Functional and Structural Analysis of RET Receptor Tyrosine Kinase and Its Complexes

Yixin Liu

Division of Biochemistry
Department of Biosciences
Molecular and Integrative Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Functional and structural analysis of RET receptor tyrosine kinase and its complexes

Yixin Liu

Division of Biochemistry
Department of Biosciences
Molecular and Integrative Biosciences
Faculty of Biological and Environmental Sciences
and
Doctoral Programme in Integrative Life Science
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki at 3 pm on 12th May 2021.

Helsinki, 2021
Supervisor  Professor Adrian Goldman, Ph.D.
Molecular and Integrative Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Thesis Committee  Docent Ville Paavilainen, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Professor Kari Keinänen, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Reviewers  Professor Lari Lehtiö, Ph.D.
Faculty of Biochemistry and Molecular Medicine
University of Oulu
Oulu, Finland

Adjunct Professor Tiina A. Salminen, Ph.D.
Faculty of Science and Engineering
Åbo Akademi University
Turku, Finland

Opponent  Professor Helen Walden, Ph.D.
Institute of Molecular Cell & Systems Biology
University of Glasgow
Glasgow, United Kingdom

Custos  Professor Kari Keinänen, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

https://ethesis.helsinki.fi/

Cover layout by Yixin Liu
Painosalama Oy, Turku, Finland
ACKNOWLEDGEMENTS

First of all, I would like to express my greatest appreciation to my supervisor Prof. Adrian Goldman for his guidance and supervision throughout the entire Ph.D. work. I am thankful for the freedom I have in conducting research work, all the project discussions and the opportunities in trying new techniques and approaches in the project. I deeply appreciated the understanding and encouragement when things were not going as planned as well as the support in my personal and career development.

I am grateful to my thesis committee members, Prof. Kari Keinänen and Docent Ville Paavilainen for the discussions and advice they gave during my Ph.D. studies and the recommendations and support for my grant and job applications. My warm thanks to Kari for his help in sorting the numerous documents during the dissertation submission process and for serving as custos for the public examination. I also wish to thank Prof. Helen Walden for accepting my invitation to be the opponent for my Ph.D. defence. I would also like to thank my thesis pre-examiners, Prof. Lari Lehtio and Adj. Prof. Tiina A. Salminen, for reviewing my thesis. The valuable suggestions and discussions were enlightening and gave me a lot of encouragement.

This Ph.D. project was carried out in the Division of Biochemistry, Molecular and Integrative Biosciences in the Faculty of Biological and Environmental Sciences. I am grateful for the excellent research facilities and equipment provided by the department and group leaders. I would like to thank the Doctoral School of Health Sciences and Doctoral Programs of Integrative Life Sciences and Brain & Mind for organizing the lectures, seminars and conferences that are invaluable for the development of my scientific skills and personal competence.

I’m extremely grateful to Dr. Stephen Muench for being my supervisor during my research exchange at the University of Leeds, sharing his profound knowledge in electron microscopy and providing suggestions, guidance and support throughout the project. I must also thank the EM facility personnel, Dr. Anna Higgins, David Klebl from the Muench group, Brendan and Dr. Miriam Walden at the University of Leeds for their help in the cryo-EM work. Many thanks to all the co-authors and collaborators for their contributions to the studies in this work!

It has been my great pleasure to work with the past and current members of the Goldman group in Helsinki, Keni V., Orquidea DCR., Tuukka E., Bernadette G., Tuulia S., Heidi K., Violeta M., Kornelia M. as well as those in Leeds, Alex H., Andreas K., Ana P., Brendan F., Claudia S., Jack W., Jannik S., James H., Jessica B., Jacob E., Maria N. and Maren T.. I am thankful for their patience in helping me to get started in the lab when I first joined, the help along the way, and the inspiring and joyful discussions on research
and non-research topics during our group meetings, lunch breaks and outings over food, cakes and drinks. I very much appreciate the advice and help from Ville and Prof. Sarah Butcher, Prof. Mart Saarma and their group members especially Juho K., Shahid R., Ausra D., James G., Justin F., Lauri P., Masha A. and Sergei G..

Over the past several years, I have also been fortunate enough to have met a number of friends who have given me lots of support and filled my life with joy. My sincere thanks to Pu C., Man X., Jue W., Linxiao C., Yinyin W. for the climbs, chats, cards and codes. Warm thanks to my past roommates Xuan Z., Wei L., and Chao Z. for the countless delicious meals and non-science talks and for sharing the apartments with me and making them our sweet (second) home. Thanks also to my lovely neighbours Lin C., Zilan W., Xueyun F. and Jingwen X. I am also thankful to Xiaonan L. for your patience in answering all my graduation-related questions and to Hui Z., Petri N., Wenfei L., for our long chats and walks. Kiitos to Sampo C. for being my dance partner and I don’t think I would have had such a nice dancing experience in Helsinki without your company.

My special thanks to my dear husband Gregory Craven. Thank you, Greg, for all the discussions, ideas and revisions and all the cups of tea even though you knew I was never going to drink them. Thank you for your unconditioned support and love that helped me get through the toughest times. And I need to thank Mio, the most wonderful cat in the world, for the cuddles, distractions and wake-up meows. Last but not the least, I wish to express my deepest gratitude to my mother, stepfather and my family. Thank you for believing in me and I always feel so powerful thinking about you and knowing that you are always there for me.

The journey of pursuing a Ph.D. is incredible and unforgettable to me. It is about hard work, curiosity and exploration. It is about being persistent not stubborn, accepting challenges and failure and having faith in the work and myself. I cannot have done this without you, my colleagues, friends and dearest family. I thank you all from the bottom of my heart.
# TABLE OF CONTENTS

## ACKNOWLEDGEMENTS

## ABSTRACT

## LIST OF ORIGINAL PUBLICATIONS

## ABBREVIATIONS

## I INTRODUCTION

1. **RET receptor tyrosine kinase**
   1.1 Receptor tyrosine kinases ................................................................. 10
   1.2 RET receptor tyrosine kinase: Ligand-induced activation .......................... 11
       1.2.1 Structural basis of the RET extracellular domain ......................... 12
       1.2.2 Extracellular domain binding ...................................................... 14
       1.2.3 Intracellular domain activation and downstream signalling .............. 18
       1.2.4 The biological function of RET and its complexes ......................... 21
   1.3 RET dysregulation ................................................................................. 22
       1.3.1 Loss-of-function ............................................................................. 22
       1.3.2 Gain-of-function ............................................................................. 22

2. **Therapies targeting RET and its complexes** ....................................... 26
   2.1 Kinase domain targeting ..................................................................... 26
   2.2 Extracellular domain targeting ......................................................... 29

3. **Structural investigation using electron microscopy** .............................. 31
   3.1 Electron microscopy sample preparation ............................................. 32
       3.1.1 Negative stain EM sample preparation ........................................... 32
       3.1.2 Cryo-electron microscopy sample preparation ............................... 33
   3.2 Electron microscopy data acquisition and processing ........................... 35

4. **Studies of full-length RET and other RTKs** ......................................... 37
   4.1 Membrane protein expression, solubilization and purification .............. 37
   4.2 Structural studies of full-length RTKs .................................................. 39
   4.3 Structural insights into full-length RET activation ............................... 44

## II AIMS OF THE STUDY

## III MATERIALS AND METHODS

1. **DNA constructs** .................................................................................... 47

2. **Cell Lines and Cell Culture** ................................................................ 48

3. **Methods and Assays** .......................................................................... 48

4. **Programs** ............................................................................................ 49
IV RESULTS AND DISCUSSION

1 RETWT activation by extracellular ligand binding
   1.1 Investigating the interaction between RETECD and efna-A5ECD .................. 50
   1.2 Characterization of the RETECD/GDF15/GFRALECD complex .................. 52
      1.2.1 Expression of functional human RETECD .................................. 52
      1.2.2 Structural investigation of the hRETECD/hGDF15/hGFRALECD complex .. 54

2 RET activation via the oncogenic C634R mutation
   2.1 The dimeric oncogenic hRETECD C634R mutant .................................. 59
      2.1.1 Expression of the dimeric hRETECD(C634R) ................................ 59
      2.1.2 Structural investigation of dimeric hRETECD(C634R) ..................... 61
   2.2 The liganded hRETECD(C634R) complex ........................................ 62
      2.2.1 Characterization of the hRETECD(C634R)/hGDF15/hGFRALECD complex .. 62
      2.2.2 The structural investigation of the hRETECD(C634R) complex .......... 62

3 Expression and purification of full-length RET

V CONCLUSION AND FUTURE STUDIES

VI REFERENCES

PUBLICATIONS
ABSTRACT

The RET receptor tyrosine kinase is a versatile receptor which responds to multiple signalling cues. The signalling of RET plays a central role in cell proliferation, differentiation and maintenance. Its dysregulation contributes to various human diseases through gain- and loss-of-function, such as Hirschsprung’s disease and thyroid and lung cancers. Mutations at the cysteine-rich domain (CRD) of RET lead to constitutive activation through receptor dimerization via an intermolecular disulfide bond. However, due to the lack of functional and structural studies of RET and its ligands, the molecular mechanisms of RET activation under normal and pathological conditions is unclear. In this thesis project, a combination of biochemical and biophysical tools are used to characterize the activation of RET with an emphasize on three aspects: studying (1) its extracellular domain (ECD) complex formation with ephrin As (efn-As) and the growth and differentiation factor 15 (GDF15)/the glial cell-derived neurotrophic factor receptor α-like (GFRAL), (2) the oncogenic C634R mutation, which accounts for the majority of the multiple endocrine neoplasia type 2A (MEN2A) cases and (3) the solubilization and purification of full-length RET.

To explore the binding between RET and efn-As, I first demonstrated that functional zebrafish RET\textsuperscript{ECD} (zRET\textsuperscript{ECD}) and efn-A5 (zefn-A5) can be expressed using insect cells. In contrast to previous cellular studies, I showed that zRET\textsuperscript{ECD} does not interact directly with zefn-As and postulated that another binding partner is required to mediate their interaction for the reverse signalling of efn-As. Unlike zRET\textsuperscript{ECD}, I found that human RET\textsuperscript{ECD} (hRET\textsuperscript{ECD}) expressed in insect cells is non-functional; therefore, I established a mammalian expression platform for functional expression of hRET\textsuperscript{ECD}, demonstrating its superiority to insect cell expression in which hRET is prone to misfold. Using recombinantly expressed proteins, I reconstituted the wild-type (WT) hRET\textsuperscript{ECD}/hGDF15/hGFRAL\textsuperscript{ECD} complex and was able to reconstruct an 8-Å cryo-EM map of the complex, which revealed a “butterfly”-shaped tripartite complex. For the expression of the hRET\textsuperscript{ECD}(C634R) mutant, I used a C-terminal Fc tag to successfully drive the receptor dimer formation, as the mutant protein is otherwise monomeric. I studied the resulting apo-hRET\textsuperscript{ECD}(C634R) dimer and its complex with GDF15/GFRAL\textsuperscript{ECD} using structural tools. Comparison of the cryo-EM structure of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex and the negative stain EM maps of the apo-hRET\textsuperscript{ECD}(C634R) dimer and the mutant hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} complex revealed significant conformational changes that provide a structural model for the oncogenic activation of RET. Finally, I established a membrane protein production pipeline for full-length RET that will enable future structural studies in a more biologically native system.
LIST OF ORIGINAL PUBLICATIONS

This thesis contains material from the following original articles that referred to in the text by their Roman numerals I-IV. The rights have been granted by publishers to include the material in dissertation.


IV. **Liu Y**, de Castro Ribeiro O, Craven GB, Muench SP, Goldman A. A structural basis for the constitutive activation of the oncogenic C634R mutant RET. *Manuscript.*
ABBREVIATIONS

2D  2-Dimensional
3D  3-Dimensional
AL  Activation loop
ARTN  Artemin
ATP  Adenosine triphosphate
BBB  Blood–brain barrier
BN PAGE  Blue-native polyacrylamide gel electrophoresis
CHS  Cholesteryl hemisuccinate
CLD  Cadherin-like domain
CMC  Critical micelle concentration
CRD  Cysteine-rich domain
CTF  Contrast transfer function
DDM  N-Dodecyl-β-D-maltoside
DTT  Dithiothreitol
ECD  Extracellular domain
EGFR  Endothelial growth factor receptor
EM  Electron microscopy
Ephrin A  Efn-A
FDA  The U.S. Food and Drug Administration
FH/HF  Flag-Hiss/Hiss-Flag tags
GDF15  Growth and differentiation factor 15
GDNF  Glial cell-derived neurotrophic factor
GFL  Glial cell-derived neurotrophic factor family ligand
GFRα  GDNF family receptor α
GFRAL  GDNF family receptor α-like
GPI  Glycosylphosphatidylinositol
GRL  Glycine-rich loop
HIS  Hiss tag
HSCR  Hirschsprung’s disease
ICD  Intracellular domain
IGF-1R  Type 1 insulin-like growth factor receptor
IR  Insulin receptor
KIT  Stem cell growth factor receptor Kit
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBD</td>
<td>Ligand-binding domain</td>
</tr>
<tr>
<td>LMNG</td>
<td>Lauryl maltose neopentyl glycol</td>
</tr>
<tr>
<td>MEN2</td>
<td>Multiple endocrine neoplasia type 2</td>
</tr>
<tr>
<td>MTC</td>
<td>Medullary thyroid cancer</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NS-EM</td>
<td>Negative stain electron microscopy</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Non-small-cell lung cancer</td>
</tr>
<tr>
<td>NTRN</td>
<td>Neurturin</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymer chain reaction</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PDGFR</td>
<td>Platelet-derived growth factor receptor</td>
</tr>
<tr>
<td>PSPN</td>
<td>Persephin</td>
</tr>
<tr>
<td>RET</td>
<td>Rearranged during transfection</td>
</tr>
<tr>
<td>RTK</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>SDS PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>MALS</td>
<td>Multiangle static laser light scattering</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus protease</td>
</tr>
<tr>
<td>TK</td>
<td>Tyrosine kinase domain</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane domain</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>Vascular endothelial growth factor receptor 2</td>
</tr>
<tr>
<td>WB</td>
<td>Western blotting</td>
</tr>
</tbody>
</table>
### ABBREVIATIONS OF AMINO ACIDS

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>C</td>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>D</td>
<td>Asp</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>E</td>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>F</td>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>G</td>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>H</td>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>I</td>
<td>Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>K</td>
<td>Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>M</td>
<td>Met</td>
<td>Methionine</td>
</tr>
<tr>
<td>N</td>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>P</td>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>Q</td>
<td>Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>R</td>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>S</td>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>T</td>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>W</td>
<td>Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>V</td>
<td>Val</td>
<td>Valine</td>
</tr>
<tr>
<td>Y</td>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>
I INTRODUCTION

1 RET receptor tyrosine kinase

1.1 Receptor tyrosine kinases

Across the human genome, 90 genes have been identified that encode protein tyrosine kinases, including 58 receptor tyrosine kinases (RTKs) and 32 non-receptor tyrosine kinases (1). The RTKs are subdivided into 20 families and control a variety of cellular processes, ranging from cell migration and proliferation to differentiation and maintenance (2). Despite their functional differences, RTKs share a similar structural architecture: they have an extracellular domain (ECD) for ligand recognition, a single-span transmembrane (TM) domain and an intracellular tyrosine kinase (TK) domain followed by a carboxy (C-) terminal region (Figure 1).

![Diagram representing the activation mechanism of RTKs]

**Figure 1. Schematic representation of a general activation mechanism for receptor tyrosine kinases (RTKs).** Most RTKs are monomeric in the apo-state, with some exceptions that exist in inactive dimeric or oligomeric states, for example the insulin receptor. Binding of ligands at the extracellular domain (ECD) of RTKs induces receptor dimerization, which leads to primary trans-phosphorylation at one or more tyrosine residue(s) at the tyrosine kinase (TK) domain. The primary phosphorylation activates the receptor, inducing secondary auto-phosphorylation of the TK domain. The phospho-tyrosines recruit adaptor proteins and facilitate substrate phosphorylation, inducing the downstream signalling.

Typically, the activation of the RTKs requires ligand-induced receptor dimerization, although a small number of RTKs are readily oligomeric in their apo-state. For example, the insulin receptor (IR) exists as disulfide-crosslinked dimer in the absence of its ligand
and the binding of the ligand induces significant conformational change of the receptor for its activation (3,4). In either case, the association of the ligands stabilizes the active conformation of the receptor, whose mode-of-activation can be summarized into four major categories: (i) Binding of the dimeric ligands induces receptor dimerization with no direct interaction between the receptors, for instance the TrkA receptor (5); (ii,iii) ligand binding mediates receptor dimerization with direct interactions between the receptors without (ii) or with (iii) accessory molecules contributing to the dimer interface, the examples of which are the KIT (6) receptor and the fibroblast growth factor receptors (FGFR) (7), respectively; (iv) receptor-driven dimerization with the ligand-binding site away from the dimer interface, such as EGFR (8) and IR (3,4).

Upon the binding of their corresponding ligands through the extracellular domain, RTKs undergo conformational changes, which results in TK activation via the destabilization of autoinhibitory subunits or by activating trans-autophosphorylation or a combination of both (9). There are multiple tyrosine residues in the TK domains of RTKs and the autophosphorylation takes place in a sequential manner (Figure 1). Typically, primary autophosphorylation is seen at the activation loop, which stabilizes the enzymatically activate conformation of the kinase domain to improve its catalytic activity, except for EGFR and RET, which will be explained later in this thesis. Subsequently, the first phosphorylation event induces secondary trans-phosphorylation of the tyrosine kinase at various positions and regions, including the juxtamembrane (JM) and C-terminal regions, which have also been shown to contribute to the release of auto-inhibition in i.e. KIT (10) and Tie2 (11) receptors. The phospho-tyrosines then serve as binding sites for the adaptor and signalling proteins that contain Src homology domain 2 (SH2) or phosphotyrosine-binding domain (12), thus facilitating substrate phosphorylation and initiating intracellular signalling.

1.2 RET receptor tyrosine kinase: Ligand-induced activation

This thesis focuses on the rearranged during transfection (RET) receptor tyrosine kinase. RET was first discovered in 1985 as an oncogene that is activated through DNA recombination (13). RET is the only member in the RET receptor family of RTKs. It is evolutionally conserved and its gene is found in the genome of all vertebrates as well as some invertebrates, including Drosophila melanogaster (14) and Amphioxus (15). The signalling of RET is implicated in a variety of cellular processes and is required for the development of the urogenital and nervous systems (16,17). Dysregulation of RET is a driver in multiple diseases, such as cancers, Parkinson’s disease and obesity. Relative to the other RTKs, the activation mechanism of RET is unique in that the extracellular domain of RET must bind to a pre-formed complex containing a pair of membrane-bound co-receptors in addition to the soluble ligand dimer, generating a functional tripartite complex. In vertebrates, at least six different ligands are known to signal through RET,
including glial cell-derived neurotrophic factor (GDNF) family ligands (GFLs), growth and differentiation factor 15 (GDF15) and ephrin A ligands (efn-As), making RET a highly versatile receptor.

