008). The CRAC channel consists of a tetramer formed by Stim-induced dimerization of Orai dimers. *Nature* **456**, 116–120
115. Prakriya, M. et al. (2006). Orai1 is an essential pore subunit of the CRAC channel. *Nature* **443**, 230–233
116. Maruyama, Y. et al. (2009). Tetrameric Orai1 Is a Teardrop-shaped Molecule with a Long, Tapered Cytoplasmic Domain. *J. Biol. Chem.* **284**, 13676–13685
117. Takahashi, Y. et al. (2007). Essential role of the N-terminus of murine Orai1 in store-operated Ca²⁺ entry. *Biochem. Biophys. Res. Commun.* **356**, 45–52
118. Fahrner, M. et al. (2009). Mechanistic view on domains mediating STIM1–Orai coupling. *Immunol. Rev.* **231**, 99–112
119. Fukushima, M., Tomita, T., Janoshazi, A. & Putney, J.W. (2012). Alternative translation initiation gives rise to two isoforms of Orai1 with distinct plasma membrane mobilities. *J. Cell Sci.* **125**, 4354–4361
120. Mignen, O., Thompson, J.L. & Shuttleworth, T.J. (2008). Orai1 subunit stoichiometry of the mammalian CRAC channel pore. *J. Physiol.* **586**, 419–425
121. Gwack, Y. et al. (2007). Biochemical and Functional Characterization of Orai Proteins. *J. Biol. Chem.* **282**, 16232–16243
122. Hou, X., Pedi, L., Diver, M.M. & Long, S.B. (2012). Crystal Structure of the Calcium Release-Activated Calcium Channel Orai. *Science* **338**, 1308–1313
123. Feske, S. et al. (2006). A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185
124. Gaspar, H.B., Gilmour, K.C. & Jones, A.M. (2001). Severe combined immunodeficiency—molecular pathogenesis and diagnosis. *Arch. Dis. Child.* **84**, 169–173

-
125. Huang, Y.H., Hoebe, K. & Sauer, K. (2008). New therapeutic targets in immune disorders: ItkB, Orai1 and UNC93B. *Expert Opin. Ther. Targets* **12**, 391–413
 126. Zhang, S.L. et al. (2011). Mutations in Orai1 transmembrane segment 1 cause STIM1-independent activation of Orai1 channels at glycine 98 and channel closure at arginine 91. *Proc. Natl. Acad. Sci.* **108**, 17838–17843
 127. Soboloff, J., Rothberg, B.S., Madesh, M. & Gill, D.L. (2012). STIM proteins: dynamic calcium signal transducers. *Nat Rev Mol Cell Biol* **13**, 549–565
 128. Soboloff, J., Madesh, M. & Gill, D.L. (2011). Sensing cellular stress through STIM proteins. *Nat Chem Biol* **7**, 488–492
 129. Muik, M., Schindl, R., Fahrner, M. & Romanin, C. (2012). Ca²⁺ release-activated Ca²⁺ (CRAC) current, structure, and function. *Cell. Mol. Life Sci.* **69**, 4163–4176
 130. Park, C.Y. et al. STIM1 Clusters and Activates CRAC Channels via Direct Binding of a Cytosolic Domain to Orai1. *Cell* **136**, 876–890