1.2.1 Structural basis of the RET extracellular domain

At the tertiary structural level, RET contains a single-span TM domain and a TK domain that are conserved with the other RTK families but its ECD is composed of unique structural subunits. The RETECD comprises four cadherin-like domains (CLDs 1-4) and a cysteine-rich domain (CRD) (Figure 2A). High-resolution cryo-electron microscopy (cryo-EM) structures, which were only reported at a late stage of my Ph.D. research, have revealed that RETECD exhibits an unusual “C” shape (Figure 2B, left panel) that is stabilized by calcium binding motifs and extensive disulfide bonding in addition to typical hydrophobic effect and hydrogen bonding driven folding (18). The CLDs contain approximately 110 residues each and all exhibit a cadherin-like structure consisting of five or seven β-strands, which are packed as two opposing β-sheets (19). Similar to cadherins, the activity of RETECD requires the binding of calcium ions. In agreement with previous molecular modelling (20) and functional (21) studies, two calcium binding sites have been observed in the cryo-EM structures, one between the CLD2 and CLD3, which binds up to three calcium ions, and the other one in the CRD (18), where the single cationic calcium is surrounded by multiple negatively charged amino acids (Figure 2B, right panel). The CRD of RET contains 120 residues, of which (for mammalian RET) 16 are cysteines that form eight intramolecular disulfide bonds (Figure 2C). The RETECD contains 11 glycosylation sites and is extensively glycosylated. The correct glycosylation of RETECD have been shown to be critical for RET folding and stability (22–24).

Following the CRD, the single-helical TM domain of RET is composed of 22 amino acids (aa) and links to the intracellular juxtamembrane (JM) region of 50 aa (25). After the JM, there is the TK domain of about 300 aa, which is followed by a C-terminal region of approximately 100 aa. Three isoforms of RET exist due to alternative splicing of the 3'-exon, namely RET9, RET51 and, less commonly, RET43 that contain 9, 51 or 43 additional amino acids respectively after residue 1063 at the C-terminus, each with a unique sequence (Figure 2A) (26,27). It appears that most tissues can expressed different RET isoforms but the expression level of each isoform varies, with RET9 being the most abundant while RET43 the least (26,28). Furthermore, studies have shown that splicing variants have different trafficking and maturation properties and are capable of inducing distinct autophosphorylation patterns, influencing the intracellular signalling outcomes (26,29–31).
**Figure 2.** The architecture of RET receptor tyrosine kinase. A) Diagrammatic representation of full-length RET and alternative splicing variants. B) RET extracellular domain (left panel) and expanded view showing the calcium binding sites (right panel). C) Structure of the CRD of RET. CLD: cadherin-like domain; CRD: cysteine-rich domain; TM: transmembrane domain; TK: tyrosine kinase domain; SP: signal peptide; CBS: calcium binding site.
1.2.2 Extracellular domain binding

The GFLs and GDF15 are divergent members of the transforming growth factor-β (TGF-β) superfamily. They are expressed as prepro-proteins, processed in the ER to pro-proteins, subsequently followed by proteolytic digestion and extracellular secretion to generate mature proteins (32–35). There are four GFLs, namely GDNF, neurturin (NTRN), artemin (ARTN) and persephin (PSPN). The GFLs and GDF15 share a similar fold and all function as homodimers (Figure 3, upper panel), linked via a disulfide bond. Each of the GFLs contains 7 conserved cysteines while GDF15 contains 9, one of which forms the intermolecular disulfide bond responsible for the dimer formation while the rest form intramolecular disulfide bonds. For the purpose of this thesis, GDF15 is considered part of the GFLs.

![Figure 3. Structures of the GDNF family ligands and GDF15 and their complexes with GFRα family receptors. Cartoon comparison of GDF15 (pink), GDNF (light blue), NRTN (grey) and ARTN (beige) (upper panel) as well as GDF15/GFRAL (pink and red, PDB code 5VZ4), GDNF/GFRα1 (light blue and]
green, PDB code 2V5E), NRTN/GFRα2 (silver and yellow, PDB code 5MR4) and ARTN/GFRα3 (beige and orange, PDB code 2GH0) complexes.

The GDNF family receptors α (GFRα) consist of four members, GFRα1-4, which are anchored to the cell membrane by glycosylphosphatidylinositol (GPI). It is suggested that GPI-anchored GFRαs are enriched in the lipid rafts and further recruit RET, facilitating the increased local concentration of GFRαs and RET in the lipid rafts, favouring the complex formation (36–38). GFRα-like (GFRAL), a remote member of the GFRα family, is a transmembrane protein with a 23-aa intracellular region. It is not yet known what roles the transmembrane and intracellular regions of GFRAL play. Mammalian GFRα1-3 (as well as most vertebrate GFRαs) and GFAL contain three homologous domains while GFRα4 only has two, lacking the first domain. These domains contain mainly α-helices with five intramolecular disulfide bonds within domain 2 and 3 and four in domain 1, exhibiting a unique fold (39). Although certain levels of cross-talk exist, each of the GFLs, GDNF, neuturin (NRTN), artemin (ARTN) and perspech (PSPN), selectively binds with high affinity to their respective co-receptors (Table 1), forming heterodimeric complexes (Figure 3, lower panel) (40–46). Similarly, GDF15, remotely related to the GFLs, exclusively interacts with its orphan co-receptor GFRAL (45,47–49).

### Table 1. Measured binding between GFLs and their receptor(s) using different techniques.

All the affinities were reported as dissociation constants (Kd) but the values should be individually assessed based on the different techniques used. Methods: a) surface plasmon resonance (SPR); b) cell-based binding assay using radio-labelled ligands (competition); c) cell-based binding assay using radio-labelled ligands (non-competition); d) cell-based binding assay using unlabelled ligands; e) enzyme-linked immunosorbent assay (ELISA); f) other cell-based assays (i.e., immunoblotting and immunofluorescence); g) reporter gene activation assay; h) isothermal titration calorimetry (ITC).

<table>
<thead>
<tr>
<th>GFRα1</th>
<th>GDNF</th>
<th>NRTN</th>
<th>ARTN</th>
<th>PSPN</th>
<th>GDF15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63 nM (a, GFRα1-Fc) (40)</td>
<td>1.04 nM (a, GFRα1-Fc) (40)</td>
<td>Interaction observed (f) (50)</td>
<td>No interaction (f, g) (45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 pM (b) (51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 pM (b) (52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.47 nM (e) (53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 nM (e, GFRα1-Fc) (54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GFRα2</th>
<th>GDNF</th>
<th>NRTN</th>
<th>ARTN</th>
<th>PSPN</th>
<th>GDF15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 nM (e) (53)</td>
<td>0.94 nM (a, GFRα2-Fc) (40)</td>
<td>Interaction observed (f) (55)</td>
<td>No interaction (f, g) (45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 nM in the presence of RET (e, GFRα2-Fc) (54)</td>
<td>220 nM (h) (44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 pM (b) (52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GFRα3</th>
<th>GDNF</th>
<th>NRTN</th>
<th>ARTN</th>
<th>PSPN</th>
<th>GDF15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction observed in the</td>
<td>200 nM (d) (56)</td>
<td></td>
<td>No interaction (f, g) (45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90 nM (h) (41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFRα4</td>
<td>Interaction observed (f) (55)</td>
<td>Interaction observed (f) (55)</td>
<td>Interaction observed (f) (55)</td>
<td>6 nM (a, GFRα4-Fc) (42)</td>
<td>No interaction (f, g) (45)</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>GFRAL</td>
<td>No interaction (b, g) (45)</td>
<td>No interaction (b, g) (45)</td>
<td>No interaction (b, g) (45)</td>
<td>330 pM (c) (45)</td>
<td>8 nM (a) (45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.5 pM (a, GFRα3-Fc) (47)</td>
<td>25.8 nM (a) (49)</td>
</tr>
</tbody>
</table>

Interestingly, although these GFLs are selective towards their co-receptors and their complexes are structurally different (Figure 3, lower panel), the ligand/co-receptor complexes all bind to RET for downstream signalling. At the beginning of this thesis project, no high or medium resolution structural information was available for any of the RET complexes and it was not clear how the various ligand and co-receptor pairs bind to RET. As such, studying these complexes by cryo-EM was one of the aims of this thesis project (Objective 1). Then, in late 2019 and after I had made significant progress in this regard, the cryo-EM structures of the RET/GDF15/GFRAL, RET/GDNF/GFRα1, RET/NRTN/GFRα2 and RET/ARTN/GFRα3 complexes were solved (18,58). These models showed that all the complexes exhibit a butterfly-like conformation and that the N-terminal end of RET<sup>ECD</sup> interacts with the co-receptor GFRαs while the C-terminal end of RET<sup>ECD</sup> binds to the GFLs (Figure 4A). Comparing the conformation of RET in complexes with the different ligand and co-receptor pairs reveals the relative structural rigidity of RET (Figure 4B). The angles between the two wings of the complexes are determined by the conformation of the GFLs as well as their interaction interface with the co-receptors (18,43–45,59) and such differences may play an important role in controlling the selective activation of RET in response to specific GFLs, leading to different signalling outcomes (Figure 5).
Figure 4. RET binding at the extracellular domain. A) Cartoon representation of RET/GDF15/GFRαL (grey, pink and red, PDB code 6Q2J), RET/GDNF/GFRα1 (grey, light blue and green, PDB code 6Q2N), RET/NRTN/GFRα2 (grey, silver and yellow, PDB code 6Q2O) and RET/ARTN/GFRα3 (grey, beige and orange, PDB code 6Q2S). B) Alignment of the extracellular domains of RET of four complex structures (PDB code 6Q2J, 6Q2N, 6Q2O and 6Q2S).

Figure 5. The RET receptor tyrosine kinase and its signalling network. RET is a transmembrane receptor for a variety of ligands, including four glial cell-derived neurotrophic factor (GDNF) family ligands (GFLs), namely GDNF, neurturin (NTRN), artemin (ARTN) and persephin (PSPN), growth and
differentiation factor 15 (GDF15) as well as ephrin A ligands (efn-As). GFLs and GDF15 first bind to their
gDNF family receptors α (GFRα1-4, shown as GFRA1-4 in this figure) and GFRα-like (GFRAL) and
subsequently bind to RET to form heterohexameric complexes. Upon ligand-induced activation, a range of
tyrosine residues at the tyrosine kinase (TK) domain become phosphorylated (P), leading to intracellular
signalling. Structural domains are indicated by the key. CLD: cadherin-like domain; CRD: cysteine-rich
domain; TM: transmembrane domain; GPI: glycosylphosphatidylinositol.

Apart from the GFLs, ephrin A ligands have also been reported to signal through RET
for its reverse signalling via the EphA receptors (60,61). The ephrin A ligands (efn-As)
belong to the ephrin ligands family, which are classified based on their binding
preference towards the EphA family receptors (62). The signalling of Eph receptors and
ephrin ligands is bidirectional and occurs at cell junctions. In particular, Eph A receptors
bind with efn-As in trans on the interacting cell membranes to transduce signal from
either EphAs or efn-As (Figure 5) (63,64). Signal transmission from the efn-expressing
cells to the Eph receptor-expressing cells is called the “forward” signalling while the term
“reverse” signalling is used for signalling into the efn-expressing cell. Similar to the
GFRαs, efn-As are also GPI-anchored membrane proteins. Due to the lack of an
intracellular domain, efn-As require the engagement of other transmembrane proteins to
signal. Interestingly, RET was identified as a receptor for efn-As through cellular assays,
mediating the reverse signalling of efn-As for motor neuron pathfinding (60,61,65). However, it was not clear whether efn-As and RET interact directly, which requires
further investigation.

1.2.3 Intracellular domain activation and downstream signalling

The structures of protein kinases can be divided into 12 subdomain features according to
their sequence (66). The first subdomain contains a glycine-rich loop (GRL), covering
the ATP binding site. The αC helix is located in the third subdomain and is essential in
maintaining a kinase-active conformation. The sixth subdomain contains a catalytic loop
“HRDLAARN” while the seventh and eighth subdomains harbour a 30-aa activation loop
(AL). When a kinase is activated, the AL adopts an active conformation for substrate
binding, which then interact with the “HRD” motif in the catalytic loop. The AL usually
starts from a conserved “DFG” motif. In the active conformation, the aspartic acid residue
of the DFG motif is poised to bind the Mg²⁺ at the ATP binding site while the
phenylalanine residue points outwards. However, in the inactive conformation, the
position of the two residues is swapped, resulting in the phenylalanine residue occupying
the active site. Thus, the orientation of the DFG motif is generally used to describe the
inactive (DFG-out) and active (DFG-in) conformations of a kinase (67).

All the TK domains of RTKs share a similar fold, containing both N- and C-lobes and are
structurally similar in their active conformations, which is required for catalytic activity.
However, the structures of the inactive RTK\textsuperscript{TK}s are more varied; therefore, the activation mechanisms of the TK domains vary from receptor to receptor. There are three major \textit{cis}-autoinhibitory mechanisms of the TK domain, including the inhibition \textit{via} the (i) the activation loop (AL), (ii) the juxtamembrane (JM) and (iii) the C-terminal tail. In each case, the activation loop or certain parts of the JM or C-terminal tail regions (or a combination of each) adopts an auto-inhibitory conformation which hinders kinase activity, either by stabilizing the inactive conformation (10) or hindering substrate binding (68). In response to ligand binding, RTKs undergo conformational changes to activate. As described earlier, most RTKs show early auto-phosphorylation at its AL to release \textit{cis}-auto-inhibition and improve kinase catalytic activity. For instance, in the case of the insulin receptor, three tyrosine residues Y1158, Y1162 and Y1163 located at the AL of the TK domain exhibit an auto-inhibited conformation in the apo state, blocking the active site of the kinase (Figure 6A). Binding of its ligand, insulin, at the extracellular domain of IR activates the receptor and induces the trans-autophosphorylation of Y1158, Y1162 and Y1163, leading to significant conformational change of the AL and opening up the active site for substrate binding (69,70). Apart from the AL, the \textalpha\textsubscript{C}-helix also reorients to stabilize the active conformation.

There are 18 and 16 tyrosine residues in RET\textsubscript{51}\textsuperscript{ICD} and RET\textsubscript{9}\textsuperscript{ICD}, respectively, with Tyr1090 and Tyr1096 being unique to RET\textsubscript{51}. Among these tyrosines, 14 have been identified as auto-phosphorylation sites and are spread out across the ICD: Tyr687 is located in the JM region, Tyr752, 806, 809, 826, 900, 905, 928 and 981 are in the kinase domain and Tyr1015, 1029, 1062, 1090, 1096 in the C-terminal tail (71–73). The dynamic phosphorylation and dephosphorylation of these tyrosine residues regulate the activity and binding interactions of RET\textsuperscript{ICD}. Two tyrosine residues, Tyr900 and Tyr905, located in the AL, were identified as late auto-phosphorylation sites and in comparison, Tyr687 in the JM region and Tyr1062 in the C-terminal tail showed early auto-phosphorylation (74). Common for protein kinases, phosphorylation of the AL tyrosines is crucial for maximising kinase activity. In the case of RET, its kinase activity indeed shows at least a 3-fold increase \textit{in vitro}, mediated \textit{via} Tyr905 phosphorylation (73,74). To date, only the active (or close to active) conformation (\textit{i.e.} DFG-in) of RET\textsuperscript{ICD} has been identified based on the crystal structures of the TK domain in both the AL-non-phosphorylated (73) and phosphorylated (73–75) forms (Figure 6B). Aligning these RET\textsuperscript{TK} structures reveals a high-level of structural conformity, with the main conformational changes being observed at the GRL (“GRL-open” and “GRL-closed” states), suggesting the flexibility of the GRL region. In the “GRL-closed” state, the GRL functions as a tether, stabilizing the AL in an inactive conformation.
Further studies have shown that the JM region of RET\textsuperscript{ICD} is important in regulating RET\textsuperscript{TK} activity and suggested that this region may function to suppress kinase activity \cite{25,76}, which can be reversed through Tyr687 phosphorylation. Crystal structures of RET\textsuperscript{TK} in the presence of the JM region, reported by Plaza-Menacho \textit{et al.} \cite{25}, show a rearrangement of the AL that is induced by serine phosphorylation at Ser909. Phospho-Ser909 displaces phospho-Tyr905 from its salt-bridged interaction with R770 in the $\alpha$C helix (Figure 6B, right panel). Interestingly, they demonstrated that phosphorylation of Ser909 is a late autophosphorylation event, indicating that RET is capable of serine phosphorylation. Overall, it is not yet clear what the biological significance of these different kinase conformations are or what role the C-terminal region plays in regulating RET\textsuperscript{TK} activity. A comprehensive understanding of the roles of individual tyrosine phosphorylation events in RET is still lacking.
Following the activation of RET\textsuperscript{TK}, the phosphorylated tyrosines function as docking sites for different cytosolic effector proteins to transmit the downstream signalling cascade, either by serving as substrates of RET phosphorylation or as adaptor proteins (Figure 5). Phospho-Tyr\textsuperscript{905} is the binding site for the growth factor receptor-bound protein 7 (Grb7) and Grb10, which activates the downstream Rat sarcoma proteins (RAS)/mitogen activated protein kinase (MAPK) pathways. Phospho-Tyr\textsuperscript{752} and phosphor-Tyr\textsuperscript{928} are the docking sites for the signal transducer and activator of transcription 3 (STAT3) for signal transduction \emph{via} the Janus kinase (JAK)-STAT signalling pathway. Multiple adaptor proteins, including the insulin receptor substrate (IRS) and the Shc adaptor protein, bind to phospho-Tyr\textsuperscript{1062} and recruit other signalling molecules to activate several signalling pathways, including the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) and RAS/MAPK (77,78) pathways. Activation of these signalling cascades affects gene expression and mediates a variety of cellular processes, including cell maintenance, differentiation, proliferation and migration (79,80).

1.2.4 The biological function of RET and its complexes

The expression of GFLs and their co-receptors is tissue-dependent, which allows simultaneous regulation of divergent RET signalling pathways (37,81–85). Upon GFL stimulation through their co-receptors, RET receptor tyrosine kinase activates and its signalling plays an important role in the development of the nervous system and kidneys, and mediates energy intake and expenditure. GDNF is a central player for midbrain dopaminergic neuron (86–88) and motor neuron survival (65,89,90), as well as the development of the enteric nervous systems (91–94), the parasympathetic neurons (95,96) and kidney (97–99). Apart from GDNF, NRTN and PSPN also play an important role in dopaminergic neuronal survival (84,100,101). ARTN and NRTN promote the parasympathetic and enteric neuronal development (102,103), but their effect is less significant compared to GDNF (104). In addition, it also appears that ARTN is responsible for controlling nociceptor sensitivity (105), sensory (106) and sympathetic neuron development and maintenance (107,108). GDF15/GFRAL was identified as a regulator of food intake and energy expenditure under emergency feeding pathways (45,47–49). Previous studies have shown that treating rodents and monkeys with recombinant GDF15 resulted in reduced food intake and weight loss (109) and similar results were also seen in GDF15 deficient mice, where mice fed on normal diet gain more weight than the control (110). Moreover, the serum concentration of GDF15 is elevated under stress conditions (111,112) and may rise up to 200-fold in patients with advanced cancer, which correlates with reduced food intake and weight loss (113). In addition to the energy control, GDF15 may also function to protect dopaminergic neurons (114,115).
1.3 RET dysregulation

1.3.1 Loss-of-function

The loss-of-function mutations of RET are associated with the majority of Hirschsprung’s disease (HSCR) cases. There are three main mechanisms accounting for the loss of function: 1) RET truncation at the C-terminal region leading to impaired kinase activity; 2) incomplete post-translational modification (glycosylation), immature RET production and failure to localize on the plasma membrane; and 3) mutation-induced impairment of kinase activity (Table 2). Although a large number of RET gene variants in HSCR have been identified with predicted functions, the biological function of some of them is yet to be verified (23,116–119). Apart from HSCR, the loss of RET function has also been shown to disrupt parkin and GDNF-mediated neuron maintenance and regeneration, which is implicated in the Parkinson’s disease (PD) (88,120). The roles that RET plays in various diseases highlight its therapeutic importance.

1.3.2 Gain-of-function

1.3.2.1 Fusion: chimeric RET oncoproteins

The chromosomal rearrangement of the RET gene accounts for a special group of RET fusion proteins, where a 5’ gene of another protein is fused with the 3’ RET kinase gene. If the 5’ gene encodes a protein dimerization domain, such apposition will lead to constitutively activated dimeric RET kinase domains, resulting in the oncogenic fusion proteins. Since the discovery of the first fusion RET oncoprotein, called the coiled-coil domain containing 6 gene (CCDC6)-RET, 18 different RET-fusion proteins have been identified. Overall, RET-fusion proteins have an occurrence of 20 to 40% in papillary thyroid carcinoma (PTC), 1 to 2% in non-small-cell lung cancer (NSCLC) (80,121–125) and less than 0.5% in other cancers including breast cancer (126) (Table 2).

1.3.2.2 Over-expression

Another oncogenic mechanism is over-expression-induced aberrant RET activation. Increased RET expression or transcription has been reported in the cases of pancreatic cancer (127), acute myeloid leukemia (128), breast cancer (129) and a subtype of NSCLC (130), as a result of gene amplification, impairment of regulatory mechanisms and transcription factor upregulation (129,131).

1.3.2.3 Gain-of-function point mutations

Gain-of-function point mutations of RET lead to the constitutive activation of the receptor without the binding of its ligands. Such mutations are known to be the primary inherited
oncogenic drivers of multiple endocrine neoplasia type 2 (MEN2) syndrome, with a strong correlation between the disease phenotype and the specific mutations (80,132). MEN2 is a germline multi-tumour syndrome, primarily characterized by medullary thyroid cancer (MTC), parathyroid tumours, and pheochromocytoma. Two subtypes exist in MEN2, namely MEN2A and MEN2B and the occurrence of MEN2A is around 19 times higher than that of MEN2B (133). MEN2A is further classified into four clinical categories, classical MEN2A, MEN2A with cutaneous lichen amyloidosis (CLA), MEN2A with HSCR and familial medullary thyroid cancer (FMTC).

Point mutations in the ICD of RET lead either to FMTC/MEN2A, the least aggressive subtype, or to MEN2B, the most severe and aggressive phenotype. The well characterized M918T and A883F mutations exclusively lead to MEN2B by activating both RET monomer and dimers, of which M918T accounts for the majority of cases (134–136). M918 is located between the substrate binding pocket and the activation loop of the TK domain of RET and the M918T mutation disrupts the tether between the GRL and AL, which results in a conformational change of the AL, favouring ATP binding. Unlike M918, A883 sits next to the activation loop and mutation at this residue is likely to function by destabilizing the inactive form of the kinase domain and may be less aggressive (137). Other point mutations in the ICD, including E768, L790, Y791, V804 and S891, in addition to the mutation at G533 in the ECD, result mainly in FMTC by different mechanisms (Table 2) (73,80,132,138–142). Moreover, although some single mutations exhibit a relatively mild phenotype, tandem mutations, such as V804/Q781, V804/E805, V804/Y806 and V804/S904, contribute to increased kinase activity and more aggressive MEN2B-like phenotype (143,144).

As described earlier, the cysteine residues in the CRD of RET all form intramolecular disulfide bonds; therefore, the mutation of one of the cysteine residues to any other residue will leave the other cysteine unpaired, which leads to RET homo-dimerization via intermolecular disulfide bond formation. Examples of these CRD point mutations, including C609, C611, C618, C620, C630 and C634, have been identified as the major cause of MEN2A (85–90%). The C609, C611, C618, C620 and C630 mutations are considered moderately activating, because they impair the expression and maturation of RET and hence limited the level of RET activation (23,145,146). Additionally, mutations at C609, C611, C618 and C620, also known as Janus mutations, may also induce a loss-of-function phenotype of RET as a result of the impaired expression, maturation and plasma membrane localization, characterized by MEN2A with HSCR (136,147,148).

Among all the cysteine mutations, the C634 mutation is considered the most strongly activating and clinically aggressive, leading to a higher risk of parathyroid tumours, and pheochromocytoma (136,149,150). In vitro studies have showed that ligand-independent dimerization of full-length RET is induced by C634R mutation and that the
phosphorylation level of RET(C634R) is considerably elevated relative to its wild-type counterpart (151,152). However, previous studies have failed to express the extracellular domain of RET(C634R) mutant in its dimeric form, raising questions over the mechanism of action. It was unclear why RET(C634R) ECD did not form a dimer when recombinantly expressed (21,153), while the dimeric full-length RET(C634R) mutant could be readily detected, and this represents an interesting topic for investigation.

**Table 2. Mechanisms of RET dysregulation in diseases.** ECD: extracellular domain; ICD: intracellular domain; FMTC: familial medullary thyroid cancer; MEN2: multiple endocrine neoplasia type 2; PTC: papillary thyroid carcinoma; NSCLC: non-small-cell lung cancer; CMML: chronic myelomonocytic leukemia; HSCR: Hirschsprung’s disease; TCA: total colonic aganglionosis.

<table>
<thead>
<tr>
<th>Location</th>
<th>Mutation/Fusion</th>
<th>Effect</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain-of-function mutations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECD</td>
<td>G533C</td>
<td>Moderately activating.</td>
<td>FMTC</td>
</tr>
<tr>
<td></td>
<td>C609</td>
<td>RET homodimerization via intermolecular disulfide bond. Moderately activating.</td>
<td>MEN2A</td>
</tr>
<tr>
<td></td>
<td>C611</td>
<td>RET homodimerization via intermolecular disulfide bond. Moderately activating.</td>
<td>MEN2A</td>
</tr>
<tr>
<td></td>
<td>C618</td>
<td>RET homodimerization via intermolecular disulfide bond. Strongly activating.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C620</td>
<td>RET homodimerization via intermolecular disulfide bond. Strongly activating.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C630</td>
<td>RET homodimerization via intermolecular disulfide bond. Strongly activating.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C634</td>
<td>RET homodimerization via intermolecular disulfide bond. Strongly activating.</td>
<td></td>
</tr>
<tr>
<td>ICD</td>
<td>G691</td>
<td>Possibly increases kinase activity.</td>
<td>FMTC, MEN2A</td>
</tr>
<tr>
<td></td>
<td>E768</td>
<td>Increases kinase activity. Moderately activating.</td>
<td>MEN2B</td>
</tr>
<tr>
<td></td>
<td>L790</td>
<td>Increases kinase activity and ATP binding.</td>
<td>FMTC, MEN2A</td>
</tr>
<tr>
<td></td>
<td>Y791</td>
<td>RET monomer phosphorylation. Moderately activating.</td>
<td>FMTC</td>
</tr>
<tr>
<td></td>
<td>V804</td>
<td>Increases ATP binding. Moderately activating.</td>
<td>FMTC</td>
</tr>
<tr>
<td></td>
<td>A883</td>
<td>Destabilizes the inactive conformation.</td>
<td>MEN2B</td>
</tr>
<tr>
<td></td>
<td>S891</td>
<td>Conformational change and RET monomer phosphorylation.</td>
<td>FMTC, MEN2A</td>
</tr>
<tr>
<td></td>
<td>M918</td>
<td>Increases kinase activity, ATP binding and stabilize the active conformation.</td>
<td>MEN2B</td>
</tr>
<tr>
<td></td>
<td>V804, Q781</td>
<td>Conformational change and increases kinase activity.</td>
<td>MEN2B-like</td>
</tr>
<tr>
<td></td>
<td>V804, E805</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V804, Y806</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V804, S904</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RET fusion</strong></td>
<td>ICD</td>
<td>CCDC6(H4)-RET</td>
<td>Increases RET kinase expression and dimerization.</td>
</tr>
<tr>
<td></td>
<td>PRKAR1A-RET</td>
<td></td>
<td>PTC</td>
</tr>
<tr>
<td></td>
<td>NCOA4(ELE1)-RET</td>
<td></td>
<td>PTC, NSCLC, breast cancer</td>
</tr>
<tr>
<td></td>
<td>(RET/PTC3)</td>
<td></td>
<td>PTC</td>
</tr>
<tr>
<td></td>
<td>NCOA4(ELE1)-RET*</td>
<td></td>
<td>PTC, NSCLC</td>
</tr>
<tr>
<td></td>
<td>GOLGAS(RFG5)-RET</td>
<td></td>
<td>PTC</td>
</tr>
<tr>
<td></td>
<td>TRIM24-RET</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRIM33-RET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Tumor Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KTN1-RET</td>
<td>PTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFG9-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERC1(ELKS)-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCM1-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFP(TRIM27)-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOOK3-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIF5B-RET</td>
<td>NSCLC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUFY2-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUX1-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA1468-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLIP1-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYOSC-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPHA5-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PICALM-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRMD4A-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIF13A-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAC-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RELCH-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCR-RET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR1OP-RET</td>
<td>CMML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RasGEF1A-RET</td>
<td>Breast cancer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loss-of-function mutations</th>
<th>RET/PTC4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECD</strong></td>
<td></td>
</tr>
<tr>
<td>G93</td>
<td>Largely impairs RET maturation.</td>
</tr>
<tr>
<td>V145</td>
<td>HSCR (TCA)</td>
</tr>
<tr>
<td>S32</td>
<td>Impairs RET maturation.</td>
</tr>
<tr>
<td>L56</td>
<td>HSCR (less severe phenotypes)</td>
</tr>
<tr>
<td>R76L (insertion)</td>
<td></td>
</tr>
<tr>
<td>R77</td>
<td></td>
</tr>
<tr>
<td>W85</td>
<td>Insufficient glycosylation and impaired RET maturation.</td>
</tr>
<tr>
<td>S100</td>
<td>C-terminal truncation and loss of kinase activity</td>
</tr>
<tr>
<td>R114</td>
<td>Impairs RET maturation.</td>
</tr>
<tr>
<td>F174</td>
<td></td>
</tr>
<tr>
<td>R180</td>
<td></td>
</tr>
<tr>
<td>E252</td>
<td>C-terminal truncation and loss of kinase activity</td>
</tr>
<tr>
<td>Y263</td>
<td>C-terminal truncation and loss of kinase activity</td>
</tr>
<tr>
<td>P270</td>
<td>Impairs RET maturation.</td>
</tr>
<tr>
<td>F393</td>
<td></td>
</tr>
<tr>
<td>R475</td>
<td></td>
</tr>
<tr>
<td><strong>ICD</strong></td>
<td></td>
</tr>
<tr>
<td>E762</td>
<td>Partially impaired kinase activity.</td>
</tr>
<tr>
<td>S765</td>
<td>Loss of kinase activity</td>
</tr>
<tr>
<td>S767</td>
<td>Partially impaired kinase activity.</td>
</tr>
<tr>
<td>R770</td>
<td>C-terminal truncation and loss of kinase activity</td>
</tr>
<tr>
<td>Q860</td>
<td></td>
</tr>
<tr>
<td>R873</td>
<td>Loss of kinase activity</td>
</tr>
<tr>
<td>F893</td>
<td></td>
</tr>
<tr>
<td>R897</td>
<td></td>
</tr>
<tr>
<td>E921</td>
<td></td>
</tr>
<tr>
<td>R972</td>
<td>Partially impaired kinase activity.</td>
</tr>
<tr>
<td>P973</td>
<td>Decrease RET expression.</td>
</tr>
<tr>
<td>M980</td>
<td>Partially impaired kinase activity.</td>
</tr>
</tbody>
</table>

*R* RET/PTC4, a variant of NCOA4(ELE1)-RET fusion which is associate with both RET/PTC3 and RET/PTC4.
2 Therapies targeting RET and its complexes

RET plays essential roles in a variety of cellular processes and its malfunction is implicated in different diseases as described earlier, such as PTC, NSCLC and MEN2. As such, it represents an important therapeutic target for treating these diseases. As a receptor tyrosine kinase, modulators targeting RET can be characterized into two categories: intracellular tyrosine-kinase modulators (TKM) and extracellular domain modulators (EDM).

2.1 Kinase domain targeting

Historically, protein kinases were viewed as undruggable because it was assumed that selective inhibition would be impossible given the high level of structural homology in the ATP binding site for a family of 518 proteins (154). However, the FDA approval of Imatinib in 2001, which was the first selective protein kinase inhibitor to be approved, marked the beginning of a highly successful era of kinase drug discovery (155). There are currently 52 protein kinase inhibitors in clinical use, of which 28 target RTKs, and these have had a profound impact on our ability to treat a variety of diseases, most notably cancers (156). Most kinase inhibitors are ATP-competitive and typically contain an aromatic hinge binding motif that hydrogen bonds to the hinge region (spanning residues 805-807 for RET), mimicking the adenine core of ATP. Kinase inhibitors are typically classified as either Type I, which bind to the active conformation of a kinase with the conserved DFG motif facing in, or Type II, which target the inactive conformation with the DFG facing outwards (157). However, in the case of RET this nomenclature is less applicable because it has never been observed in a DFG out conformation (apart from that shown in homology modelling (158)) but instead its activity appears to be regulated by a conformational shift in the GRL and the juxtamembrane region (Figure 6B). The tyrosine kinase domain of RET has high sequence similarity with the other TKs and is especially structurally conserved at the ATP-binding pocket. As a result, early examples of RET\textsuperscript{T} inhibitors, which targeted the GRL closed conformation, were only partially RET selective. For example, vandetanib was developed in 2002 as a dual vascular endothelial growth factor receptor 2 (VEGFR2) and EGFR inhibitor but was withdrawn from clinical trials to treat advanced NSCLC after it failed to show an improvement in patient survival (159,160). However, vandetanib was later found to also potently inhibit RET\textsuperscript{T} (161) and was ultimately approved as a treatment for MTC in 2011 under orphan drug designation (162,163). Overall, multi-tyrosine kinase inhibitors have been surprisingly successful in treating RET associated diseases, with vandetanib, cabozantinib, lenvatinib and sorafenib already approved to treat thyroid cancers and other examples progressing in late stage clinical trials (Table 3).
Table 3. Inhibitors that interact with the tyrosine kinase domain of RET. VEGFR: vascular endothelial growth factor receptor; MET: hepatocyte growth factor receptor; KIT: stem cell growth factor receptor Kit; AXL: tyrosine-protein kinase receptor UFO; FGFR: fibroblast growth factor receptor; PDGFR: platelet-derived growth factor receptor; FLT: FMS-related tyrosine kinase; EGFR: epidermal growth factor receptor; ALK: anaplastic lymphoma kinase; IAK: Janus kinase; IRAK: interleukin-1 receptor-associated kinase; CSF-1R: colony stimulating factor 1 receptor; DDR: discoidin domain receptor tyrosine kinase; PI3K: phosphoinositide 3-kinase; CDK: Cyclin-dependent kinase; ARK: AMPK-related protein kinase; FDA, the U.S. Food and Drug Administration; TC: thyroid cancer; NSCLC: non-small-cell lung cancer; MTC: medullary thyroid cancer; DTC: differentiated thyroid cancer; IBS: irritable bowel syndrome.

<table>
<thead>
<tr>
<th>Inhibitor (reference)</th>
<th>Target protein</th>
<th>FDA</th>
<th>FDA approval for RET associated disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabozantinib (XL184) (164)</td>
<td>VEGFR2, MET, RET, KIT, AXL</td>
<td>Y</td>
<td>Y (TC) Phase II (NSCLC)</td>
</tr>
<tr>
<td>Lenvatinib (E7080) (165)</td>
<td>VEGFR1-3, FGFR1-4, PDGFR, KIT, RET</td>
<td>Y</td>
<td>Y (TC)</td>
</tr>
<tr>
<td>Ponatinib (AP24534) (166,167)</td>
<td>ABL, PDGFRα, VEGFR1,2, FGFR1, SRC, RET(WT/V804M), KIT</td>
<td>Y</td>
<td>Phase II (MTC)</td>
</tr>
<tr>
<td>Sorafenib (BAY43-9006) (168)</td>
<td>RAF1, BRAF, VEGFR1-3, PDGFRβ, FLT3, KIT, RET</td>
<td>Y</td>
<td>Y (DTC)</td>
</tr>
<tr>
<td>Sunitinib (SU11248) (169)</td>
<td>VEGFR2, PDGFRβ, KIT, IRE1α, FLT3, RET</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Vandetanib (ZD6474) (170)</td>
<td>VEGFR2/3, EGFR, RET</td>
<td>Y</td>
<td>Y (TC)</td>
</tr>
<tr>
<td>Nintedanib (BIBF1120) (171)</td>
<td>VEGFR1-3, FGFR1-4, PDGFRα/β, SRC, FLT3, RET</td>
<td>Y</td>
<td>Phase II (MTC, DTC)</td>
</tr>
<tr>
<td>Alectinib (CH5424802) (172)</td>
<td>ALK(WT/L1196M), RET</td>
<td>Y</td>
<td>Phase II (NSCLC)</td>
</tr>
<tr>
<td>Regorafenib (BAY73-4506) (173)</td>
<td>VEGFR1-3, PDGFRβ, KIT, RET, BRAF, RAF1</td>
<td>Y</td>
<td>Phase II (MTC)</td>
</tr>
<tr>
<td>Fedratinib (TG101348) (174)</td>
<td>JAK2(V617F/WT), FLT3, RET</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Danusertib (PHA-739358) (175)</td>
<td>Aurora A, ABL, TrkA, RET, FGFR1</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Agerafenib (RXDX-105) (176)</td>
<td>BRAF(V600E/WT), RAF, ABL1, KIT, RET, PDGFRβ, VEGFR2</td>
<td>N</td>
<td>Phase I (NSCLC)</td>
</tr>
<tr>
<td>Dovitinib (TK1258) (177)</td>
<td>FLT3, KIT, FGFR1-3, PDGFRα/β, VEGFR1,2, Trk, CSF-1R, RET</td>
<td>N</td>
<td>Phase II (tumours with mutations or translocation)</td>
</tr>
<tr>
<td>Apatinib (YN968D1) (178)</td>
<td>VEGFR2, PDGFRβ, RET, Sre</td>
<td>Y</td>
<td>Phase II (NSCLC)</td>
</tr>
<tr>
<td>Luminespib (AUY922) (179)</td>
<td>EGFR, BRAF, ALK, ROS1, RET</td>
<td>N</td>
<td>Phase II (NSCLC)</td>
</tr>
<tr>
<td>Sitravatinib (MGCD516) (180)</td>
<td>VEGFR2, MET, AXL, RET, Trk, DDR, PDGFRα, KIT</td>
<td>N</td>
<td>Phase I (NSCLC)</td>
</tr>
<tr>
<td>AST487 (181)</td>
<td>FLT3, RET, VEGFR2, KIT, ABL</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2-D08 (182)</td>
<td>AXL, IRAK4, ROS1, MLK4, GSKβ, RET, VEGFR2, PI3Kα</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>BAW2881 (183)</td>
<td>VEGFR1-3, PDGFRβ, KIT, TIE2, RET</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erdfatinib (JNJ-42756493) (184)</td>
<td>FGFR1-4, RET, CSF-1R, PDGFRα/β, FLT4, KIT, VEGFR2</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>ON123300 (185)</td>
<td>CDK4, ARK5, PDGFRβ, FGFR1, RET, FYN</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>CS-2660 (JNJ-38158471) (186)</td>
<td>VEGFR2, RET, KIT</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>WHI-P180 (187)</td>
<td>RET, VEGFR2</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>GSK3179106 (188)</td>
<td>RET</td>
<td>N</td>
<td>Phase I (IBS)</td>
</tr>
</tbody>
</table>
More recently, inhibitors that specifically target RET have also been identified, including selpercatinib (191) and pralsetinib (190). These RET-specific inhibitors show exceptional selectivity and potency for RET, its oncogenic mutants and fusion proteins compared to the non-selective inhibitors (Figure 7) (192–194). Very recently, high resolution crystal structures of selpercatinib and pralsetinib in complex with RET\(^{TK}\) were reported (195). Comparing these two structures with the vandetanib-RET complex reveals the difference in binding mode which accounts for the increased RET selectivity (Figure 7): Vandetanib solely binds at the primary adenine pocket along the hinge region and in a GRL-closed conformation. However, both selpercatinib and pralsetinib extend deeper into the cavity, opening up a cryptic pocket which locks RET in the GRL-open conformation. Notably, a slight outward shift of D892 at the DFG motif is observed in the selpercatinib-RET complex, providing the first example of Type II like behaviour in a RET inhibitor (Figure 7D). From these models, it appears that the RET-selective inhibitors, such as selpercatinib and pralsetinib, bind RET\(^{TK}\) in its GRL-open conformation while the multi-tyrosine kinase inhibitors dock RET\(^{TK}\) in the GRL-closed conformation (Figure 6 and Figure 7). Considering the uniqueness of the conformational change of the GRL of RET\(^{TK}\), it is consistent that other kinases are unable to accommodate the binding of selpercatinib and pralsetinib, resulting in the high level of selectivity observed.

In 2020, FDA approved the use of selpercatinib, followed by pralsetinib, in the treatment of RET-fusion oncogenic protein specific NSCLC and RET mutation-induced MTC. This highlights the tractability and promising future for RET-mutant and/or fusion specific drugs. However, acquired resistance mutations have been recently identified in patients who were treated with these selective inhibitors (192,196) and this is anticipated to become an increasing problem as use of these inhibitors becomes more widespread. The development of next-generation inhibitors targeting the resistance mutants of RET is needed, as well as alternative inhibitor modalities, for example targeting the extracellular domain of RET.

<table>
<thead>
<tr>
<th>Selpercatinib (LOXO-292) (189)</th>
<th>RET(WT/V804M,L/S891A/M918T/A883 F), RET(WT/M918T)</th>
<th>Y</th>
<th>Y (NSCLC, MTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pralsetinib (BLU-667) (190)</td>
<td>RET, BRAF, S6K, SRC</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Figure 7. Inhibitor-binding sites at the tyrosine kinase domain of RET. Lateral view and hydrogen bond contacts in the complex structures between (A) vandetanib (purple, PDB code 2IVU), (B) Pralsetinib (yellow, PDB code 7JU5) and (C) Selpercatinib (grey, PDB code 7JU6) with RET\textsuperscript{TK} in their binding pockets, which are coloured in green, brown and silver respectively. A cryptic pocket (dotted circle) was revealed in the structures of RET\textsuperscript{TK} with Pralsetinib (B) and Selpercatinib (C), but not with vandetanib (A). (D) Enlarged view of the inhibitor-binding site after superimposing the complex structures of vandetanib /RET\textsuperscript{TK} (purple/green, PDB code: 2IVU), pralsetinib/RET\textsuperscript{TK} (yellow/brown, PDB code 7JU5) and Selpercatinib/RET\textsuperscript{TK} (grey/silver, PDB code 7JU6). The structures of Pralsetinib and Selpercatinib extend into the pocket, showing a clear collision with the glycine-rich loop (green) of RET\textsuperscript{TK} in the complex with vandetanib. IC\textsubscript{50} values are based on previously a published study (192) which were measured at physiological ATP concentration.

2.2 Extracellular domain targeting

Since RET binds its ligands and co-receptors extracellularly, targeting the extracellular domains of the RET complexes represents an alternative to the conventional RET\textsuperscript{TK} regulation for the treatment of RET-associated diseases (apart from RET-fusions). Unlike
the tyrosine kinase domain, the extracellular domain of RET shares little sequence similarity to the other receptor tyrosine kinases; therefore, RET\textsuperscript{ECD} regulators are more likely to show very high selectivity for RET. Furthermore, while kinase inhibitors are extremely well characterised, there are very few examples of direct activators of kinase domains. For the treatment of RET deficient diseases like PD and HSCR, activation of the intracellular kinase domain by RET\textsuperscript{ECD} agonism would be highly valuable.

A lot of research has focused on the application of GFLs to up-regulate GFL signalling as a disease treatment. Using Parkinson’s disease as an example, administering GFLs has long been considered to be a potential treatment for dopaminergic neuron maintenance through agonism of RET (197). However, GFLs cannot pass the blood–brain barrier (BBB) and direct administration of GFLs to the brain is required for effective target engagement. Although the results of a Phase II clinical trial, where GDNF was intermittently infused into the brain, showed no significant difference between GDNF and placebo on PD symptoms within the time monitored, all patients treated with GDNF had improvement in their symptoms after 18 months (198,199). The trial showed a promising future for GDNF treatment but concerns still exist on the long-term administration of GFLs, protein quality control, off-target effect and the high price associated with protein production. To overcome those problems, targeted gene therapy has been applied using adeno-associated viral type-2 vector (AAV2) to transfer GFLs to certain locations in the brain for their over-expression. However, in the case of NTRN, initial treatment using AAV2-NRTN led to severe side effects (200) and follow-up studies did not show significance of the treatment in improving PD symptoms (201). Gene therapy techniques are still developing regarding safety and efficiency and the future application of gene therapy to RET agonists may still have promise (202,203).

To down-regulated GFL signalling, the application of neutralizing antibodies targeting GFLs (204) or the GFL co-receptors (205,206) has also gained attention. For instance, treatment with an anti-GFRAL antibody in mice disrupted GDF15/GFRAL/RET signalling and rescued the GDF15-induced cancer anorexia-cachexia (206). Noticeably, antibody therapies targeting RET are currently underdeveloped compared to several of the other RTKs: For example, monoclonal antibodies (mAbs) targeting the EGFR family have received a lot of attention, with four examples currently in clinical use, and other mAbs under evaluation, as treatments for cancers, including NSCLC (207). These EGFR mAbs function through a combination of mechanisms, including disrupting EGF binding, hindering EGFR association with other HER family receptors as well as facilitating EGFR degradation. Although no RET mAb has been reported so far to treat RET-associated diseases, an antibody-drug conjugate (ADC) therapy targeting RET, using an anti-Ret antibody conjugated to a cytotoxic agent, has been under pre-clinical assessment for the treatment of breast cancer and showed promising anti-tumour activity (208). Additionally, antibodies that specifically recognize either monomeric RET\textsuperscript{ECD} and
dimeric RETECD in complex with its ligands were reported in 2020 (206). These strategies highlight the potential of clinical anti-RET antibody therapies for the treatment of RET-dysregulation-associated diseases, which recognize different conformations of RET, including its constitutively activated RETECD mutant dimers.

Apart from the macromolecules, peptides (57) and small molecules are also gaining attention as RETECD modulators. In 2003, XIB4035 was first small molecule identified as a GFRα1 agonist and its application led to RET activation in cell-based assays (209), with the activation being dependent on the presence of GFRα1 (210). More than a decade later, another two classes of agonists targeting RETECD, BT (211,212) and Q (213) compound series, were discovered through high-throughput screening. Another class of agonist with a different scaffold was identified through structure-based in silico screening (214). One of the BT compounds, BT13, showed promising results, maintaining and promoting the growth of dopamine neurons and the secretion of dopamine (215,216). Unlike the GFLs, the small molecules can pass the BBB and didn’t show significant toxicity (211). The development of these compounds is still underway for improved efficacy and potency with reduced toxicity. Compared to the protein-based therapies, the small-molecule modulators targeting the extracellular domain complexes have received less attention and investigation, and no small molecule inhibitors have been identified that target RETECD directly, which is, to a large extent, restricted by the lack of high-resolution structural information to allow structure-based drug design.

3 Structural investigation using electron microscopy

X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy (EM) are the three major techniques for the structural determination of biological macromolecules. The high molecular weight of the RET domains makes them unsuitable for investigation by NMR, which is typically limited to the study of proteins with molecular weight less than 30 kDa (217). Furthermore, the extensive glycosylation and flexibility of the extracellular and transmembrane domains, makes the structural investigation of RET difficult using X-ray crystallography. Whereas there are a wealth of crystal structures of the tyrosine kinase domain of RET, there has been significantly less success in studying RETECD by crystallography, with only two publications of truncated RETECD crystal structures, one of which used a moderately truncated zebrafish construct and the other a highly truncated human construct (23,218). However, in the last decade, great advances have been made in cryo-EM techniques that now enable such challenging protein classes to be structurally studied at high resolution. Late on in my Ph.D. research, when I had already obtained cryo-EM images of the RET/GDF15/GFRAL complex (Study IV), four structures of humanRETECD in complex with its ligands were solved at
3.4-4.4 Å resolution using cryo-EM (18,58). This study highlights the impact cryo-EM is having on the structural biology of challenging protein classes and in particular, that cryo-EM is poised to elucidate additional RET states, many of which are yet to be structurally studied, such as full length RET and its oncogenic variants.

During EM data collection, a focused and accelerated electron beam that is emitted from the top of the microscope interacts with the sample, causing the electrons to scatter. The scattered electrons are focused, magnified and recorded by detectors, which is visualized as a micrograph, a 2-dimentional (2D) projection of the biological sample. Different micrographs represent snapshots of sample particles in particular orientations and the projections can be further processed and reconstituted into a 3-dimentional (3D) model. Because biological samples are deposited on EM grids with random orientations, the more views (or orientations) captured in single-particle EM, the more likely the reconstructed 3D model can accurately represent the real structure of the biological sample. To reduce the scattering of the electrons by molecules other than the specimen, high-vacuum is applied in the microscope; therefore, the samples must be prepared to withstand such environment with minimized radiation damage (219,220). To achieve this, several approaches have been developed for sample preparation, including negative staining and plunge freezing to create vitreous ice for cryo-EM, which are used in this thesis project.

3.1 Electron microscopy sample preparation

Negative stain EM grids typically contain a carbon film on a copper grid support while cryo-EM grids also use gold film and a gold grid support. A variety of grid types have been developed and their suitability is sample-dependent (221–225). Because the surface of EM grids is intrinsically hydrophobic, it is important to make the surface hydrophilic for the application of aqueous samples by either glow-discharging or plasma cleaning (225,226).

3.1.1 Negative stain EM sample preparation

To prepare EM grids using negative staining, the sample is first incubated on the glow-discharged grid and excess solution is blotted, prior to the addition of a droplet of negative staining solution (Figure 8A). After staining, the grid is blotted till dry. Typically uranyl acetate is used as the staining reagent, though uranyl formate, tungsten or molybdate salt solutions may behave better for certain samples (220). Because the heavy stains are electron dense and they coat the background of the grids, excluding the sample particles, more electrons are scattered by the stained area than the particles of interest, resulting in a dark background with bright particles. Compared to cryo-EM, the use of negative stain EM (NS-EM) enables relatively simple and quick analysis of sample size and
homogeneity as well as structural evaluation through data processing. However, there are drawbacks and limitations using this approach. Firstly, although most samples can withstand the coating by negative stains, their integrity may be impaired during the preparation. Next, the stain may flatten or only partially coat the particles, resulting in distorted or incomplete structures. Thirdly, the resolution of the reconstructed models using NS-EM data is generally limited to ~18 Å (221). As the contrast of the images arise from the heavy stain rather than the particles and the stain may penetrate or unevenly coat the particles, high resolution structural information cannot be obtained; therefore, the vitrification of sample is preferred for high-resolution structural investigation (227), which is the basis of cryo-EM grid preparation.

![Sample preparation and data processing procedures for electron microscopy](image)

**Figure 8. Sample preparation and data processing procedures for electron microscopy.** Schematic typical grid preparation of an EM specimen by A) negative staining and B) plunging freezing for cryo-EM. Procedures may vary depending on different samples and approaches used. RT: room temperature. C) The workflow of EM image processing. Adapted from Tegunov and Cramer (228).

### 3.1.2 Cryo-electron microscopy sample preparation

Protein samples for EM studies are in aqueous solutions and vitrification of the sample allows the preservation of sample particles in native conditions while remaining stable in the vacuumed environment. Overall, the grid preparation procedure for cryo-EM has many similarities to that of negative stain EM. However, instead of staining and blotting to dry completely, cryo-EM samples stay hydrated in solution and the sample droplet on the grid is blotted briefly to a thin layer before rapidly plunge-freezing in liquid ethane or propane, which is temperature controlled using liquid nitrogen at approximately −174 °C
Here, liquid nitrogen is not used for the plunge-freezing because the temperature is close to its boiling point and its heat capacity is low; therefore, plunging the grids would cause rapid nitrogen evaporation, risking the formation of cubic ice. Instead, liquid ethane or propane are used, which have freezing points of −182.8 °C or −188 °C respectively and heat capacities that are almost five times higher than liquid nitrogen. The EM specimens are kept at liquid nitrogen temperature at all times after preparation. Rapid freezing of the grids allows the biological samples to be trapped in a thin layer of amorphous ice in a near-native storage environment, preserving their native conformations. Obtaining amorphous ice is important to obtain micrographs of good quality as ice crystals damage the EM sample and diffract electrons (221,227,229). Unlike NS-EM, which uses heavy atoms to improve the signal with low resolution, cryo-EM relies on electron diffraction by the sample particles, of which the signal to noise ratio is significantly weaker. Other sample preparation methods have been developed such as cryo-negative stain EM, where the samples are negative-stained, while being maintained in a hydrated and frozen state (230–232). However, non-stained cryo-EM is still necessary to obtain high resolution information due to the limitation of the stained samples as discussed above.

There are several factors to consider when optimizing cryo-EM sample preparation for high-resolution reconstruction, including buffer composition, sample purity, stability and concentration, EM grid type and ice thickness (220–222,233). Reagents such as glycerol and detergent can reduce image contrast; therefore, the concentration of those reagents should be reduced to minimum. Next, the protein concentration and grid types should be determined empirically for each sample to obtain a good particle distribution, where there are sufficient particles per micrograph while each particle is well separated to allow efficient particle picking and subsequent reconstruction. When the protein samples are unstable, various approaches have been successfully applied to stabilize heterogenous proteins and complexes, such as chemical cross-linking, including a gradient fixation method termed “GraFix” (234,235) as well as the addition of stabilizing modulators (236). Additionally, the choices of grid type, blotting parameters and methods should also be determined on a case-by-case manner, which directly affects the ice thickness (224) and indirectly influences the air-water interface effect. Generally, the ice thickness should be thick enough to just cover a monolayer of particles (237,238). If the ice layer is too thin, it may affect sample integrity and too thick ice layers lead to reduced contrast, uneven defocus levels or overlapping particles (239). Another factor that greatly interferes with particle distribution and quality is the air-water interface. The volume of the sample applied to a cryo-EM grid is relatively small (3-5 µl) and the proteins that are exposed to the hydrophobic air are prone to denature. During the blotting step in specimen preparation, the volume decreases while the surface to volume ratio increases, which leads to the absorption of samples to the air-water interface, exacerbating sample
denaturation and aggregation (240). Although many proteins retain their integrity at the air-water interface, protein particles are likely to adopt preferred orientations on the grid as a result of the effect, ultimately hindering reconstruction (239). Such effect is not as big a concern in NS-EM grid preparation, thanks to the use of grids with a continuous carbon support, which supports the absorption and even distribution of the sample on to the grid. However, the continuous carbon support is not favoured in cryo-EM as it worsens the signal-to-noise ratio (SNR) and is susceptible to movement caused by radiation damage (241). To overcome the air-water interface effect, multiple approaches have been developed and have shown success in various cases, including the addition of detergent at low concentrations, such as nonyl phenoxypolyethoxylethanol (NP-40) (242), lauryl maltose neopentyl glycol (LMNG) (243) or CHAPSO (244), the addition of stabilizing modulators (243), cross-linking (234,245), the usage of continuous or graphene oxide support film (241,246,247) and decreasing the time between sample application and plunge-freezing (239,240,248–250). Although the above-mentioned methods have shown success in some cases, there is no universal method that suits all and the conditions must be optimized for individual proteins empirically.

3.2 Electron microscopy data acquisition and processing

Following successful grid preparation, micrographs can then be collected on an electron microscope for sample evaluation and reconstruction (Figure 8C). For image acquisition, although a higher electron dose is positively related to the contrast, it negatively effects the radiation damage and specimen drifting; therefore, the electron dose to use is decided based on the sample properties, data collection modes and camera types (222). Typically, a total electron dose of 50 – 60 e/Å² is split to 40 – 50 image frames during imaging in movie mode on the direct electron detector devices (DDD) and the dose-fractioned frames are stacked into a series of movies. The alignment of the movie frames allows the correction of beam-induced movement. After motion correction and movie averaging, the micrographs are subject to contrast transfer function (CTF) estimation and correction. As described previously, the contrast of an EM micrographs mainly arises from the interference of the scattered and directly transmitted electrons. Because the biological samples used for cryo-EM, such as viruses and proteins, mainly consist of light elements (C, N, O and H) that diffract electrons weakly and are in common with the buffer components, the SNR of the images is low when collected using a low electron dose. There are several approaches to increase the SNR, such as the application of phase plate (251,252) and energy filter (220) and the most commonly used one is by defocusing, which effectively enhances the contract by introducing phase shifts. For thin EM specimens, the image formation is described by an oscillating contrast transfer function (CTF) and the determination of CTF relies on multiple parameters defined by a given microscope set up as well as defocus values and astigmatism of the individual
micrographs and can be achieved by comparing the model CTF and observed CTF oscillations (Thon rings) and adjusting the parameters. Because the CTF oscillates, it has zeros where the information is lost; therefore, the micrographs should be taken at different defoci for CTF correction to fully restore information of a undistorted object (220,222,253). Additionally, it is worth noting that while increasing the defocus value gives a rise in contrast, the high-resolution information is lost. Hence, the defocus range should be limited to a certain range (i.e. \(-0.8 \) to \(-3.0 \ \mu \text{m}\)) so that both high- and low-resolution information is obtained.

Normally, particle picking can begin once the micrographs are motion corrected and CTF corrected. The process can be done manually or in a semi-automated or automated fashion. Manual particle picking can be useful at an early stage of reconstruction when little structural information is known about the object of interest. However, it is labour-intensive and is prone to user bias. The effectiveness of particle picking relies on the quality of micrographs as well as the algorithms when using software for auto-picking (254,255). No matter in which manner the particle picking is performed, special attention should be drawn to minimize user or template induced bias (256). More recently, the integration of deep learning into cryo-EM data processing has gained remarkable popularity and such implementation has significantly improved the reliability of automated particle picking (257–260), enabling on-the-fly analysis in an automated data processing pipeline (261). Particle picking identifies and records the coordinates of the particles, which are then extracted for alignment and 2D classification (262,263). The 2D class averages are indicative of data quality and can be used as a data clearing step to remove bad particles. Subsequently, particle stacks of the selected 2D class averages are grouped for initial model building based on the projection-slice theorem (reviewed in (264)), followed by 3D classification of the selected group of particles typically using an initial model as a reference map. This is then followed by 3D class analysis, before finally proceeding to structure refinement and post-processing (Figure 8C). The resolution of the final map is determined based on the Fourier Shell Correlation (FSC) curve, which is essentially a correlation between two half maps computed based on two theoretically independent sets of data from the complete dataset used for the final reconstruction (265). To determine the overall resolution of an EM map, a cut-off value of the FSC curve is chosen based on different criteria (222,266), with the current most widely accepted one being the “gold-standard” where a cut-off of FSC curve at 0.143 is used (267). Furthermore, as EM is commonly used for structural investigation of large biological molecules and complex systems, different parts or regions of these objects are likely to exhibit different flexibility and stability; therefore, the resulting reconstructed EM map often shows different local resolution with the rigid and stable regions at higher resolution than those flexible and disordered parts. This necessitates the calculation of local resolution, which will in turn help us better understand and evaluate EM maps (268,269).
4 Studies of full-length RET and other RTKs

Similar to the studies of other RTKs, numerous structural investigations have been done on the truncated versions of RET, including the CLDs, the extracellular domain and intracellular kinase domain. However, in order to fully understand the activation mechanism of RET and other RTKs, i.e. how the extracellular-domain binding affects the conformation of the transmembrane and intracellular domains, structural investigation of RTKs in their full-length form is essential.

4.1 Membrane protein expression, solubilization and purification

Despite the abundance of membrane proteins in the proteomes of all living organisms (270,271), only 2174 membrane protein structures have been solved (https://blanco.biomol.uci.edu/mpstruc/, accessed on 19th December, 2020) and this represents only around 1.2% of all the deposited structures in the Protein Data Bank resb.org (272). The disproportionately low number of membrane protein structures can be attributed to multiple levels of difficulties associated with membrane protein research: their low expression level and purification yield, low solubility and stability. Therefore, careful selection and optimization of experimental conditions, such as expression hosts, solubilization and purification methods, are typically required to obtain sufficient purified product for structural investigation (273–275).

Although Escherichia coli remains the most popular expression system, insect and mammalian expression systems are gaining more and more attention, especially for the expression of eukaryotic membrane proteins (276). As different biosynthesis machinery and post-translational modifications are utilized in different expression hosts, the choice of the optimal system for each protein must be empirically determined. Typically, insect and mammalian cells require more specific culturing conditions and are more expensive than bacteria and yeast cells; therefore, the choice of expression systems is also a balance of cost-effectiveness and the expression yield of functional proteins. Following protein expression, the membrane protein of interest needs to be isolated from the membrane through a solubilization step. Among all the solubilization methods that are available, the use of detergents is the most popular (277,278). Detergents are small amphiphilic molecules, which contain a hydrophilic head and a hydrophobic tail. They are classified as ionic, nonionic or zwitterionic, based on their hydrophilic group composition. The amphiphilic nature of detergents enables them to disrupt lipid bilayers and extract membrane proteins from their native environment, with the hydrophobic tails covering the hydrophobic regions of membrane proteins (i.e. transmembrane domain) and the hydrophilic heads pointing towards the aqueous environment (Figure 9).
Figure 9. Solubilization of membrane proteins. Detergent micelles are formed at a concentration above the critical micelle concentration, with the hydrophobic tails pointing towards the hydrophobic patches of the membrane protein and the hydrophilic heads pointing outwards. Bicelles are formed with long-chain lipids and a short-chain detergent rim. Liposomes are spherical lipid vesicles which membrane proteins can be reconstituted into, mimicking native membranes. Amphipols are polymers that are usually used on detergent-solubilized proteins to replace the detergent micelles. They associate with membrane proteins with high affinity and tightly wrap the transmembrane domain. Nanodisc solubilize membrane proteins by embedding into the lipid bilayers with two circular belts of helical membrane scaffold proteins (MSPs). The size of nanodisc is dependent on the size and ratio of MSPs and lipids used. Styrene maleic acid (SMA) and diisobutylene-maleic acid (DIBMA) co-polymers as well as the Saposin-lipoprotein nanoparticles (SapNPs) extract the membrane proteins into disc-like structures. Unlike the SMA/DIBMA co-polymers that isolate the membrane proteins with native lipids, SapNPs reconstitute the proteins (typically pre-solubilized by detergents or liposomes) with supplemented lipids and their size can be adjusted according the membrane protein targets. Peptidiscs utilize specially engineered bi-helical peptides to bind to the hydrophobic regions of the membrane proteins, forming a “customized” peptide belt.

To solubilize membrane proteins from the lipid bilayer using detergents, the formation of detergent micelles is required, a phenomenon that occurs when the concentration of the detergents reaches the critical micelle concentration (CMC). The CMC value of each detergent varies according to its physical properties and depends on the ionic strength and pH of the aqueous environment (278). Most ionic detergents are too harsh for solubilizing membrane proteins without denaturing the protein. Zwitterionic and nonionic detergents, however, are both suitable for membrane protein solubilization, with nonionic detergents generally being the mildest. A detergent screening step is usually recommended to identify the optimal detergent for a particular membrane protein, which often requires a compromise between extraction efficiency and structural determination suitability (279–283). In cases where a compromise cannot be found, a detergent exchange step can be
implemented. Currently, nonionic detergents, including n-Dodecyl-β-D-maltoside (DDM), LMNG and digitonin, are the most commonly used for the isolation and structural determination of membrane proteins with or without the addition of the cholesterol derivative cholesteryl hemisuccinate (CHS) (284). The inclusion of CHS is favoured for many proteins to improve stability and activity, mimicking the native membrane environment (282,285,286).

Apart from detergents, a variety of tools have been developed for membrane isolation in recent years which have shown a lot of success for structural studies (Figure 9), including bicelles (287), amphipols (288), liposomes (289,290), nanodiscs (291), styrene–maleic acid co-polymers (SMAs) (292,293), disobutylene-maleic acid (DIBMA) (294), lipoprotein nanoparticles (295) and peptidiscs (296,297). Notably, the solubilization and purification steps for membrane protein production often result in high levels of protein loss, as a result of low solubilization efficiency, protein denaturation and aggregation (273). These solubilization methods are typically expensive because of the high cost of many of these solubilizing agents, which are often needed in large quantities to work on a suitable scale for structural investigations of low yielding membrane proteins. Hence, it is of significant importance to optimize effectively the solubilization method and purification protocol for a given membrane protein.

4.2 Structural studies of full-length RTKs

During the last decade, we have witnessed remarkable achievements in the structural determination of membrane proteins, especially of the multi-span transmembrane proteins, such as GPCRs (298) and ion channels (299). However, high-resolution structures of RTKs, which only contain a single-span transmembrane domain, are significantly underrepresented. To date, only a few medium to high-resolution cryo-EM and low-resolution negative stain EM structures of full-length RTKs have been solved (Table 4).

<table>
<thead>
<tr>
<th>Name</th>
<th>Solubilization method</th>
<th>Structural Method</th>
<th>Resolution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>Triton X-100 → DDM</td>
<td>NS-EM</td>
<td>2D classes</td>
<td>(300)</td>
</tr>
<tr>
<td>EGFR/EGF</td>
<td>Triton X-100 → DDM</td>
<td>NS-EM</td>
<td>2D classes</td>
<td>(300,301)</td>
</tr>
<tr>
<td>PDGFRβ/PDGF</td>
<td>LMNG (GraFix)</td>
<td>NS-EM</td>
<td>27 Å</td>
<td>(302)</td>
</tr>
<tr>
<td>KIT/SCF</td>
<td>DDM (GraFix)</td>
<td>NS-EM</td>
<td>25 Å, 28 Å</td>
<td>(303)</td>
</tr>
<tr>
<td>IR</td>
<td>CHAPS → nanodisc</td>
<td>NS-EM</td>
<td>2D classes</td>
<td>(304)</td>
</tr>
<tr>
<td>IR/insulin</td>
<td>CHAPS → nanodisc DDM + CHS</td>
<td>NS-EM</td>
<td>2D classes</td>
<td>(304)</td>
</tr>
<tr>
<td>IGF-1R/IGF</td>
<td>DDM</td>
<td>Cryo-EM</td>
<td>4.3 Å</td>
<td>(306)</td>
</tr>
<tr>
<td>IGF-1R/insulin</td>
<td>DDM → Amphipol A8-35</td>
<td>Cryo-EM</td>
<td>7.7 Å</td>
<td>(307)</td>
</tr>
<tr>
<td></td>
<td>DDM → Amphipol A8-35</td>
<td>Cryo-EM</td>
<td>4.7 Å</td>
<td>(307)</td>
</tr>
</tbody>
</table>
In 2008, Mi et al. (301) reported the first negative stain EM snapshots of full-length EGFR in complex with EGF and, three years later, a more thorough study which included NS-EM snapshots of liganded and unliganded EGFR (300). The NS-EM 2D classes averages revealed conformational changes of EGFR upon ligand binding. Firstly, the conformation of EGFR\textsuperscript{ECD} switched from a tethered-monomer to a EGF-bound symmetric dimer, and the 2D classes of the unliganded and EGF-bound EGFR correlate well with the previously published crystal structures of the auto-inhibited EGFR\textsuperscript{ECD} monomer (308) and the activated EGFR\textsuperscript{ECD}/EGF dimer (309), respectively (Figure 10A). Furthermore, distinct conformations of the EGFR TK domain were observed, showing either an elongated asymmetric dimer or a round symmetric dimer. The asymmetric dimer revealed an active TK conformation, which could be stabilized by active-conformation binding inhibitors (300), while the symmetric dimer represents an inactive TK conformation.

In 2014 and 2015, Opatowsky et al. (303) and Chen et al. (302) reported low resolution NS-EM models of KIT receptor/stem cell factor (SCF) (Figure 10B) and platelet-derived growth factor (PDGFRβ)/PDGF-B (Figure 10C) complexes, respectively. These represent the first 3D models of any full-length RTK to be published and provide significant insight into RTK biology. Both KIT and PDGFRβ belong to the Type III RTK family, also known as the PDGFR family, and they comprise five immunoglobulin-like (Ig-like) domains extracellularly, a TM domain and an intracellular TK domain. There is a high-degree of similarity between the NS-EM models of activated KIT and PDGFRβ dimers. Both receptors showed a “zipper-like” activation mechanism with the domain 2 and domain 3 (D2 and D3) associating with their corresponding ligands and the D4 and D5 being responsible for receptor dimer self-association. In agreement with an X-ray crystallographic structure of apo-KIT (6), 2D NS-EM classes showed that unliganded KIT receptor adopts an elongated monomeric conformation (303). Upon ligand binding, KIT\textsuperscript{D5} undergoes a significant conformational change, bringing the C-term of the KIT\textsuperscript{ECD}s into close proximity (Figure 10B). Similar to that of EGFR, the TK domains of ligand-bound KIT and PDGFRβ exhibit an asymmetric kinase dimer conformation, suggesting the kinase domain asymmetry may also be required for RTK activation. Although high resolution structural studies are still underway, low resolution EM studies as well as the investigation of the extracellular domain complexes (18,58,309–312) have nonetheless provided valuable insights into RTK biology.
Figure 10. Structural insights into RTK activation using NS-EM. A) Representative negative stain EM 2D classes of unliganded (the first on the left) and EGF bound (right) EGFR. Adapted from Mi et al. (300). EGFR\textit{\Delta tail} represents full-length EGFR without a C-terminal unstructured tail. Conformational heterogeneity is shown in the different classes of EGF-bound EGFR. ECD and KD are circled (black and orange, respectively). Crystal structures of the unliganded EGFR monomer (PDB code 1NQL) and EGF-bound EGFR (PDB code 3NJP) are circled, shown beneath the 2D classes. B) Left panel: negative-stain EM 2D classes of apo-KIT and liganded KIT and a density map of KIT receptor dimer with two stem cell factor (SCF) with the ECD circled. Adapted from Opatowsky et al. (303). Right panel: crystal structure of the monomeric KIT\textsuperscript{SCD} (PDB code 2EC8) and an extracellular domain complex of KIT/SCF (PDB code 2E9W). Superimposed D5 of the apo- and liganded LIT is enlarged. C) Negative-stain EM 2D classes and a density map of PDGFRβ/PDGFB complex with the structure of the ECD complex as well as the TM fitted. Adapted from Chen et al. (302). D: domain.

Of all the RTKs, the activation mechanisms of IR and IGF-1R are currently the best elucidated by virtue of the high-resolution cryo-EM studies of the full-length proteins that were reported from 2019-2020 (Figure 11) (305–307). IR and IGF-1R share high sequence and structural similarity with each other but differ from the other RTKs in that
they both exist as disulfide linked homodimers in their apo-forms (9) and remain as dimers when activated. In addition to the dimer assembly in the crystal structure of apo-IR (PDB code 4ZXB) (3), NS-EM 2D class averages of apo-IR revealed that the receptors adopt an auto-inhibitory conformation with the C-terminal regions of the ECD and transmembrane domains of the IR dimer pointing away from each other and thus keeping the TK domains separated (304). In response to ligand binding, the transmembrane and kinase domains are brought to close proximity (Figure 11) through conformational changes, locking IR in a “T”-shaped active conformation (Figure 12A) (305,313,314).

**Figure 11. Structural insights into RTK activation using cryo-EM.** A) Left panel: cryo-EM map of the insulin receptor (IR) with four insulin (EMDB code EMD-20522). Right panel: coloured cryo-EM map of the ECD complex of IR/insulin. Density of the TM domain (dotted box) fitted with two transmembrane helices. Adapted from Uchikawa et al. (305). B) Left panel: cryo-EM map of the type 1 insulin-like growth factor receptor (IGF-1R) with one IGF1 (EMDB code EMD-20524). Right panel: the structure of IGF-1R/IGF1 fitted inside the cryo-EM map of the ECD complex. Density of the detergent micelle and TM domain after reconstruction are marked with a dotted box. Adapted from Li et al. (306)

Unlike the IR protomers, which showed symmetric conformational changes following the binding of four insulin molecules, IGF-1R protomers exhibited asymmetric hinge motion that led to the formation of an asymmetric complex (Figure 12B) (306,307), presumably due to the association of only one IGF at the primary binding site (315). The rearrangement of the extracellular and transmembrane domains upon ligand binding in both cases highlights the structural flexibility of RTKs and the important role the TM domain plays in receptor activation. However, it is noteworthy that for IR and IGF-1R, although the full-length form of the purified proteins was used in the structural determination by cryo-EM, only the extracellular domains could be reconstructed at high resolution. While the densities of the transmembrane domains can be resolved to low resolution, those of the
intracellular domains are completely unresolved in the cryo-EM maps (Figure 11), due to the structural heterogeneity and flexibility of the transmembrane and TK domains, which is consistent with the observation of different TK conformations observed in the NS-EM studies of EGFR (Figure 10A) and KIT (303).

Figure 12. Activation mechanisms of IR and IGF-1R. Cartoon representation of IR activation upon insulin stimulation (A) and insulin- or IGF1-induced IGF-1R activation (B). A) The two IR protomers are coloured in blue and red, respectively. The four insulin molecules bind at two distinct sites, which are coloured in grey and yellow. The crystal structure of the ECD of the unliganded IR (dotted circle, PDB code 4ZXB) and cryo-EM structure of the ECD of the insulin-bound IR (dotted circle, PDB code 6PXV) are shown beneath. Nanobodies in the original crystal structure are removed for clarity. B) The two IGF-1R protomers are coloured in light blue and brown, respectively. In the original published structures, the assignment of α-CTs was not conclusive as the corresponding densities are not continuous with the main chain; therefore, α-CTs, after ligand binding, are coloured half-blue and half-brown in this figure. The crystal structure of the ECD of the unliganded IGF-1R (dotted circle, PDB code 5U8R) and cryo-EM structure of IGF1-bound IGF-1R (dotted circle, PDB code 6JK8) are shown beneath. IR: insulin receptor; IGF-1R: the type 1 insulin-like growth factor receptor; α-CT: the C-terminus of the α-chain.
4.3 Structural insights into full-length RET activation

The investigation of full-length RET has so far been limited to cellular studies and no structure or low-resolution NS-EM snapshots of full-length RET have been reported. Through previous structural and functional studies of RET^{ECD} with or without the presence of its ligands, an equilibrium has been observed between apo RET monomers and apo dimers (18). The apo RET^{ECD} dimer conformation appeared to be autoinhibitory, but is not conclusive due to the structural heterogeneity revealed by the cryo-EM 2D classes (Figure 13A). Similar to that described for apo IR and apo IGF 1R, the N-terminal domains of apo RET were shown to closely associate with each other, keeping the C-terminus of RET^{ECD}, and thus the TM and TK domains, separated. Binding of the GFLs and their co-receptors induces conformational change at the ECD and draws the C-termini of the ECDs into close proximity, favouring the dimerization of the RET^{TM} domains. However, the last dozen residues at the C-terminus of RET^{ECD} have not been resolved in any of the RET complex structures, probably due to the flexibility at the terminal region of RET^{ECD}. The distance measured between the terminal residues of each RET^{ECD} protomer in the complexes, for instance RET^{ECD}/GFRAL^{ECD}/GDF15 (PDB code 6Q2J), is approximately 41 Å. Considering the number of the unresolved residues, such a distance indicates that the final amino acids of RET^{ECD} are likely to be very close to each other in the active conformation of RET.

Apart from the studies of RET activation under normal physiology, little information is known about the molecular mechanism of RET oncogenic activation in MEN2A, caused by single cysteine mutations at RET^{CRO} (i.e. how RET changes its conformation to adopt an intermolecular disulfide bond). Previous investigations into the effects of RET^{TM} mutations have shown that alterations to the composition of the RET^{TM} can prevent RET dimerization, including S649, S653 and A649/A641 point or double mutations (316,317) (Figure 13B). These results provided valuable information about the interactions that drive RET conformational changes and dimerization, but a more complete understanding of RET activation is yet to be unveiled. As mutations at the RET^{CRO}, especially C634, drive MEN2A and are close to the end of RET^{ECD}, high-resolution structures of the entire RET^{ECD}, RET^{C634R} covalently-linked dimer, and full-length RET are required to shed light on the self-association of RET^{TM} and to more fully understand RET biology.
**Figure 13. Activation mechanisms of RET receptor tyrosine kinase.** A) Cartoon representation of proposed RET activation mechanisms. The two RET monomers are coloured in yellow and grey. GFRαs and GFL dimers are coloured in red and pink, respectively. Lower panel: RET activation under normal physiological conditions. A representative 2D class, a proposed cryo-EM model of apo-RET<sup>EC</sup> dimer with two RET structures fitted (dotted circle, figures adapted from Li et al. (18)) and a cryo-EM structure of RET<sup>EC</sup>/GFRα1/GDNF (dotted circle, PDB code 6Q2N) are shown beneath. Upper panel: constitutive activation of RET through oncogenic mutation. A hypothesized model of RET dimerization via an intermolecular disulfide linkage across the two cysteine-rich domains (dotted circle). B) An example of the association of two RET<sup>TM</sup> in the plasma membrane and the hypothetical amino acid arrangement of RET<sup>TM</sup> adapted from Kjær et al. (316). Key amino acids responsible for the association are coloured in olive and red (marked by stars). TK: tyrosine kinase domain; TM: transmembrane domain.
II AIMS OF THE STUDY

The aim of this thesis project was to gain deeper understanding of the activation mechanisms of RET receptor tyrosine kinase using a combination of biochemical and biophysical tools.

The specific objectives were:

1. To characterize the binding of RET for its extracellular domain ligands, with emphasis on ephrin A5 (Study I) and GDF15/GFRAL (Study II and IV).

2. To understand RET oncogenic activation through C634R mutation-induced RET\textsuperscript{ECD} dimerization (Study IV).

3. To establish a pipeline for the expression and solubilization of full-length RET (Study III).
III MATERIALS AND METHODS

DNA constructs (Table 5), cell lines (Table 6), assays and methods (Table 7) as well as programs (Table 8) that have been used in this thesis project are listed in the tables below. Detailed description of each method can be found in the original publications (I-IV) as indicated.

1 DNA constructs

Table 5. List of the constructs used in this thesis project. EUG: Ecdysteroid UDP-glucosyltransferase.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cloned residues</th>
<th>Signal peptide</th>
<th>Vector</th>
<th>Referred as</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>FH-Thrombin-zfna-(A^ECD)</td>
<td>21–204</td>
<td>EUG</td>
<td>pK503.9</td>
<td>zfna-(A^ECD)</td>
<td>I</td>
</tr>
<tr>
<td>FH-Thrombin-zfna-(A_2ECD)</td>
<td>17–174</td>
<td>EUG</td>
<td>pK503.9</td>
<td>zfna-(A_2ECD)</td>
<td>I</td>
</tr>
<tr>
<td>FH-Thrombin-zRET(ECD)</td>
<td>22–626</td>
<td>EUG</td>
<td>pK503.9</td>
<td>zRET(ECD)</td>
<td>I</td>
</tr>
<tr>
<td>FH-Thrombin-zGFR(\alpha_1ECD)</td>
<td>31–351</td>
<td>EUG</td>
<td>pK503.9</td>
<td>zGFR(\alpha_1ECD)</td>
<td>I</td>
</tr>
<tr>
<td>F-zGDNF</td>
<td>90–236</td>
<td>EUG</td>
<td>pK503.9</td>
<td>zGDNF</td>
<td>I</td>
</tr>
<tr>
<td>mEPH(A_4LBD)-Thrombin-Fc</td>
<td>29–210</td>
<td>GP64</td>
<td>pMA152a</td>
<td>mEPH(A_4LBD)</td>
<td>I</td>
</tr>
<tr>
<td>hRET(ECD)-TEV-H</td>
<td>28–635</td>
<td>Igx</td>
<td>pcDNA3.1</td>
<td>RET (mammalian)</td>
<td>II, IV</td>
</tr>
<tr>
<td>hRET(ECD)-TEV-HF</td>
<td>28–635</td>
<td>Igx</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)-TEV-HF</td>
<td>28–635</td>
<td>CD33</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C634R)-TEV-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C634R)-TEV-Gs-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C634R, C87R, C216S)-TEV-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C634R, C87R, C216S)-TEV-Gs-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C6430A, C643R)-TEV-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-Fc</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(N336Q)-TEV-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(N336Q)</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(N343Q)-TEV-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(N343Q)</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(N468Q)-TEV-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(N468Q)</td>
<td>II</td>
</tr>
<tr>
<td>hRET(ECD)(C634R, C87R, C216S)-Thrombin-Gs-Fc-HF</td>
<td>1–635</td>
<td>Native</td>
<td>pcDNA3.1</td>
<td>RET(C634R)-thr</td>
<td>IV</td>
</tr>
<tr>
<td>Fc-Thrombin-hGDF15</td>
<td>195–208</td>
<td>Igx</td>
<td>pcDNA3.1</td>
<td>Fc-GDF15</td>
<td>II, IV</td>
</tr>
<tr>
<td>Fc</td>
<td>195–208</td>
<td>Igx</td>
<td>pcDNA3.1</td>
<td>Fc-GDF15</td>
<td>II, IV</td>
</tr>
<tr>
<td>Fc-Thrombin-hRET(ECD)</td>
<td>28–635</td>
<td>EUG</td>
<td>pK503.9</td>
<td>RET (insect)</td>
<td>II</td>
</tr>
<tr>
<td>hGFRAL(ECD)-TEV-HS</td>
<td>19–351</td>
<td>EUG</td>
<td>pK503.9</td>
<td>GFRAL</td>
<td>II, IV</td>
</tr>
<tr>
<td>hRET9-TEV-eGFP-H</td>
<td>28–1072</td>
<td>EUG</td>
<td>pK503.9</td>
<td>GFRAL</td>
<td>III</td>
</tr>
<tr>
<td>FH-Thrombin-hRET9-TEV-eGFP-H</td>
<td>28–1072</td>
<td>EUG</td>
<td>pK503.9</td>
<td>GFRAL</td>
<td>III</td>
</tr>
</tbody>
</table>
2 Cell Lines and Cell Culture

Table 6. List of cell lines used in this thesis project. DMEM: Dulbecco’s modified Eagle medium; FBS: fetal bovine serum; A&A: anti-biotic and anti-mycotic; NEAA: non-essential amino acids; SR: serum replacement; LM: lipid mixture.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Manufacturer</th>
<th>Culture condition</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>XL-10 Gold</td>
<td>Stratagene</td>
<td>Luria-Bertani medium (LB) with vector-specific anti-biotics</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Spodoptera frugiperda (SF9)</td>
<td>Thermo Fisher Scientific</td>
<td>SF900 II; Insect-XPRESS</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Trichoplusia ni High Five (Hi5)</td>
<td>Thermo Fisher Scientific</td>
<td>Insect-XPRESS</td>
<td>I, II, IV</td>
</tr>
<tr>
<td>Human embryonic kidney 293T cells (HEK293T)</td>
<td>American Type Culture Collection, CRL-3216</td>
<td>Maintenance: DMEM, 10% FBS Transfection: DMEM, 4% FBS, 1 x A&amp;A; DMEM, 1 x SR, 1 x LM, 1 x A&amp;A</td>
<td>II, IV</td>
</tr>
<tr>
<td>Chinese hamster ovary cells (CHO)</td>
<td>A kind gift from Prof. Mart Saarma (University of Helsinki, Finland)</td>
<td>Maintenance: DMEM, 10% FBS, 1 x NEAA Transfection: DMEM, 4% FBS, 1 x NEAA, 1 x A&amp;A</td>
<td>II</td>
</tr>
<tr>
<td>Chinese hamster ovary K1 cells (CHO-K1)</td>
<td>A kind gift from Dr. Helena Vihinen (University of Helsinki, Finland)</td>
<td>Maintenance: DMEM, 10% FBS, 1 x NEAA Transfection: DMEM, 4% FBS, 1 x NEAA, 1 x A&amp;A</td>
<td>II</td>
</tr>
</tbody>
</table>

3 Methods and Assays

Table 7. List of methods used in this thesis project. PCR: polymer chain reaction; PEI: polyethyleneimine; SEC-MALS: Size exclusion chromatography-coupled multicycle static laser light scattering.

<table>
<thead>
<tr>
<th>Method</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning (PCR, enzymatic ligation, site-directed mutagenesis, restriction enzyme digestion)</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>DNA extraction</td>
<td></td>
</tr>
<tr>
<td>Mini-prep with silica spin columns</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Isopropanol precipitation for Bacmid DNA</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>DNA maxi/giga prep with isopropanol precipitation for transfecting mammalian cells</td>
<td>II, IV</td>
</tr>
<tr>
<td>Protein expression</td>
<td></td>
</tr>
<tr>
<td>Baculovirus-infected insect cells</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Mammalian cell expression system (transient expression with PEI)</td>
<td>II, IV</td>
</tr>
</tbody>
</table>
Protein purification
- Affinity purification using Ni-NTA resin I, II, III, IV
- Affinity purification using anti-Flag resin II, III, IV
- Affinity purification using Strep-Tactin resin I, II, IV
- Affinity purification using Protein A/G resin I, II, IV
- Size exclusion chromatography I, II, IV
- Cell membrane preparation III
- Detergent solubilization for membrane proteins III

SDS PAGE I, II, III, IV
Western blotting I, II, III, IV
SEC-MALS II, IV

Protein-protein interaction characterization
- Mass spectrometry I
- Native PAGE I, II, IV
- Pull down assay with Ni-NTA resin I
- Pull down assay with Protein A/G resin I, II
- Pull down assay with anti-Flag resin I
- Biolayer interferometry system (BLItz) with Ni-NTA biosensor I
- Biolayer interferometry system (BLItz) with anti-human IgG (AHC) biosensor II, IV

Negative stain EM (grid preparation, data collection and processing) IV
Cryo-EM (grid preparation and data collection and processing) IV

4 Programs

Table 8. List of programs and applications used in this thesis project.

<table>
<thead>
<tr>
<th>Program</th>
<th>Reference</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP4 package: General use of program tools</td>
<td>(318)</td>
<td>IV</td>
</tr>
<tr>
<td>Coot: Model building</td>
<td>(319,320)</td>
<td>IV</td>
</tr>
<tr>
<td>MacPyMOL: Analysis and visualization of structures</td>
<td>Version 2.2.0, Schrodinger LLC.</td>
<td>I, II, IV</td>
</tr>
<tr>
<td>UCSF Chimera: Analysis and visualization of structures and EM densities</td>
<td>(321)</td>
<td>IV</td>
</tr>
<tr>
<td>GraphPad Prism 6/Prism 8: Data analysis and curve fitting</td>
<td>Prism 6/Prism 8</td>
<td>I, II, IV</td>
</tr>
<tr>
<td>RELION 3: EM data processing</td>
<td>(254)</td>
<td>IV</td>
</tr>
<tr>
<td>EMAN2.3: EM data processing</td>
<td>(322)</td>
<td>IV</td>
</tr>
<tr>
<td>crYOLO: Automated particle picking</td>
<td>(257)</td>
<td>IV</td>
</tr>
<tr>
<td>Fiji (Image J): PAGE gel and western blotting image analysis</td>
<td>(323)</td>
<td>I, II, III, IV</td>
</tr>
</tbody>
</table>

49
IV RESULTS AND DISCUSSION

1 RET\textsuperscript{WT} activation by extracellular ligand binding

1.1 Investigating the interaction between RET\textsuperscript{ECD} and efn-A5\textsuperscript{ECD}

Based on cellular assays, including colocalization and immunoprecipitation, RET was previously postulated to be a receptor for efn-A5 “reverse” signalling (60,61). However, in vitro study was lacking to elucidate whether the interaction between efn-A5 and RET was direct. Mammalian RET is prone to misfold during protein production (this thesis - Study II)(23,24); therefore, in this study, zebrafish RET\textsuperscript{ECD} (zRET\textsuperscript{ECD}) and zebrafish efn-A5\textsuperscript{ECD} (zefn-A5\textsuperscript{ECD}) were used to investigate their interaction, since the Eph receptors, efn ligands and RET are all evolutionarily conserved among vertebrates.

Using the baculovirus expression vector system (BEVS), soluble zRET\textsuperscript{ECD} and zefn-A5\textsuperscript{ECD} were recombinantly expressed and the proteins purified separately. To verify the purified products were functional, the binding capacity of zefn-A5\textsuperscript{ECD} for mouse EphA4 ligand-binding domain (mEphA4\textsuperscript{LBD}) and zRET\textsuperscript{ECD} for zebrafish GDNF/GFRα1\textsuperscript{ECD} (zGDNF/zGFRα1\textsuperscript{ECD}) was determined using bio-layer interferometry, native PAGE and pull-down assays. I showed that mEphA4\textsuperscript{LBD} could be efficiently pulled down by zefn-A5\textsuperscript{ECD} and vice versa. Using bio-layer interferometry, the measured binding affinity of mEphA4\textsuperscript{LBD} for zefn-A5\textsuperscript{ECD} was 0.18 μM ± 0.01 μM, similar to that of hEphA4\textsuperscript{LBD} and hefn-A5\textsuperscript{ECD} (324), confirming the functionality of the purified zefn-A5\textsuperscript{ECD} (Figure 14A). Additionally, analysis of the results of BN PAGE (Figure 14C) and mass spectrometry showed that zGDNF/zGFRα1\textsuperscript{ECD} formed a complex with zRET\textsuperscript{ECD}, which verified the ligand-binding capacity of zRET\textsuperscript{ECD}. Notably, while zefn-A5\textsuperscript{ECD} stayed monomeric in solution, zRET\textsuperscript{ECD} appeared as a combination of monomeric, dimeric and, to a less extent, higher oligomeric species. Whereas the purified zefn-A5\textsuperscript{ECD} and zRET\textsuperscript{ECD} were active towards their respective positive control binding partners, no binding was detected between zRET\textsuperscript{ECD} and zefn-A5\textsuperscript{ECD} using either bio-layer interferometry (Figure 14B) or BN PAGE (Figure 14C) with or without the addition of mEphA4\textsuperscript{LBD}, zGDNF and zGFRα1\textsuperscript{ECD}. The results clearly indicated that zefn-A5\textsuperscript{ECD} does not interact with zRET\textsuperscript{ECD} directly and suggests that the “reverse” signalling of efn-A5 may be more complicated than previously hypothesized (60).
In my study, zefn-A5EC, zGFRα1EC and zRETEC did not form a complex. As these proteins are evolutionarily conserved, it is unlikely that their mammalian homologs form a signalling complex either. Instead, the association of other modulators is likely a prerequiste for the formation of the RET and efn-A signalling complex. In 2014, Chai et al. reported that mouse efn-A5, mouse GFRα1 and mouse RET co-immunoprecipitated with Celsr3/Fzd3, which suggests that Celsr3/Fzd3 may mediate the signalling of efn-A5 with GFRα1 and RET (61). Furthermore, Mandai et al. reported that the leucine-rich repeat and immunoglobulin (LRIG) family receptor, Linx, interacts with Celsr3/Fzd3, TrkA, p75 neurotrophin receptor (NTR) and RET, but not EphA receptors (325). Interestingly, among these Linx-interacting receptors, TrkA and p75NTR have also been shown to bind to efn-As (326,327) and to crosstalk with RET (328,329) under the stimulation of neurotrophins. In addition to the above mentioned receptors, integrins have...
also been reported to directly interact with RET (330) and, directly or indirectly, associate with efn-A5 in the presence of EphA receptors (331). The wealth of efn-A5-associated ligands, receptors and co-receptors that have now been identified illustrate the complexity of the cellular regulation of signalling transduction machinery (Figure 15). In future studies, it will be interesting to identify exactly which ligands and/or co-receptors are necessary for the mediation of the “reverse” signalling of efn-A ligands through RET.

![Figure 15. Schematic model of RET and efn-A signalling.](image)

The efn-A ligands may interact with RET/GDNF/GFRα1 and Celsr3/Fzd3 for its reverse signalling, together with Linx, β integrins, Trk and p75NTR may have potential roles in mediating efn-A signalling with RET. For clarity, interactions that involve efn-A, RET and Linx are linked with blue, grey and red lines, respectively (for the interactions between efn-A, RET or Linx, only one of the colors was used).

### 1.2 Characterization of the RET$^{ECD}$/GDF15/GFRAL$^{ECD}$ complex

#### 1.2.1 Expression of functional human RET$^{ECD}$

In late 2017, GFRAL was identified to be the orphan receptor of GDF15 and that the GDF15/GFRAL complex was found to signal through RET, regulating energy intake and expenditure for neurons in the hindbrain area postrema (AP) and nucleus of the solitary tract (NTS) (45,47–49). At the time that this study was initiated, it was known that all GFLs and GFRαs form complexes with RET with the same 2:2:2 constitution but no high-resolution structural studies of these RET complexes had been reported and it was not clear how these different ligand and co-receptor pairs bind to and activate RET. Therefore, in this study, I aimed to recombinantly produce the wild-type human RET$^{ECD}$ (hRET$^{ECD}$(WT)), human GDF15 (hGDF15) and human GFRAL$^{ECD}$ (hGFRAL$^{ECD}$) in vitro and investigate the structure of the hRET$^{ECD}$(WT)/hGDF15/hGFRAL$^{ECD}$ complex using cryo-EM.
In order to produce sufficient amount of material for the structural investigation, it is essential to recombinantly express the proteins in the properly folded active conformation with a satisfactory yield. Although the baculovirus expression system has shown a lot of success in expressing functional RTKs, including zRETECD (this thesis – Study I), full-length hEGFR (301), hEphAECD (332) and hTrkECD (333), I found that hRETECD(WT) expressed using the baculovirus-infected Sf9 cells contained a mixture of monomeric, dimeric and higher oligomeric species (Figure 16A).

![Figure 16. Analysis of recombinantly expressed hRETECD(WT). A) SDS-PAGE image showing hRETECD(WT) expressed using insect (RET (insect)) and mammalian cells (RET (mammalian)) under reducing and non-reducing conditions. B) Sensograms of hRETECD(WT) expressed using insect cells (RET (insect)) at different concentrations binding to immobilized Fc-GDF15/GFRAL. C) BN PAGE image showing RET (mammalian) formed complex with GDF15/GFRAL (marked with a red star) but RET (insect) could not. The bands corresponding to the GDF15/GFRAL complex are marked with black stars.](image)

Analysis of the hRETECD(WT) expressed in Sf9 cells using biochemical and biophysical techniques demonstrated that the protein was misfolded, containing significant disulfide-
mismatching as demonstrated by comparing reducing and non-reducing SDS-PAGE (Figure 16A), and was inactive with respect to binding to hGDF15/hGFRAL ECD (Figure 16B,C). While this non-functional hRET ECD(WT) has undergone the secretion pathway when expressed using Sf9 cells, it was not secreted when expressed in Hi5 insect cells that had been successfully used for the production of functional zRET ECD(WT) (this thesis - Study 1). On the contrary, when expressed using the mammalian cells, hRET ECD(WT) was monomeric in solution and its activity was confirmed using BN PAGE, demonstrating that it formed the expected complex with hGDF15/hGFRAL ECD (Figure 16A,C). After optimizing the construct by varying the type of signal peptide (Igκ, CD33 or native peptide) and the location of the affinity tags (at the N terminus or C-terminus), a final construct was selected for large scale expression of hRET ECD(WT) for structural studies, which contains the native signal peptide of hRET with C-terminal HIS and Flag tags.

The extracellular domain of hRET contains 28 cysteine residues that form 13 disulfide bonds, with the remaining two cysteines being unpaired. The large number of disulfide bonds involved in the correct folding of RET ECD gives rise to the potential for protein misfolding. This may explain why zRET ECD(WT), which contains fewer cysteine residues and different disulfide arrangements than hRET ECD(WT), was successfully expressed using Hi5 cells and was functional. However, although the majority was monomeric, purified zRET ECD(WT) still showed a certain level of heterogeneity (Figure 14C), possibly due to disulfide-mismatch similar to that of hRET ECD(WT) (Figure 16C) when expressed in insect cells. The results show that mammalian cells perform better in expressing cysteine-rich RET ECD than insect cells and highlights how the choice of protein expression system can make a significant difference in functional recombinant protein production, especially for challenging eukaryotic proteins.

1.2.2 Structural investigation of the hRET ECD/hGDF15/hGFRAL ECD complex

Complex reconstruction

In order to obtain sufficient material for structural studies, hRET ECD(WT) and Fc-hGDF15 were expressed in HEK293T cells in large scale (>1 L) using either roller bottles or multi-layered flasks and hGFRAL ECD was expressed using Hi5 cells in suspension culture. The purified proteins were subsequently mixed in an equimolar ratio to reconstitute the complex. I used SEC-MALS to assess the oligomeric state of the reconstructed complex and the elution fractions were further analysed by BN PAGE. The major peak, which eluted between 9–11 ml correlated to the desired hRET ECD(WT)/hGDF15/hGFRAL ECD hexameric complex, showing a calculated MW of ~393 kDa (theoretical unglycosylated complex MW = 253 kDa, expected glycosylation ~100 kDa) (Figure 17A). The earlier fractions contained higher-order oligomers and the
later fractions contained excess hRET<sup>ECD</sup>(WT) monomer that had not formed a complex. The peak fractions containing hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> were then assessed using SDS PAGE, demonstrating that the dimeric hGDF15 was be reduced to monomeric form under reducing conditions while both hRET<sup>ECD</sup> and hGFRAL<sup>ECD</sup> remained monomeric with or without the addition of the reducing agent. Analysis of the intensity of the bands corresponding to each protein component was consistent with hRET<sup>ECD</sup>, hGDF15 and hGFRAL<sup>ECD</sup> forming a complex with an equimolar ratio.

**Figure 17. The formation of the hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> complex.** A) Characterization of the oligomeric state of hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> using SEC-MALS (upper panel). Elution fractions from SEC-MALS (marked with a black rectangle) were collected and analysed using BN PAGE (lower panel). The peak fractions of hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> are highlighted with dotted rectangles. B) SDS PAGE analysis of hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> from the peak fractions under non-reducing (left) and reducing conditions (right).

**EM grid optimization**

The hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> complex was prepared on a larger scale, purified by gel filtration and used to prepare both negative stain EM and cryo-EM grids. A good distribution of particles that were homogenous in their dimensions was observed on the NS-EM grids, showing different views of the complex (Figure 18A). However, in the initial cryo-EM screening trials, I found the particles were largely aggregated on the cryo-EM grids (Quantifoil Cu, Figure 18B), despite the positive observations in the NS-EM micrographs using the same batch of sample. A follow-up cryo-EM grid optimization trial varying the kinds of grid support and protein concentration showed that the hRET<sup>ECD</sup>(WT)/hGDF15/hGFRAL<sup>ECD</sup> complex particles tended to aggregate on the Quantifoil copper grids and to a lesser extent on graphene-oxide coated copper grids.
(Figure 18B,E). Pleasingly, however, the particles were well distributed on both UltrAuFoil and Quantifoil gold grids when deposited at a protein concentration of ~0.1 mg/ml. The specimen prepared on the Quantifoil gold grids showed superior contrast relative to the UltrAuFoil grids and was therefore selected for data collection (Figure 18C,D).

Figure 18. Cryo-EM sample optimization by altering grid support. A) Negative stain EM image of the hRET<sup>EC</sup> (WT)/hGDF15/hGFRAL<sup>EC</sup> complex. Right panel is an enlarged view of the area marked with a black square. Image was collected on a F12 at 120kV and 30,000 magnification. B-E) Cryo-EM images of the particle distribution of the same hRET<sup>EC</sup> (WT)/hGDF15/hGFRAL<sup>EC</sup> sample using different grid supports. Images were collected on a Titan Krios electron microscope (300 kV) at 75,000 nominal magnification.

Solving the cryo-EM structure of hRET<sup>EC</sup> (WT)/hGDF15/hGFRAL<sup>EC</sup>

Following specimen optimization, a large dataset was collected for 3D reconstruction. After three rounds of 2D and 3D classification, bad particles were removed and a set of 142083 particles were used for auto-refinement. The final reconstructed model converged at 8.0 Å resolution with C2 symmetry. Segmentation of the map using Chimera identified 12 major components (Figure 19B). Docking of the crystal structure of the dimeric GDF15/GFRAL<sup>EC</sup> (PDB code 5VZ4) shows that the GDF15 dimer fits well in the central density (Figure 19C, green), which led to the assignment of the densities of two GFRAL<sup>EC</sup> molecules (Figure 19C, blue), and revealed that the location of each
component matched well with the segmented map. A conformational change of GFRAL\textsuperscript{ECD} was observed by comparing the location of the helices of GFRAL\textsuperscript{ECD} in the docked crystal structure with the corresponding density map, where GFRAL\textsuperscript{ECD} appeared to exhibit an outward rotational movement upon RET binding (Figure 19C, right panel).

![Figure 19](image)

Figure 19. Cryo-EM reconstruction of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex. A) Segmented refined cryo-EM 3D model showing different coloured domains of hRET\textsuperscript{ECD}(WT), hGDF15 and hGFRAL\textsuperscript{ECD}. Density that potentially belongs to GFRAL\textsuperscript{D1} is marked with a dotted circle. D: domain. B) A crystal structure of hGDF15/hGFRAL\textsuperscript{ECD} (PDB code 5VZ4) fitted in the segmented cryo-EM map showing the conformational difference of GFRAL\textsuperscript{ECD} in the GDF15/GFRAL\textsuperscript{ECD} and the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complexes. C) A modified structure of the hRET\textsuperscript{ECD}(WT)/GDF15/GFRAL complex (purple, PDB ID: 6Q2J) fitted in the reconstructed density (orange). A base glycan GlcNAc was added to fit the density at the glycosylation sites using Coot.

Unfortunately in late 2019, when I was pursuing this project, a high resolution cryo-EM structure of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} was published at 4.1 Å (PDB 6Q2J) (18). This atomic model could be unambiguously docked into our refined 3D map (Figure 19C) with only minor differences at certain loop regions. The fitting of the 6Q2J structure
into my map verified that the reconstruction was correct and that the eight segmented densities, apart from the ones corresponding to a GDF15/GFRAL\textsuperscript{ECD} dimer, were indeed different domains of the two RET\textsuperscript{ECD}(WT)s (Figure 19). Analysis of the fitted map revealed that the CLD1 and CLD2 of RET\textsuperscript{ECD}(WT) associated with domain 3 of GFRAL in the complex while the RET\textsuperscript{CRD} interacted with GDF15. Notably, domain 1 of GFRAL was not resolved in either the published structure (18) or in our reconstructed map. However, a small density next to domains 2 and 3 of GFRAL was observed (Figure 19A, right panel with dotted circle). The location of the extra density agrees well with that of a large density in a recently reported negative stain EM map of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex without a domain assignment (218), which suggests that such density could be attributed to domain 1 of GFRAL. As this domain locates outside the complex and does not seem to interact with any other protein components in the complex, it may be less conformationally stable and thus could not be solved in the EM maps. Additionally, the densities of the glycans at all the 11 glycosylation sites of hRET\textsuperscript{ECD}(WT) (18,23,58) were observed in my reconstructed map (Figure 19C), confirming the observations of the SDS PAGE gel (Figure 17B) that the purified hRET\textsuperscript{ECD}(WT) was heavily glycosylated. One or two base glycans were added to the docked hRET\textsuperscript{ECD}(WT) protomers of the complex (PDB code 6Q2J) to fit in the corresponding densities using Coot (Figure 19C). The results demonstrated that the proteins produced were active and formed the tripartite hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex with a molar ratio of 2:2:2 and that the pipeline used for the structural study of the RET complexes was effective.

Further analysis of the three 3D classes that were generated during 3D classification revealed changes of the angle between the two wings of the “butterfly” shaped RET\textsuperscript{ECD}(WT) complex (Study IV – Figure 6A). To model the movement, the structures of hRET\textsuperscript{ECD}, hGDF15 dimer and hGFRAL\textsuperscript{ECD} were docked into each 3D map and aligned onto the hGDF15 dimer. Different degrees of movement were observed in the individual components: The biggest movement was seen in hRET\textsuperscript{ECD} and the smallest in the hGDF15 dimer (Study IV – Figure 6B). Overall the movement represents an outwards twist of the complex that originates from the central GDF15/GFRAL\textsuperscript{ECD} components. The big movement of RET\textsuperscript{ECD} is consistent with its versatility in binding to conformationally different ligand/co-receptor pairs (Figure 4). Similar observations have also been reported recently for the RET\textsuperscript{ECD}/NTRN/GFRα2 complex (58). Unfortunately, these conformational changes could not be analysed in atomic detail based on the moderate resolution maps in this study and further work is needed to elucidate whether the movement in the complex has any biological significance.
2 RET activation via the oncogenic C634R mutation

2.1 The dimeric oncogenic hRET\textsuperscript{ECD} C634R mutant

2.1.1 Expression of the dimeric hRET\textsuperscript{ECD}(C634R)

The C634R mutation of RET is one of the major driving mutations for MEN2. Structurally, C634 is positioned at the C-terminal end of the CRD, close to the TM domain, and forms a disulfide bond with C630 in wild-type RET. The C634R mutation renders C630 unpaired, which leads to homodimerization of RET(C634R) via the formation of an intermolecular C630-C630 disulfide bond. In a cellular context, this dimerization has been shown to drive constitutive ligand-independent activation of RET signalling but there is not a clear structural understanding of the activation mechanism. Two groups have previously reported that, when hRET\textsuperscript{ECD}(C634R) is expressed recombinantly, it is isolated solely in its monomeric form and that it does not dimerize even at high concentration (21,153). They hypothesized that ligand binding may be critical for RET(C634R) dimerization, although this is not consistent with the observation that full-length RET(C634R) auto-phosphorylates in the absence of the ligands (21,58). Therefore, I aimed to isolate hRET\textsuperscript{ECD}(C634R) in its dimeric form and to undertake a structural investigation into this mutant.

To express hRET\textsuperscript{ECD}(C634R), I first tried using the same expression conditions and expression vector that I had previously used successfully for the expression of hRET\textsuperscript{ECD}(WT). After introducing the C634R mutation by site-directed mutagenesis, I expressed and purified hRET\textsuperscript{ECD}(C634R) and found that it was indeed monomeric, as confirmed by it having the same molecular weight on non-reducing SDS-PAGE as hRET\textsuperscript{ECD}(WT). This raised the question of why full-length RET(C634R) dimerizes efficiently but not the ECD of RET(C634R)? Unlike the full-length receptor, which is locally concentrated in the cell membrane, recombinantly expressed RET\textsuperscript{ECD} is secreted into the medium at a low concentration. Furthermore, the cell culture medium used for mammalian cell culture, for example DMEM, typically creates a highly oxidising environment for the secreted proteins. As disulfide bond exchange is mediated by free thiols, I hypothesized that the combination of the low concentration of secreted RET\textsuperscript{ECD} and the oxidative environment of the medium would favour all the C630s to oxidise independently (eg to sulfinic/sulfenic/sulfonic acids) and thus prevent the formation of intermolecular disulfide bonds. Therefore, I decided to add an Fc tag at the C-terminus of hRET\textsuperscript{ECD}(C634R), which would bring two RET\textsuperscript{ECD} molecules in close proximity via the two disulfide bonds at the hinge region of the Fc dimer, creating an increased local protein concentration that would favour the formation of the C630-C630 disulfide bond (Figure
Either a TEV or Thrombin protease cleavage site was incorporated in the constructs between the hRET\textsuperscript{ECD}(C634R) gene and Fc for tag removal.

![Figure 20. Schematic representation of the method for producing the dimeric RET\textsuperscript{ECD} C634R mutant.](image)

Fc tag is cleaved by either TEV or Thrombin protease.

Analysis using SDS PAGE and WB showed that purified Fc-tagged hRET\textsuperscript{ECD}(C634R) was dimeric and, following protease cleavage, the untagged hRET\textsuperscript{ECD}(C634R) existed as a mixture of monomer and dimer, which could be separated using size-exclusion chromatography (Figure 21). The untagged dimeric hRET\textsuperscript{ECD}(C634R) fraction remained dimeric after purification and could be reduced into monomers in the presence of a reducing agent, proving that dimerization was driven by disulfide bond formation (Figure 21, right panel). Furthermore, expression of a double mutant carrying both C630A and C634R mutations produced only monomeric RET, which confirmed that the RET C634R dimer was indeed linked through a C630-C630 disulfide bond (Study II – Supplementary information). Taken together, these results show that the addition of an Fc tag does drive the dimerization of hRET\textsuperscript{ECD}(C634R).

The efficiency of the Fc tag cleavage was compared using TEV and thrombin proteases. Interestingly, a RET\textsuperscript{ECD}-TEV cross-linked product was observed after tag cleavage using the TEV protease, but no cross-linked RET\textsuperscript{ECD} C634R-Thrombin was observed. The use of Thrombin protease for hRET\textsuperscript{ECD} C634R production thus increased the purification efficiency and the yield of purified product compared to the TEV protease.

![Figure 21. Purification of the hRET\textsuperscript{ECD}(C634R) dimer.](image)

SDS PAGE gel images of the fractionated untagged-hRET\textsuperscript{ECD} C634R sample after Thrombin cleavage from gel filtration (GF) on a Hiload 16/600 Superdex 200 column (left panel) and concentrated sample from the peak fractions under reducing and non-reducing conditions (right panel). Sample before gel filtration is labelled with S.
2.1.2 Structural investigation of dimeric hRET<sup>ECD</sup>(C634R)

The purified hRET<sup>ECD</sup>(C634R) mutant was used for a structural investigation using negative stain EM (Study IV – Figure 1A,B). The collected micrographs were first inspected and processed in EMAN2.3 and a set of 3490 particles were manually picked. The 2D class averages were diverse, representing different views of the sample (Study IV – Figure 1C). Among the 2D classes, some showed a distinctive “S” shape which suggested a potential C2 symmetry. Subsequent initial model and refinement were performed with symmetry and the projections of the reconstructed model agreed well with the 2D classes (Study IV – Figure 1E). The final 3D density was refined to a final resolution of 22 Å and provided a tight fit for two copies of RET<sup>ECD</sup> (Study IV – Figure 1F). The two “C”-shaped RET<sup>ECD</sup> protomers connect to each other through the disulfide bond at the CRDs at the centre of the density, exhibiting an “S” shaped top view.

Limited by the resolution, an atomic model of the hRET<sup>ECD</sup>(C634R) dimer could not be built based on the density map. Instead, the analysis of the reconstructed map relied on positioning of the rigidly docked structures that I adapted from PDB code 6Q2J. The measured distance between the last residues of the two docked RET protomers (residue P622s) was 18 Å (Study IV – Figure 1F) The two RET protomers in the 3D density of the hRET<sup>ECD</sup>(C634R) dimer adopted an open configuration with the N-terminal domains pointing away from each other. The CRDs interact closely with each other, which I anticipate would, in the context of full-length RET, bring the TM domains into close proximity and induce the dimerization and activation of the ICDs. No other direct interaction was observed between the two protomers. This novel configuration reveals that hRET<sup>ECD</sup>(C634R) undergoes significant conformational change to allow the ligand-independent receptor dimerization compared to the auto-inhibitory conformation of a wild-type hRET dimer (18). The configuration of the hRET<sup>ECD</sup>(C634R) dimer is also different from that of the two RET<sup>ECD</sup>(WT) protomers when in complex with GFLs/GFRαs, with the distance between the C-termini of the RET<sup>ECD</sup> protomers being more than two times shorter in the apo-hRET<sup>ECD</sup>(C634R) dimer. Such difference may result in an altered activation mechanism in the TM domain and ICD. Indeed, a previous study has found that the activation by ligand-independent RET dimerization leads to enhanced AKT phosphorylation but with a much later onset compared to that induced by ligand-dependent RET activation (334). In line with this result, signalling of the ligand-independent RET dimer was also found to stimulate cell proliferation more aggressively than for wild-type RET, suggesting that the RET mutant dimer may trigger unique intracellular processes.
2.2 The liganded hRET<sup>ECID</sup>(C634R) complex

2.2.1 Characterization of the hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> complex

Although cellular studies have shown that hRET(C634R) can activate in the absence of its ligands, it is not clear whether this oncogenic dimer is capable of ligand-binding. Indeed, no such in vitro characterization has been reported using the purified proteins. Therefore, I used soluble hRET<sup>ECID</sup>(C634R) to study its binding potential with its ligand (hGDF15) and co-receptor (hGFRAL<sup>ECID</sup>) utilizing a combination of biochemical and biophysical assays. The results showed that dimeric hRET<sup>ECID</sup>(C634R) does form a complex with hGDF15/hGFRAL<sup>ECID</sup>, with a slightly weaker binding affinity as hRET<sup>ECID</sup>(WT). Surprisingly, the hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> complex is twice the size of the hRET<sup>ECID</sup>(WT)/hGDF15/hGFRAL<sup>ECID</sup> complex (Figure 22A,B) and the protein components in the complexes share equimolar ratios (Figure 22C). Since hRET<sup>ECID</sup>(WT), hGDF15 and hGFRAL<sup>ECID</sup> form a complex at 2:2:2 molar ratio, the difference of the apparent MW between the two complexes suggests that hRET<sup>ECID</sup>(C634R), hGDF15 and hGFRAL<sup>ECID</sup> form a 4:4:4 complex.

![Figure 22](image)

**Figure 22.** The formation of the hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> complex. A) BN page showing difference between the hRET<sup>ECID</sup>(WT) (black star) and hRET<sup>ECID</sup>(C634R) (red star) complexes. B) Characterization of the oligomeric states of hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> and hRET<sup>ECID</sup>(C634R) dimer using SEC-MALS. C) SDS PAGE analysis of hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> from the SEC-MALS peak fractions under non-reducing (left) and reducing conditions (right).

2.2.2 The structural investigation of the hRET<sup>ECID</sup>(C634R) complex

Next, negative stain EM was used to assess the conformational homogeneity of the hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> complex and to investigate its structure. Compared to the apo-hRET<sup>ECID</sup>(C634R) dimer, the particles of the hRET<sup>ECID</sup>(C634R)/hGDF15/hGFRAL<sup>ECID</sup> complex appeared wider and thicker but were similar in their longest dimension (Figure 23). With a set of 21776 particles, 2D classification was performed and an initial model was built using EMAN2.3 with C1 point symmetry. A clear “X” shape was observed among several 2D classes (Study IV – Figure
7B), which was also reflected by the initial model. One copy of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex (PDB code 6Q2J) could be docked into the first half of the density map (Study IV – Supplementary Figure 3). The second half of the density was less well resolved than the first half, possibly due to the flexibility of the protein complex and potential degradation during the NS-EM grid preparation. Analysis of both the 2D class averages and the initial C1 symmetric model revealed that the mutant hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} complex was likely constructed from two back-to-back hexameric complexes each with a similar configuration and C2 symmetry as the WT complex, giving the overall structure D2 symmetry. Therefore, D2 symmetry was applied in the subsequent 3D refinement, leading to a final model at 23 Å resolution (Study IV – Figure 7C). Reprojections of the model match well with the observed 2D classes (Study IV – Figure 7D). Two copies of the hRET\textsuperscript{ECD}(WT)/hGDF15/hGFRAL\textsuperscript{ECD} complex could be docked back-to-back into both halves of the density, occupying the majority of the 3D volume (Study IV – Figure 7E). The back-to-back fitting is consistent with each of the 2:2:2 complexes being linked by two C630-C630 disulfide bonds and overall gives a “twin-butterfly”-shaped dodecameric hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} complex, which is consistent with the results described earlier where I used BN PAGE and SEC MALS to show that the three proteins interact in a 4:4:4 ratio (Figure 22).

![Figure 23. Comparison between the apo-hRET\textsuperscript{ECD}(C634R) dimer and hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} using negative stain EM. Representative micrographs of (A) the hRET\textsuperscript{ECD}(C634R) dimer and (B) the hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} complex, showing their particle distribution. Scale bar: 200 nm.](image)

To pursue a higher resolution reconstruction, a large dataset was collected using cryo-EM of the hRET\textsuperscript{ECD}(C634R)/hGDF15/hGFRAL\textsuperscript{ECD} complex. After motion correction and CTF estimation using RELION3, 288927 particles were auto-picked and were used for
2D classification. The 2D classes showed a potential degradation of the complex and a high degree of conformational heterogeneity (Study IV – Figure 8B). Comparison between the 2D classes of the hRET<sup>EC</sup>D(WT) and hRET<sup>EC</sup>D(C634R) complexes revealed several unique class configurations. However, I could not obtain useful initial models using this dataset, which showed maps with a high level of noise, possibly due to the flexible disulfide linkers between the two RET protomers within a hRET<sup>EC</sup>D(C634R) dimer that connects the two complex subunits.

The results clearly showed that the hRET<sup>C634R</sup> dimer could bind to its ligands through a novel mechanism. In comparison to the hRET<sup>EC</sup>D(WT), the different binding mode of hRET<sup>EC</sup>D(C634R) is likely due to the spatial restriction imposed by the intermolecular disulfide bond at the C-termini. The findings support previous observations in the cellular assays where the auto-phosphorylation level of full-length hRET<sup>C634R</sup> increased after stimulation by GDNF (335,336). Because only the extracellular domains of hRET(C634R) and hGFRAL were used in this study, the formation of the tripartite complex did not have the restriction of the transmembrane domain and the cell membrane that would exist in a cellular setting. Therefore, in the context of the full-length receptors, I hypothesize that neighbouring hRET<sup>C634R</sup>s may form a side-by-side “twin-butterfly” complex with the ligands and co-receptors (Study IV – Figure 9).

3 Expression and purification of full-length RET

The results from the structural investigations of the extracellular domain of RET are very useful for understanding how ligand binding and the oncogenic mutations drive the conformational changes and complex formation of RET. However, to understand fully the relationship between the structure and signalling of RET, these structural studies must be extended to the full-length protein so that the way the extracellular domain influences the activity of the kinase domain can be revealed. The aim of this study was to provide a pipeline for the expression and purification of full-length RET in a biologically relevant system for EM investigations. Detergent solubilization is the most wildly used approach in solubilizing membrane proteins and has been successfully applied to the structural studies of several full length RTKs (Table 4). For this study, full-length hRET was expressed using baculovirus-infected Sf9 cells, as this work was undertaken prior to our knowledge that hRET activity is sensitive to the choice of expression host (Study II). To visualize the location of the full-length hRET during purification, an eGFP tag was incorporated at the C-terminus of RET (Study III). Under the fluorescence microscope, a green fluorescent ring was detected, demonstrating that RET-eGFP was embedding properly into the cell membrane (Figure 24).
Figure 24. Full-length RET cell membrane expression visualized using a GFP tag. Transfected S/9 cells showing the expression and localization of full-length RET with a GFP tag at the plasma membrane. Images were taken under white light (left), excited at 488 nm (middle) and superimposition of the previous two images on a laser confocal microscope (40 ×).

However, after lysing the cells by high pressure homogenization and pelleting the membrane proteins by ultracentrifugation, only a small fraction of RET (~10%) could be resolubilised directly using common detergents such as DDM and LMNG, resulting in the isolation of RET at low concentration in large volumes of detergent solution. In order to enrich the fraction of viable RET prior to detergent solubilization, an adapted sucrose cushion approach was implemented (337,338). In this approach, soluble proteins are isolated in the sucrose light fraction, highly aggregated membrane proteins are excluded in the membrane pellet and non-aggregated membrane proteins can be isolated from the sucrose heavy fraction (Figure 25).
To compare the solubility of full-length RET prepared with or without the sucrose cushion method, two identical samples were prepared. After ultra-centrifugation (Figure 25, step 3), the sucrose light fraction was disposed of. Then the sucrose heavy fraction was isolated alone to give the sucrose cushion fraction (S). For the total membrane fraction (T), the sucrose heavy fraction was resuspended with the membrane pellet. A parallel detergent screen was then performed on both the sucrose cushion fraction and the total membrane fraction by diluting them into a range of detergents as 2% solutions containing 0.2% CHS. After another round of ultra-centrifugation, the amount of detergent soluble RET-eGFP was determined by in-gel fluorescence and western blotting using an anti-HIS antibody. Overall, the solubility of RET isolated in the sucrose cushion fraction (S) was significantly improved compared to that of RET from the total membrane fraction (T), demonstrating that success of the approach (Figure 26A,B). In the case of LMNG with CHS, the solubility of RET was nearly 100% (a 6-fold increase) when prepared using the sucrose cushion method (Figure 26C). The result was consistent when the expression was scaled up and I demonstrated that RET solubilized in LMNG with CHS could be subsequently purified through affinity purification.

The challenges of membrane protein solubilization include poor protein solubilization efficiency, the high cost of the solubilizing agents, and the instability and low activity of the solubilized proteins. In this project, a sucrose cushion method was successfully adapted for the isolation of detergent-soluble RET, which significantly improved the solubility and reduced detergent consumption during the solubilization.

Figure 25. Membrane protein purification pipeline using sucrose cushion. Step 1: Prepare cell lysate in light sucrose buffer in an ultracentrifuge tube. Step 2: Add the heavy sucrose buffer to the bottom of the tube. Step 3: After ultracentrifugation, soluble proteins (in light sucrose fraction) are separated from membrane proteins (in heavy sucrose fraction and pellet). This step can be repeated for improved purity. Step 4: Detergent solubilization of the samples in the heavy sucrose fraction. Step 5: Ultracentrifugation to remove detergent-insoluble fraction. Repeat Steps 1-3 for increased purity (optional). MP: membrane protein.
process. This method could also be adapted in the production of other challenging membrane proteins. In this study, full-length hRET was expressed using the baculovirus-infected Sf 9 cells prior to our knowledge on the hRET activity dependency on the expression host (Study II). Although the function of the solubilized full-length RET in Study III is yet to be assessed, it nonetheless showed the effectiveness of the sucrose cushion approach in isolating detergent-soluble proteins. In future studies, it will be interesting to apply the same method in the solubilization of full-length RET expressed using the mammalian cells and compare the results with that expressed using the insect cells.

![Figure 26. Full-length RET solubilization. A,B) Detergent solubilization screening of full-length RET-eGFP (180 kDa) in the total membrane fraction (A) and the heavy sucrose cushion fraction (B). Four representative detergents (2%) with 0.2% CHS were used to solubilize RET and the results were shown in the fluorescence gel and anti-HIS WB images. C) RET solubilization table. The solubility of RET under each condition is calculated based on the ratio of band intensities of the detergent solubilised fraction (S) to the total protein before ultracentrifugation (T).](image-url)
V CONCLUSION AND FUTURE STUDIES

The RET receptor tyrosine kinase has a wealth of interacting ligands and co-receptors and the understanding of the activation mechanism of the RET signalling complexes was limited due to the lack of *in vitro* functional and structural studies. This thesis project focussed on functional and structural investigations of the RET/efn-A system, the RET/GDF15/GFRαL signalling complex and the oncogenic RET C634R mutant.

**Study I** showed that the insect cell baculovirus expression system is suitable for the recombinant expression of functional zRET<sup>ECD</sup> and zefn-A5, which was verified by successfully preparing the positive control complexes zRET<sup>ECD</sup>/zGDNF/zGFRα1 and zefn-A5/mEphA4. Contrary to previously published studies, **Study I** demonstrated that zRET<sup>ECD</sup> and zefn-A5 do not directly interact, even in the presence of zGDNF, zGFRα<sub>1</sub><sup>ECD</sup> and mEphA4<sup>LBD</sup>. In order to transduce reverse signalling from efn-As, other mediating proteins, such as Celsr3/Fzd3 and Linx, are required to bind to RET and efn-As, forming a larger signalling complex. *In vitro* validation of this hypothesis is challenging and was beyond the scope of this work because Celsr3/Fzd3 are multipass transmembrane proteins that are difficult to express and purify recombinantly and the formation of such a complex may be sensitive to the membrane environment. This emphasizes the need for future development of recombinant purification platforms for these full-length membrane proteins including RET. However, modern cell-based techniques may prove useful for the investigation and validation of the composition of this complex, such as crosslinking mass spectrometry and proximity ligation assays.

Unlike zRET<sup>ECD</sup>, **Study II** shows that hRET<sup>ECD</sup> expressed using the insect cell expression system is inactive as a result of disulfide mismatch-induced misfolding. Compared to the insect cells, mammalian cells were found to be superior for functional expression of the monomeric hRET<sup>ECD</sup>. The result emphasizes the significance of the choice of the expression host for recombinant protein expression, especially for cysteine-rich proteins. I used HEK293T cells to express hRET<sup>ECD</sup> and Fc-hGDF15, and insect cells for the production of hGFRαL<sup>ECD</sup>. The hRET<sup>ECD</sup>/hGDF15/hGFRαL<sup>ECD</sup> complex was successfully reconstituted by combining the separately purified proteins. A cryo-EM map was reconstructed to a final resolution of 8.0 Å, which agrees well with a published hRET<sup>ECD</sup>/hGDF15/hGFRαL<sup>ECD</sup> structure (**Study IV**). Interestingly, three 3D classes identified during data processing suggest conformational flexibility of the complex. Due to the limitation of the 3D maps, the movements could not be analysed in atomic detail. To pursue high resolution maps in the future reconstruction, more cryo-EM grids should be prepared and optimized for large-scale data collection.

Despite its clinical importance, less is known about the molecular mechanism of the RET oncogenic activation induced by mutations at the RET<sup>CRD</sup> compared to the activation
wild-type hRET^{ECD} by ligand complexation. Study II is the first to show the successful expression of the dimeric hRET^{ECD}(C634R) mutant by utilizing a novel C-terminal Fe-tag approach, which lays a solid basis for the investigation of the oncogenic activation of C634R-mutant RET. In Study IV, expression of the hRET^{ECD}(C634R) dimer was further optimized and a NS-EM map of the dimeric ECDs was reconstructed. This map revealed an “S”-shaped configuration of dimeric hRET^{ECD}(C634R) with the C-termini of the “C”-shaped RET protomers closely associated at the centre. Further binding studies demonstrated that hRET^{ECD}(C634R) can interact with hGDF15/hGFRAL^{ECD} to form an unexpected “twin-butterfly” complex with a molar ratio of 4:4:4. Analysis of the NS-EM maps revealed no significant change of the overall shape of each hRET^{ECD}(C634R) protomer in comparison to that of the hRET^{ECD}(WT) monomer. Because of the limited resolution, it is not clear whether small conformational changes exist in the sub-domains of hRET^{ECD}(C634R)/hGDF15/hGFRAL^{ECD}. Although a cryo EM map of the hRET^{ECD}(C634R)/hGDF15/hGFRAL^{ECD} complex was pursued in this work, significant structural flexibility was observed in the 2D classes, which prevented a high-resolution model of the full complex from being resolved (Study IV). In future work, alternative approaches such as chemical cross-linking and the addition of stabilizing binders could be investigated when studying the hRET^{ECD}(C634R) complexes using cryo-EM. In addition to the hGDF15/hGFRAL^{ECD} complex investigated in Study IV, it will be interesting to characterize whether the other GFLs/GFRαs interact with hRET^{ECD}(C634R) through a similar mechanism.

Lastly, while studies on the RET^{ECD} have provided invaluable insights into the mechanism of RET activation via ligand binding, our knowledge of the full-length receptor activation is still limited due to the lack of structural information for the intact RET and its expression and efficient solubilization is the first problem to tackle. Therefore, a sucrose cushion method, as described in Study III, was adapted and developed for the isolation of detergent-soluble full-length RET. The solubility of RET from the isolated sucrose fraction was 3-6 times higher than that of RET in the total membrane that was prepared without the using this method. Functional assays will be required to evaluate the function of the purified full-length RET in the future, which can subsequently be used for complex reconstitution and structural investigation using cryo-EM. Furthermore, Study III provides detailed membrane protein production protocols that can be adapted for the expression and purification of a wide range of membrane proteins using E. coli and insect cells. The methods described are expected to be tested for the production of other membrane proteins, including RTKs.
VI REFERENCES

4. Ward, C. W., Lawrence, M. C., Streletsoy, V. A., Adams, T. E. & McKern, N. M. The insulin and 
5. Wehrman, T. et al. Structural and mechanistic insights into nerve growth factor interactions 
6. Yuzawa, S. et al. Structural basis for activation of the receptor tyrosine kinase KIT by stem cell 
   1134 (2010).
10. Mol, C. D. et al. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine 
13. Takahashi, M., Ritz, J. & Cooper, G. M. Activation of a novel human transforming gene, ret, by 
14. Hahn, M. & Bishop, M. J. Expression pattern of drosophila ret suggests a common ancestral origin 
    between the metamorphosis precursors in insect endoderm and the vertebrate enteric neurons. 
15. D’Aniello, S. et al. Gene expansion and retention leads to a diverse tyrosine kinase superfamily 
18. Li, J. et al. Cryo-EM analyses reveal the common mechanism and diversification in the activation 
    of RET by different ligands. Elife 8, e47650 (2019).
20. Anders, J., Kjaer, S. & Ibáñez, C. F. Molecular modeling of the extracellular domain of the RET 
    receptor tyrosine kinase reveals multiple cadherin-like domains and a calcium-binding site. J. Biol. 
21. Goodman, K. M. et al. RET recognition of GDNF-GFRα1 ligand by a composite binding site 
23. Kjaer, S., Hanrahan, S., Totty, N. & McDonald, N. Q. Mammal-restricted elements predispose 
24. Liu, Y., de Castro Ribeiro, O., Robinson, J. & Goldman, A. Expression and purification of the 
    extracellular domain of wild-type humanRET and the dimeric oncogenic mutant C634R. Int. J. 
25. Plaza-Menacho, I. et al. RET functions as a dual-specificity kinase that requires allosteric inputs 
42. Masire, S. et al. Mammalian GFRα-4, a divergent member of the GFRα family of coreceptors for glial cell line-derived neurotrophic factor family ligands, is a receptor for the neurotrophic factor persephin. J. Biol. Chem. 275, 39427–39434 (2000).
50. Wu, Z.-S. et al. Prognostic significance of the expression of GFRα1, GFRα3 and Syndecan-3,


143. Cranston, A. N. et al. RET is constitutively activated by novel tandem mutations that alter the active site resulting in multiple endocrine neoplasia type 2B. Cancer Res. 66, 10179–10187 (2006).
170. Morabito, A. et al. Vandetanib (ZD6474), a dual inhibitor of vascular endothelial growth factor receptor (VEGFR) and epidermal growth factor receptor (EGFR) tyrosine kinases: Current status

77
244. Chen, J., Noble, A. J., Kang, J. Y. & Darst, S. A. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and
production and purification enabling membrane protein structural biology. Biochem. Cell Biol. 94,
507–527 (2016).
277. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: Not just a
278. le Maire, M., Champeil, P. & Møller, J. V. Interaction of membrane proteins and lipids with
281. Gewering, T., Januliene, D., Ries, A. B. & Moeller, A. Know your detergents: A case study on
285. Sonoda, Y. et al. Tricks of the trade used to accelerate high-resolution structure determination of
286. Hanson, M. A. et al. A specific cholesterol binding site is established by the 2.8 Å structure of the
287. Dürr, U. H. N.; Soong, R. & Ramamoorthy, A. When detergent meets bilayer: Birth and coming of
292. Morrison, K. A. et al. Membrane protein extraction and purification using styrene–maleic acid
294. Postis, V. et al. The use of SMALPs as a novel membrane protein scaffold for structure study by
Methods 13, 345–351 (2016).
296. Carlson, M. L. et al. The Peptidisc, a simple method for stabilizing membrane proteins in detergent-
297. Angiulli, G. et al. New approach for membrane protein reconstitution into peptidiscs and basis for
their adaptability to different proteins. Elife 9, e53530 (2020).
298. Danev, R. et al. Routine sub-2.5 Å cryo-EM structure determination of B-family G protein-coupled